NORTHEAST DECISION SCIENCES INSTITUTE CONFERENCE PROCEEDINGS

40ᵀᴴ ANNUAL MEETING

Montréal, Canada   **April 14 – 16, 2011**

Proceedings Editor

Albert E. Avery, Towson University

Program Chair

Minoo Tehrani, Roger Williams University

AUDIT FEES, EXTERNAL AUDITOR'S REPUTATION AND CHARACTERISTICS OF THE DIRECTORS' BOARD: THE CASE OF BELGIAN CONTEXT

Azhaar LAJMI
Institut Supérieur de Gestion de Tunis, Tunisia
Doctorante
11 Rue Grmbalia Soukra, Ariana 2073, Tunis.
azhaar.lajmi@isg.rnu.tn

Marjène GANA
Institut Supérieur de Commerce et de Comptabilité de Bizerte. Tunisia
Maître de conférences
gana.ym@gnet.tn

ABSTRACT

This study investigates the effects that governance mechanisms have on the quality of external audit on a sample of 97 Belgian listed companies during over the period 2003-2007. More specifically, governance mechanisms are described by the board of directors (composition, independence, duality and frequency of meetings). The results suggest that the size and the independence of directors' board are the most important factors to understand the choice of the external audit.

Keywords: governance, audit fees, Big four, board of directors.

I. INTRODUCTION

Governance mechanisms permit us to limit the discretion of managers and enhance business efficiency. However, recent financial scandals and accounting changes all around the world (Enron and WorldCom in the United States, Parmalat in Italy, Lernout & Hauspie in Belgium, etc.) and the irregular financial statements of certain listed companies are evidence of a certain relaxation of the managers in professional ethics and calls into question the effectiveness of governance mechanisms. This sparked a confidence crisis calling for the introduction of enhanced surveillance systems to ensure the credibility of financial statements.

Charreaux (1997) defines governance as "the set of mechanisms that effectively circumscribe the powers of influence and leadership decisions which otherwise control their own behavior and define their discretionary space".

Thus, the concept of corporate governance, based on willingness for transparency, first allows better management and then reconciliation of possibly divergent interests within the firm. Therefore it is essential, after the recent world economy being recently turmoil, to restore confidence to the different stakeholders.

Fama and Jensen (1983) identified two important functions that are assigned to the board of directors: the evaluation of decisions and supervision of officers.

These objects could be achieved by the presence of independent directors, separation of the functions of CEO and chairman of the board, the function of the board in terms of meetings conducted, Moreover, the structure of corporate governance would provide insurance to stakeholders that the CEO is well supervised and should act in their interest. Several authors have studied the structure of board of directors including in their research the structure of audit committees (Smith 2006). In addition, Abbott et al. (2003) and Carcello et al. (2002) showed that firms with good internal governance structures require a high audit quality and thus they pay higher fees. De Angelo (1981b) defines audit quality as the probability that an auditor discover a violation in the financial statements of the company, and reveals it publicly.

The issue of audit quality is at the heart of the theory of the agency and governance and has been widely discussed in both Anglo-Saxon and European environments. Recent studies (Boo and Sharma 2008; Hay et al. 2008; Piot 2005; Krishnan and Visvanathan 2009; Knechel and Willekens 2006; Salleh et al. 2006) investigate the impact of some factors of governance on the external audit quality. In fact, the agency theory explains the use of an external audit of quality as a mechanism to reduce managerial opportunism and limit conflicts of agency between the different actors of the firm.

In this connection, the agency theory focuses on a firm where the separation between ownership and control is important. This is the case of the Belgium context. In such a situation, the opportunistic behavior of managers becomes clear and conflicts of agency occur. In this context, the quality of the external audit plays a major role in reducing conflicts and in protecting shareholders' interests (Jensen and Meckling 1976).

Through our literature review, we found that external audit is an important mean of solving agency problems between the partners of the firm by producing more credible financial statements. In addition, an external audit quality is likely to increase users' confidence in financial and accounting reporting.

However, and because of the complexity of the audit process, the majority of studies were based on indicators or substitutes for measuring the external audit quality by using an indirect approach for assessment (DeAngelo 1981; DeFond 1992; Piot 2005). The most commonly used criterion to attain this quality was the reputation of the audit firm meaning its affiliation to one of the "Big four" ("Big eight" in 1980). Previous research has shown that large audit firms have both human and physical resources that allow them to guarantee the quality of financial and accounting reporting. Therefore, companies that have agency problems and important agency costs are more likely to engage reputable audit firms (Firth and Smith 1992). Moreover, the financial market has more confidence in companies that employ a Big4 audit type because such auditors guarantee credible financial statements and allow business partners to take right decisions.

Other studies have used alternative attributes of the external audit quality, such as the auditor industry specialization (Beasley and Petroni 2001), the amount of fees paid (O'Sullivan 2000), the size of the audit firm measured by the number of customers (Al-Ajmi 2008). Carcello et al. (2002) reported that the audit fees reflect the costs associated with effective auditors. These costs also vary depending on the size of the audit firm and its complexity (Simunic 1980).

In addition, other researchers have used the fees as a measure of audit quality (Bariotta 2000; O'Sullivan 2000; Carcello et al. 2002; Abbott et al. 2003 ; Salleh et al. 2006; Yatim et al. 2006; Hay et al. 2008): firm reputation and audit fees are signals of audit quality.

Regarding the relationship between these two concepts, Hay et al. (2008) examined the existence of "complementarities" or "substitutability" between corporate governance and external audit. Their objective was to study the relationship between corporate governance and audit fees.

In general, the literature shows that the relationship between external audit and corporate governance mechanisms (outside directors and audit committee) is complementary. The independent administrators ask for more audit and better governance mechanisms to produce high quality financial statements.

This study contributes to line stream of research on the impact of corporate governance on the external audit quality of the Belgian listed companies. More precisely, the main objective is to understand the influence of the board of directors on the external audit quality (complementarities or substitutability). In addition, we will highlight the importance of the relationship between the board and audit quality measured by fees paid to auditors and by their affiliation to one of the «Big four».

We organize our article as follows. First, we present the state of the art and the theoretical framework related to the board of directors characteristics and the choice of auditor. Then we formulate our hypotheses, while identifying variables and proxy for audit quality. In the third section, we discuss methodological aspects (sample presentation and research methodology). In section VI, empirical results are presented and discussed. The last section summarizes the study and provides concluding comments and implications.

II. BOARD OF DIRECTORS CHARACTERISTICS AND AUDITOR CHOICE

Throughout our study, we identified four essential characteristics of the board of directors: independence, the dual structure of the management board, size and diligence.

The board of directors' independence

The inside directors are hierarchically dependent on the managers. However, according to the Vienot II report, an independent outside director is one who has no relationship with the company or its affiliates that could compromise the exercise of his freedom of judgment. In other words, this director doesn't have any role in the firm except being an administrator, controlling the management and protecting the shareholders' interests. The practices of good governance- especially those cited in the Lippens code (2005) in Belgium- require the existence of a significant number of independent directors on the board of directors. In Belgium, under the Code of Corporate Governance (2009), we assume that at least half of the board should consist of non-executive directors (a non-executive director is one who does not have executive responsibilities in the company) and at least three of them should be independent (free from any relationship work related, or any other with the company.
Thus, it becomes relevant to know if the independence of the board constitutes a complementary or substitutable control mechanism for external audit quality.

Independence is assessed by the percentage of independent members of the board. Several papers have focused on the study of the presence of external members influence on the board of directors to appoint better auditors. Contrary to the substitution argument, the literature generally supports the idea that the relationship between external audit and corporate governance mechanisms, such as the independence of the board of directors and the audit committee, is complementary.

According to Hay et al. (2008), a company, whose stakeholders want to improve their control and governance, may begin with the appointment of competent independent directors. These tend to ensure the interests of other stakeholders and have to protect their own reputation and therefore they tend to choose an external audit quality.
Based on the theory of complementarities, Beasley and Petroni (2001) examined the relationship between external members of the board and the choice of external auditor. They noted that the independent external members of the board, which have greater surveillance of management, influence the board to choose a better quality of auditors.

In addition, Carcello et al. (2002) pointed out that independent boards of directors tend to choose an external audit which offers a good quality of control, or a «Big four» audit firm because administrators seek to protect their capital, avoiding any situation affecting their legal liability and protect the interests of shareholders.

In New Zealand, Hay et al. (2008) showed that the measures of internal control and corporate governance are positively related to audit fees. This is consistent with the theory of complementary control measures. In the United Kingdom, O'Sullivan (2000) found that the independence of the directors has a positive impact on external audit quality measured by audit fees. In the American context, Abbott et al. (2001) and Carcello et al. (2002) supported the fact that the independence of the board of directors has a positive impact on audit quality measured by the remuneration of the auditors. Moreover, they showed that a more independent board of directors tends to protect its reputation and to protect the interests of shareholders by requiring a higher audit quality and paying higher fees. The successive works of Lennox (2005) and Goodwin-Stewart and Kent (2006) confirmed the positive relationship between board independence and audit quality, respectively approximated by the membership of international «Big four» auditors and audit fees. They showed that the independent members contribute to the effectiveness of the control process and influence decisions regarding the choice of the external auditor.

These empirical results confirm the hypothesis of complementarities between independence of the directors and external audit quality with a positive relationship between governance mechanisms and external audit quality. Researchers supporting the substitutability hypothesis which establish a negative relationship between internal audit or corporate governance and external audit are few. This brings us to our first hypothesis:

H_1: There is a positive relationship between independent directors of the board and external audit quality.

Duality of directors' board

In addition to the independence of the directors' board, its structure is also one of the aspects that characterize corporate governance. Two forms of board structure deserve attention: the duality and separation of the functions of chairman and chief executive officer (CEO) of the firm. In fact, the duality of the board implies that the same person holds both the chairman function and CEO function. However, dissociation implies the existence of two individuals each occupying one position.

In the literature, these two aspects appear to be controversial because some authors opt for this separation of functions while others recommend a combination of these two roles by the same person.

In Belgium, the Lippens Code (2005) followed by the Belgian Code of Corporate Governance (2009) argue that a clear separation of responsibilities at the head of the firm is drawn between the conduct of the board and the conduct of firms activities. Thus, the chairman of the board and CEO functions cannot be occupied by the same person. In this context we appreciate the duality of the board by a dummy variable that equals 1 if the same person holds both functions of CEO and chairman.

Moreover, an examination of board duality claims that this duality may have a significant effect on the demand for audit quality. It is necessary to note that the board has the responsibility to engage external auditors in the case of a dual structure of the board (Ashbaugh and Warfield 2003). In fact, according to the agency theory, several authors like Abdullah (2004) and Peel and Clatworthy (2001) indicated that company's shareholders tend to choose a higher quality of audit when companies are managed by a person who is also the president of the board of directors. Thus, the existence of the duality of boards is considered as a sacrifice of shareholder interests in favor of company managers, which will lead to agency conflicts. These agency problems encourage owners to review the management of the company, to deepen the external control mechanisms and so to recruit highly skilled auditors. Furthermore, Dechow et al. (1996) and Fama and Jensen (1983a, 1983b) considered that when a board of directors is chaired by the CEO, the auditor needs to make more efforts and subsequently apply high fees. Specifically,

Dechow et al. (1996) showed that firms identified as manipulators of their results usually have a dual structure of their board. This suggests that the combination of these functions is likely to compromise the effectiveness of the board in their tasks of control (management, accounting). Therefore, the auditor spends more time and more control effort and so, requires higher fees.

Other researchers have suggested that companies, experiencing a duality structure, pay low fees for audit and subsequently, don't look for high reputation audit firms (Yatim et al. 2006).

The study of Tsui et al. (2001) examined the relationship between the duality of the board and audit fees in Hong Kong. The results supported the hypothesis that firms with a separation of the functions of CEO and chairman are associated with significantly lower audit fees.

However, in the American context, Krishnan and Visvanathan (2009) used in their study the variable "non duality" to indicate the separation between the positions of chairman of the board and CEO, and showed that this variable is significant and positive: the separation between these two functions reduces the risk of failure of corporate governance, and so reduces the auditor's fees. Nevertheless, in the United Kingdom, O'Sullivan (2000) didn't find any result that supports that the duality of the board of directors has a significant impact on audit fees. Similarly, in a subsequent study and in the same context, O'Sullivan and Diacon (2002) didn't find any significant relation between the duality of the board and the audit fees for insurance firms. Salleh et al. (2006) provided the same evidence in the Malaysian context. Our second hypothesis is expressed as follow:

H_2: There exists a positive relationship between duality structure of the board of directors and external audit quality.

Size of the board of directors

The size of the board refers to the number of inside and outside directors. The agency theory advocates that the smaller the board, the better it is, meaning that which is composed of only a few administrators. Jensen (1993) argued that the effectiveness of the board requires in particular the choice of its appropriate size. He revealed that large boards composed of a large number of administrators result in the domination of managers and create conflicts. Moreover, it will be difficult to find a consensus in decision making between managers causing difficulties in protecting shareholders' interests.

Empirical studies of the relationship between size of the board of directors and audit quality, gave controversial results. Indeed, size of the board could affect the process of financial reporting and auditing process. Generally, large boards have less effective controllers. As a consequence, external auditors spend more time doing their job, and high fees are charged (Beasley 1996). This author found that the size of the board significantly affects the risk of fraud in financial statements. So, the auditor devotes more time to his control mission, which leads to higher audit fees.

Klumpies and Gul (1999) stated that large size of the board of directors is associated with a low audit quality measured by the low fees paid to auditors. Consequently, the chances of employing reputable auditors are reduced. Krishnan and Visvanathan (2009) found that board size and its independence are two insignificant variables on the audit quality measured by audit fees. Similarly, Yatim et al. (2006) found non-significant results regarding the relationship between the size of the board and audit fees. Hence, our third hypothesis is:

H_3: There exists a positive relationship between size of the board of directors and external audit quality.

Diligence of the board of directors

The literature on governance has focused on board independence and size as attributes that are valuable to the board of directors. Our study adds to this line of research the number of meetings conducted annually by the board as an attribute of his diligence. Thus, the diligence of the board of directors is approximated by the number of meetings conducted annually.

The intensity of activities of the board of directors may contribute to the effectiveness of its supervisory functions in particular concerning the financial reporting process. Some researchers have argued that boards that meet frequently are more often likely to perform their duties diligently and are beneficial to shareholders (Lipton and Lorsch 1992; Byrne 1996). Therefore, the board of directors that is more diligent in exercising its supervisory duties should normally improve the monitoring of the financial reporting process. It would then be expected that boards that meet frequently are negatively associated with external audit fees (Yatim et al. 2006).

In the American context, Carcello et al. (2002) found that board diligence is associated with higher fees. In the same context, Krishnan and Visvanathan (2009) showed that the two variables measuring the activity of the board of directors and audit committee (number of meetings conducted annually by the board and the audit committee) are significant and associated positively with audit fees.

However, in the Malaysian context, the result of Yatim et al. (2006) indicates that diligence of the board measured by the number of meetings conducted per year is not significantly associated with audit fees. Consequently, we present our last hypothesis:

H_4: There exists a negative relationship between diligence of the board of directors and external audit quality.

III. RESEARCH METHODOLOGY

Having developed the conceptual framework and presented our hypothesis, we present in the following section methodological aspects of our study. We successively describe our sample and dependent and independent variables.

Sample and study period

The sample of our study includes ninety-seven (97) Belgian companies listed on the Euronext Brussels Stock Exchange over the period 2003[1]-2007. These are companies for which we were easy to access information. Data were collected manually from annual reports available on the companies' website. Contrary to other studies, we do not exclude any sector.

Definition and measures of variables

Dependent variables: reputation of the auditor and audit fees
Given the difficulties of observing the audit quality, some researches have used the firm's reputation, that is to say whether it belongs to "BigN" (DeFond 1992; Craswell et al. 1995; Beasley and Petroni 2001; Piot 2005) or its size, while others use the fees as a measure of audit quality (Bariotta 2000; O'Sullivan 2000; Carcello et al. 2002;

---

[1] We chose to start our analysis in 2003 because of the financial scandals like the one which affected the Belgian company Lernout & Hauspie. Besides, the legislative reforms in Belgium concerning governance was introduced in 2002 and continued thereafter with the royal decrees of April 4, 2003 to arrive at the Lippens Code 2005 and Code of Good Governance in 2009.

Abbott et al. 2003; Salleh et al. 2006; Yatim et al. 2006; Hay et al. 2008): firm size and audit fees become signals of quality.

-Audit fees (**FAUD**): is a metric variable approximated by the natural logarithm of audit fees (Carcello 2002; Craswell et al. 1995; Simunic 1980). Palmrose (1986) showed a significant positive relationship between audit fees and audit company size.

-Reputation of the auditor (**RAUD**): is a binary variable equal to 1 if audit firm belongs to the "Big four" and equal to 0 otherwise. Like previous research (Francis and Wilson 1988; DeFond 1992), we hypothesize that large audit firms, who belong to the "Big Four" provide services of better quality because the "Big Four" are required to maintain their reputation by being independent and providing a good control of quality (Watts and Zimmerman 1983).

Independent variables

They are relative to the board of directors described as follow:

-Independence of the board of directors (**INDE**): this is the proportion of non-executive members on the board (Beasley and Petroni 2001; Carcello et al. 2002). In this context, some authors such as Abbott and Parker (2000), Beasley and Petroni (2001), Ireland and Lennox (2002) and Lennox (2005) argued that audit quality measured by the type of audit firm or its reputation, increases with the percentage of outside directors.

-Duality of the board (**DUAL**): is the accumulation of the functions of CEO and Chairman of the board. The duality of the board is a dummy variable that takes the value of 1 if the same person holds both positions, and 0, if not (Baliga et al. 1996; Daily and Dalton 1994).

-Size of the board of directors (**SIZE**): following Yermack (1996), this variable is measured by the total number of directors on the board. Jensen (1993) found that a large board causes difficulty in reaching a consensus on important decisions. Therefore, protection of interests of shareholders would be reduced. Yatim et al. (2006) and Krishnan and Visvanathan (2009) found no significant results concerning the relationship between the board size and audit quality. A board consisting of a large number of directors makes the process of sharing information and decision-making more difficult. In this context, a large board affects its decision to choose a better quality of external audit.

-Diligence of the board (**DILI**): This is the number of meetings conducted annually by the board of directors. Vafeas (1999) showed that a large number of meetings contributes to the effectiveness of the board. According to Carcello et al. (2002), a more diligent board is positively associated with audit quality measured by audit fees. Yatim et al. (2006) established that the relationship between the diligence of the board and audit fees is negative.

It seems appropriate to add other explanatory variables describing the financial director (age and time of occurrence) that may just affect the quality of external audit. These variables are successively:

-AGE and **TPRE**: They are supposed to represent the possible entrenchment of CFO. The attention to the CFO reflects the important role given to him in the firm (Zorn, 2003).

Two theoretical frameworks prevail: on one side, the time spent and / or aging may promote negative entrenchment of CFO as he will perhaps reduce the procedures highlighting his incompetence (Shleifer and Vishny 1989; Paquerot 1996; Pigé 1998). On the other hand and referring to the positive entrenchment perspective, the time spent and / or aging of the CFO may have enabled him to consolidate its network of knowledge (Maati 1999). For our part, we suppose the existence of a positive relationship between aging of the CFO and his time of presence and audit quality.

To enhance the explanatory power of our model, we consider the following control variables:

- Leverage of audited company (**LTDTA**) is the debt ratio that is measured by the long term debt to total assets ratio. Leverage is generally used to proxy for the agency costs of financial debt. Seen from the perspective of agency theory, debt can be considered as a solution to the conflict management -shareholders (Jensen and Meckling 1976), and this by its contractual obligation. The auditor comes precisely to assert respect with the terms of the contract. Therefore, the role of the auditor increases more and more when the debt increases. Its is the same

when conflicts between shareholders and debtholders. In this sense, Piot (2003, 2005) and Velury and al. (2003) argue that it is useful to use a better audit quality ensuring the credibility of financial statements and the respect with contract terms. The relationship between leverage and audit quality would be positive.

- Profitability of audited company (**PROF**): it is the companies' ability to generate returns from its assets. It is a proxy for firm performance that it is defined by the profit before tax and interest earned on total assets. To measure this variable, we are inspired by studies of Joshi and Al-Bastaki (2000), Goodwin-Stewart and Kent (2006) and Skinner and Srinivasan (2010). According to these authors, there is a significant and positive relationship between profitability and audit fees. The managers of the most profitable companies are therefore encouraged to report the quality of accounting information through a quality of audit.

IV. DESCRIPTIVE STATISTICS

Table 1 presents descriptive statistics for variables used above. The Panel A of this table presents the continuous variables, while Panel B is for dummies.

According to the Panel A of table 1, the audit quality measured by audit fees is highly volatile. To estimate the regression's models, the natural logarithm of this variable is used to reduce its scatter. Regarding our second attribute of the quality of external audit presented in Panel B, we note that the majority of Belgian listed companies (78.52%) are audited by firms with a high reputation. This clearly shows that there is a high degree of concentration of the audit market in Belgium. The «Big 4» firms are auditing approximately 78.52% of the sample of Belgian listed companies during the period 2003-2007.

Table 1: Descriptive statistics of variables of the board of directors (2003-2007)
The descriptive statistics of our sample composed of 97 listed Belgian companies are presented in Table 1.
Panel A-Continuous variables: descriptive statistics for continuous variables
(Number of observations, mean, standard deviation, minimum and maximum) are presented in Panel A.

Variables	Mean	Standard deviation	Minimum	Maximum
FAUD	12.385	2.218	6.697	19.489
INDE	0.427	0.173	0	0.929
SIZE	10.233	4.353	4	28
DILI	7.788	4.150	1	35
AGE	45.590	7.518	30	65
TPRE	3.601	0.982	0	5.529
DLTTA	0.177	0.153	0	0.705
PROF	0.053	0.126	-0.629	0.904

Panel B - Dichotomous variables: descriptive statistics of binary variables (frequency and percentage) are presented in Panel B.

Variables	Modality	Frequency	Percentage
RAUD	1	329	78.52%
	0	90	21.48%
Total		419	100%
DUAL	1	102	24.46%
	0	315	75.54%
Total		417	100%

In addition, we find that, on average, boards of companies in our sample consist of 10 members, 42.7% of them are independent non-executive directors with a minimum of 0% and a maximum of 92.9%. This result suggests that, overall; the boards of directors of Belgian listed companies comply with best practices recommended by the Belgian Code of Corporate Governance (2009). Concerning diligence, the boards meet on average eight times per year. Furthermore, the age of the CFO is between 30 and 65 years and has been present in the company for an average of 3 years.

In addition, the weight of debt is relatively low on our entire working sample with a debt which represents an average of about 20% of total assets. As for profitability, it appears week and show an average of 5.3%.

In Panel B of table 1, the statistics show the frequencies and percentages of dichotomous variables. We find that 24.46% of the firms merge the functions of CEO and chairman, so the separation of the functions of CEO and chairman is observed in approximately 75.54% of Belgian listed firms. This conclusion is consistent with the Belgian Code of Corporate Governance (2009). In general, boards of Belgian companies do not seem to be dominated by a single person.

V. RESEARCH MODELS AND RESULTS OF REGRESSIONS

Research Models

To document the relationship between the board characteristics and external audit quality approached successively by the fees paid by the firm and by the auditor's reputation, two regression models are suggested. The first one is a linear regression model whose dependent variable is metric. The second model is a logistic regression model where the dependent variable is binary. These models are inspired from Simunic (1980) tested in the U.S. market which was subsequently modified by Salleh et al. (2006) in the Malaysian market.

Model 1:
$$\text{FAUD}_{it} = \beta_0 + \beta_1 \text{INDE}_{it} + \beta_2 \text{DUAL}_{it} + \beta_3 \text{SIZE}_{it} + \beta_4 \text{DILI}_{it} + \beta_5 \text{AGE}_{it} + \beta_6 \text{TPRE}_{it} + \beta_7 \text{DLTTA} + \beta_8 \text{PROF}_{it} + \mathcal{E}_{it}$$

Modèle 2 :
$$\text{RAUD}_{it} = \beta_0 + \beta_1 \text{INDE}_{it} + \beta_2 \text{DUAL}_{it} + \beta_3 \text{SIZE}_{it} + \beta_4 \text{DILI}_{it} + \beta_5 \text{AGE}_{it} + \beta_6 \text{TPRE}_{it} + \beta_7 \text{DLTTA} + \beta_8 \text{PROF}_{it} + \mathcal{E}_{it}$$

Where:
FAUD: audit fees is an attribute of external audit quality which is equal to the natural logarithm of fees paid to auditors.
RAUD: auditor reputation represents the second measure of external audit quality which is coded 1 if the firm belongs to one of the Big 4, coded 0, otherwise.
INDE: represents the independence of the board of directors and is measured by the proportion of non-executive outsider directors relative to the total number of directors;
SIZE: the size of the board of directors measured by the total number of directors;
DILI: the diligence of the board that is defined by the number of meetings conducted annually by the board;
AGE: age of the CFO;
TPRE: the tenure of CFO is measured by the natural logarithm of the time (months) taken by the financial director to take his position in the company;
LTDTA: the debt ratio that is measured by the long term debt to total assets ratio;

PROF: the profit before tax and interest earned on total assets.

Ɛ is the term of error;

The β are the coefficients to estimate;

The index **i** corresponds to the selected companies in our sample and the index **t** for the period of the study.

Results

The table 2 below shows the results of our multiple regressions to determine the effect of board characteristics on the quality of external audit in the context of Belgian listed companies.

The results displayed in this table demonstrate a significant (at a 1% level) positive association between independence of the board and external audit quality confirming our first hypothesis and suggesting a complementary association between the proportion of independent directors on the board and external audit quality (fees paid and reputation of the firm).

Table 2: Effect of characteristics of the board of directors on external audit quality

HAUD is the measure of the external audit quality. **RAUD**: auditor reputation represents the second measure of external audit quality. **INDE** is the independence of the board of directors. **SIZE** is the size of the board of directors. **DILI** represents the diligence of the board. **AGE** is the age of the CFO. **TPRE** is the tenure of CFO. **LTDTA** represents the debt ratio that is measured by the long term debt to total assets ratio. **PROF** the profit before tax and interest earned on total assets. + Denotes a positive relationship, - means a negative relationship, 0 indicates a non-significant relationship.*** Significant at 1%, ** significant at 5%, * significant at 10%.

Variables expected	sign	Mode1(FAUD)			Model2 (RAUD)		
		Coef.	P>|z|	Sig. obtained	Coef.	p>|z|	Sig. obtained
INDE	(+)	3.061	0.000***	(+)	16.199	0.004***	(+)
DUAL	(+)	0.475	0.049**	(+)	26.551	0.999	0
SIZE	(+)	0.282	0.000***	(+)	0.717	0.007***	(+)
DILI	(-)	- 0.000	0.964	0	0.146	0.285	0
AGE	(+)	0.028	0.025	(+)	-0.026	0.607	0
TPRE	(+)	-0.001	0.978	0	0.828	0.028**	(+)
ENDE	(+)	0.209	0.697	0	-2.088	0.494	0
PROF	(+)	-0.629	0.115	0	1.604	0.405	0
β₀		7.688	0.000		-13.384	0.007	
Significance		0.000			0.000		

Our results confirm those previously cited by Abbott and Parker (2000), Lennox (2005) (approaching the quality of the audit by the reputation / size of the firm) and also by Knechel and Willekens (2006) (measuring the quality by audit fees paid) in the Belgian context and are also consistent with those of O'Sullivan (2000) and Salleh et al. (2006) in Malaysia.

It can be thus established that the independent members of the board of directors look for a high quality of audit services to better protect their reputation and to protect the shareholders' interests. Higher audit fees are justified by the long time they spend conducting supervisory duties of high quality. The large presence of executive directors is an important attribute of audit quality that appear to affect it negatively: the executives may have high confidence in the internal control system imposed by laws and do not perceive the necessity to apply some external procedures of audit

We conclude that the independence of the board of directors has an impact on the external audit quality by influencing the choice of an auditor of great reputation "Big four" and paying higher fees for a better control. The independent directors of boards of Belgian firms want to strengthen supervision to produce better financial statements in order to preserve their reputation. Board independence and external audit are, in addition, two complementary control mechanisms. This thesis of complementarities is supported by Carcello et al. (2000) in the United States and by Hay et al. (2008) in New Zealand. Our result is also consistent with the results of Piot (2005) for a sample of Canadian firms over the period 1998-1999, and with those obtained by Adjaoud et al. (2008) in the same context over the period 2002-2004.

Concerning the CEO/Chairman functions accumulation, the results reveal that the coefficient of DUAL variable is significant and positive with audit fees but not significant reputation of auditor which invalidates our second hypothesis concerning the positive relation between the duality of the board and audit fees. This result corroborates those of Carcello et al. (2000) in the American context, O'Sullivan (2000) in the British context, Salleh et al. (2006) in the Malaysian context and Adjaoud et al. (2008) in the Canadian one. This can be explained by the fact that, according to the descriptive statistics, only 24.46% of Belgian firms combine the functions of chairman and CEO.

A positive and statistically significant impact of the size of the board of directors on audit quality (approached by audit fees and auditor reputation) is found to confirm our third hypothesis. Our finding confirms the predictions of the agency theory which suggests that a small board of directors leads to more efficiency and so more coordination and better communication between members of the board. A less demand for an external audit is therefore established (Fama and Jensen 1983).
A large board of directors corresponds to a larger company (these two variables were found to be correlated[2]) and therefore, the economic and financial issues are more important than in a smaller company. Thus, the Belgian directors on a large board influence the quality of external audit by choosing a "Big four" firm and paying higher fees. Our results also confirm the complementarities thesis of both mechanisms.

Concerning diligence of board of directors (measured by the number of meetings), we find no statistically significant effect on audit quality leading to the rejection of the fourth hypothesis. This result contradicts that found by Carcello et al. (2002) in an American context. These authors argued that a more diligent board of directors in carrying out its responsibilities requires a heightened level of monitoring process, financial information and therefore requires a high level of external audit quality. However, our result is consistent with those obtained by Yatim et al. (2006). They showed that the diligence of the board is not significantly associated with audit fees. This suggests that the choice of the audit process results from the daily management of the company and therefore does not require an increase in the number of meetings of the board.

Concerning the variables describing the financial officer – the person supposed to manage, or at least guide the auditing of the company - we observe no statistically significant result for the time spent in the company for the first model. This variable was introduced to characterize the possible entrenchment in his negative or positive vision. Several interpretations of this result can be given: either the theory of the entrenchment is not an appropriate theoretical framework for analyzing the process of making decision of the CFO or the CFO is not the person for whom the possible entrenchment should be examined concerning the audit policy of the company.
Nevertheless, regarding the variable AGE, and referring to the first model whose dependent variable is the amount of audit fees, age advancement may promote positive entrenchment of the CFO. This could increase his control procedures to show his willingness to properly manage the company in the interest of the shareholders. In other

---

[2] Accourding to the test of Variation Inflation Factor (VIF), we found that these variables are correlated because their values are > 3(Stowoly and Ding 2003). Then we had to remove the variable «size» to get a better estimate of our model.

words, this result could indicate that confidence of the CFO in external procedures is an increasing function of age. However, this result is not conclusive because of the lack of homogeneity of the effect of this variable in our various regressions. Otherwise, we do not validate our assumption.

Finally, control variables have an insignificant effect on the external audit quality. For profitability, this result does not corroborate the findings of several studies (Abbott and Parker 2000; Piot 2001; Lennox 2005). Thus, the profitability of a company, whether good or bad, does not influence the choice of the audit quality.

We also point out the overall significance of both regression models since the Chi2 test is significant at 1% level.

VI. SUMMARY, IMPLICATIONS AND LIMITATIONS

In this study, the relationship between the characteristics of the board of directors (independence, duality, size and diligence) and audit quality in Belgium (as measured by the audit fees and reputation of the firm Big/NBig) was examined.

Using a multiple regression model and a logistic regression model to analyze the various relationships, we found that the independence of the board is a significant indicator of audit fees and auditor reputation. As expected, a higher proportion of independent directors on boards is associated with higher audit fees paid by Belgian companies and with the membership of audit firms to one of the «Big four».
This is consistent with the fact that higher audit fees better reflect audit quality because auditors spend more time in their control mission to produce a detailed audit report. Moreover, the choice of the auditor depends positively on the degree of board independence: The independence of the board can thus protect the shareholders' interests and control the behavior of management by using a higher audit quality (Jensen and Meckling 1976).
The results suggested also a positive relationship between the size of the board of directors and audit quality. To this end, the directors want to strengthen controls in the firm using a better external audit quality. The non separation of functions of CEO and chairman is not significantly related to auditor reputation. The same result was highlighted in the relation between meeting frequency and audit quality.

The contribution of this research adds to our current understanding of the impact of some characteristics of the board of directors on the audit quality in the Belgian context. While audit fees have been already the subject of several studies, to our knowledge, in Belgium, there is no study using governance mechanisms and firm size as explanatory variables of external audit quality. In addition, in this study, we expend previous researches that have mainly focused on the composition of the board of directors, including variables concerning its diligence and size.

However, this study has limitations. Our work considered two measures of audit quality. This research could be improved by providing a global index of external audit quality. In addition, other variables could be introduced, particularly those concerning firm activity sectors and their risk level.
Finally, our study uses a quantitative approach without taking into account more qualitative factors, especially the level of real power of relations in the directors' board.

REFERENCES

Abbott, L.J., and S. Parker. 2000. Auditor selection and audit committee characteristics. *Auditing: A Journal of Practice and Theory:* 47-66.

Abbott, L. J., S. Parker, G. F. Peters and K. Raghunandan. 2003. The association between audit committee characteristics and audit fees. *Auditing: A Journal of Practice and Theory* 22 (2): 17-32.

Abdullah, S. N. 2004. Board composition, CEO duality and performance among Malaysian listed companies. *Corporate Governance* 4 (4): 47-61.

Adjaoud, F., C. Mamoghli, and F. Siala. 2008. Auditor reputation and internal corporate governance mechanisms: complementary or substituable? *Review of Business Resarch.*

Al-Ajmi, J. 2008. Audit and reporting delays: Evidence from an emerging market. *Advances in Accounting, incorporating Advances in International Accounting* 24: 217-226.

Ashbaugh, H., and T. Warfield. 2003. Governance mechanism: evidence from the German market. *Journal of International Accounting Research* 2: 1-21.

Baliga, B.R., R. C. Moyer, and R. S. Reo. 1996. CEO duality and firm performance: what's the fuss? *Strategic Management Journal* 17: 41-53.

Bariotta, L. 2000. The Audit Committee Handbook. John Wiley et son Inc., New York.

Beasley, M. 1996. An empirical analysis of the relation between the board of director composition and financial statement fraud. *The Accounting Review* 71(4): 443-465.

Beasley, M., and K. Petroni. 2001. Board independence and audit firm type. *Auditing: A Journal of Practice and Theory:* 97-114.

Byrne, J. 1996. And you thought CEOs were overpaid. *Business Week*, August 26th: 34.

Carcello, J.V., D. R. Hermanson, T. L. Neal, and R. A. Riley. 2002. Board characteristics and audit fees. *Contemporary Accounting Research* 19 (3): 365-384.

Charreaux, G. 1997. *Gouvernement des entreprises : théories et faits.* Economica, Paris.

Craswell, A., and S. Taylor. 1995. Auditor brand name reputation and industry specializations. *Journal of Accounting and Economics* 20: 297-322.

Daily, C.M., and D. R. Dalton. 1994. Bankruptcy and corporate governance: the impact of board composition and structure. *Academy of Management Journal* 37 (6): 1603-1617.

DeAngelo, L. 1981a. Auditor independence, low balling, and disclosure regulation. *Journal of Accounting and Economics* 3: 113-127.

DeAngelo, L. 1981b. Auditor size and auditor quality. *Journal of Accounting and Economics* 3: 183-199.

Dechow, P., R. Sloan, and A. Sweeny. 1996. Causes and consequences of earnings manipulation: an analysis of firms subject to enforcement actions by the SEC. *Contemporary Accounting Research* 13 (1): 1-36.

DeFond, M. L. 1992. The association between changes in client firm agency costs and auditor switching. *Auditing: A Journal of Practice and Theory* 11: 16-31.

Fama, E. F., and Jensen M. C. 1983a. Separation of ownership and control. *Journal of Law & Economics* 26: 301-325.

Fama, E. F., and Jensen M. C. 1983b. Agency problems and residual claims. *Journal of Law & Economics* 26: 327-349.

Firth, M., and Smith A. 1992. Selection of auditor firms by firms in the new issue market. *Applied Economics* 24: 247-255.

Francis, J. R., and E. R. Wilson. 1988. Auditor changes: a joint test of theories relating to agency costs and auditor differentiation. *The Accounting Review* 63: 663-682.

Goodwin-Stewart, J. and Kent, P. 2006. Relation between external audit fees, audit committee characteristics and internal audit. *Accounting and Finance* 46: 387-404.

Hay, D., W. R. Knechel, and H. Ling. 2008. Evidence on the impact of internal control and corporate governance on audit fees. *International Journal of Auditing* 12: 9-24.

Ireland, J., and C. Lennox. 2002. The large audit firm fee premium: A case of selectivity bias? *Journal of Accounting, Auditing and Finance.* 17 (1): 73-91.

Jensen, M. 1993. Presidential address: the modern industrial revolution, exit and the failure of internal control systems. *Journal of Finance* 48: 831-880.

Jensen, M. C., and W. H. Meckling. 1976. Theory of the firm: managerial behavior, agency costs and ownership structure. *Journal of Financial Economic* 3: 305-360.

Joshi, P. L., and H. AL-Bastaki. 2000. Determinants of audit fees: evidence from the companies listed in Bahrain. *International Journal of Auditing* 4: 129-138.

Klumpes, P., and F. Gul. 1999. The pricing of audit services, board composition and corporate control: international evidence from the life insurance industry. *Working Paper*, University of Warwick.

Knechel, W. R., and M. Willekens. 2006. The role of risk management and governance in determining audit demand. *Journal of Business Finance & Accounting* 33: 1344-1367.

Krishnan, G., and G. Visvanathan. 2009. Do auditors price audit committee's expertise? The case of accounting versus non accounting financial experts. *Journal of Accounting, Auditing and Finance :* 115-144.

Lennox, C. 2005. Management ownership and Audit Firm Size. *Contemporary Accounting Research* 22 (1) : 205-227.

Lipton, L., and J. Lorsch. 1992. A modest proposal for improved corporate governance. *The Business Lawyer* 48: 59-77.

Maati, 1999. *Le Gouvernement d'Entreprise*, Editions DeBoeck.

O'Sullivan, N. 2000. The impact of board composition and ownership on audit quality: evidence from large UK companies. *British Accounting Review* 32: 397-414.

O'Sullivan, N., and S. R. Diacon. 1994. Audit fee determination and governance structure: empirical evidence from UK insurance companies. *Geneva Papers on Risk and Insurance* 19 (70): 70-84.

Palmrose, Z. 1988. Analysis of auditor litigation and audit service quality. *Accounting Review* 63 (1): 55-73.

Piot, C. 2005. Auditor reputation and model of governance: a comparison of France, Germany and Canada. *International Journal of Auditing.* 9:21-44.

Peel, M. J., and M. A. Clatworthy. 2001. The relationship between governance structure and audit fees pre-cadbury: some empirical findings. *Corporate Governance* 9(4): 286-297.

Salleh, Z., J. Stewart, and S. Manson. 2006. The impact of board composition and ethnicity on audit quality: Evidence from Malaysian companies. *Working Paper*, Queensland University of Technology, Australia.

Simunic, D. 1980. The pricing of audit services: theory and evidence. *Journal of Accounting Research* 18: 161-190.

Shleifer, A., and R. W. Vishny. 1988. Value maximization and the acquisition process. *Journal of Economics Perspectives* 2 (1): 7-20.

Skinner, D.J., and S. Srinivasan. 2010. Audit quality and auditor reputation: Evidence from Japan. *Working paper* n°50, http://ssrn.com/abstract=1557231.

Smith, L.M. 2006. Audit committee effectiveness: did the blue ribbon committee recommendations make a difference? *International Journal Accounting, Auditing and Performance Evaluatio* 3 (2) 240-251.

Tsui, J.S.L., B. Jaggi, and F. A. Gul. 2001. CEO domination, growth opportunities and their impact on audit fees. *Journal of Accounting, Auditing and Finance:* 189-208.

Vafeas, N. 1999. Board meeting frequency and firm performance. *Journal of Financial Economics* 53: 113-142.

Watts, R., and J. Zimmerman. 1983. Agency problems, Auditing, and the theory of the firm, some evidence. *Journal of Law and Economics* 26 :613-633.

Yatim, P., P. Kent, and P. Clarkson. 2006. Governance structure, ethnicity, and audit fees of Malaysian listed firms. *Business papers*, School of Business, Bond University.

Yermack, D. 1996. Higher market valuation of companies with a small board of directors. *Journal of Financial Economics* 40: 185-211.

Zorn. 2003. Le triomphe du directeur financier : conceptions du contrôle et accession au pouvoir du directeur financier dans les entreprises américaines. *Finance Contrôle Stratégie* 6 (4) : 155-183.

Changes in Auditor Going Concern Modification Resolution

George E. Nogler, CPA, DBA, Merrimack College

(corresponding author, e-mail: noglerg@merrimack.edu)

Inwon Jang, Office of the Comptroller of the Currency

PRELIMINARY DRAFT – Please do not quote without permission.

The views expressed herein are those of the authors and do not necessarily reflect the views of the Office of the Currency, or the US Treasury Department

Changes in Auditor Going Concern Modification Resolution

Abstract: This study considers the resolution of auditor going concern modifications in the period 1997 to 2010. Hypothesis presented posit that in the period after the passage of the Private Securities Litigation Reform Act (PSLRA) of 1995, auditors provided going concern modifications only in the presence of severe financial distress. In such cases, more going concern opinion modifications would be resolved by bankruptcy than successful resolutions than in the base period observed by Nogler (1995). An additional hypothesis posits that after the events surrounding the bankruptcy of Enron and the demise of Arthur Andersen, auditors became significantly more cautious and provided going concern opinion modifications in the presence of less severe financial distress. This would be expected to be evidenced by a greater percentage of successful resolutions rather than bankruptcies than in either the base period or the PSLRA period.

Results show that, while statistically the first hypothesis cannot be accepted, the direction of the change in percentages in this period is as proposed. The second hypothesis is accepted, indicating that many more firms successfully resolve their going concern modifications than in the base period or the period after the PSLRA. This indicates that auditor decision functions viewed the relative cost of a type 2 error (no going concern modification, firm files bankruptcy) versus a type 1 error (going concern modification, file does not file bankruptcy) increased in the post Enron period.

Logistical models are presented that support the changing decision models for going concern modification resolution.

Changes in Auditor Going Concern Modification Resolution

The auditing standard governing auditor responsibility in the circumstance of client going concern issues has not changed since the issuance of Statement on Auditing Standards No. 59 (AICPA 1988) in 1988. However, substantial changes have occurred in the environment in which auditors make this judgment. In 1995, the Private Securities Litigation Reform Act provided for proportionate liability in damage awards resulting in the expectation that auditors would no longer be viewed as a "deep pockets" in cases of corporate failure (Geiger & Raghunandan 2001). In 2001-2, the failure of Enron and subsequent dissolution of its auditor, Arthur Andersen caused the auditing profession to become more conservative, that is, more likely to issue a going concern modification (Geiger et al, 2005, Nogler 2006).

The objective of this study is to test whether these changes in the environment in which auditors make going concern judgments also affected how these going concern modifications were resolved. A previous study (Nogler 1995) provides a baseline for these measurements. That study found that, of firms resolving their going concern opinions, approximately one-third subsequently received an unqualified opinion.   Of the remaining two-thirds, approximately equal numbers either filed bankruptcy or terminated operations or ceased being publicly traded in some other manner (e.g., going private, voluntary liquidation, merger or acquisition).

LITERATURE REVIEW

Resolution of Going Concern Opinions: While there is substantial research on the determinants of auditor going concern opinions and auditor behavior in this area,

there is a limited literature on the resolution of such going concern modifications. Nogler (1995) indicates that nearly equally numbers of firms receiving going concern opinions in the period 1983 to 1990 resolved their going concern opinions either by successfully receiving an unqualified opinion from their auditor (35.0%), entering bankruptcy proceedings (33.1%), or other means - merged, taken private, voluntarily liquidated, or no longer required to file with SEC - (31.9%).   This study also found that most firms which resolved their going concern opinions successfully did so by restructuring their debt or issuing stock.   The study used logistical regression.

Louwers et al (1999) modeled going concern survival using discrete time survival analysis. DTSA allows the model to retain continuing firms which were not included in Nogler's study.

Zhau (1999) retests the Nogler model, as well as other models tested in Nogler (1995) in both the post Enron and Australian settings and proposes adding new variables to increase the explanatory power in the more recent environment.

Effect of Private Securities Litigation Reform Act 0f 1995: The Private Securities Litigation Reform Act of 1995 (PSLRA) has been the subject of a number of research studies.

The SEC, in a 2000 report, asserted that changes in the legal incentives of auditors could influence auditor going concern modification issuance behavior. Specifically, they maintained that auditor protections from civil liability were causing auditors to issue going concern modifications at a slower pace. Such a conclusion is consistent with Kida (1980) who identified both risk of losing the client and risk of lawsuit

as factors auditors considered in isssuing going concern opinions to financially distressed firms under SAS 39.

Geiger & Raghunandan (2001) find that auditors were less likely to issue a going concern modification after the implementation of the PSLRA. The study considered 365 firms which filed bankruptcy. Of these firms, 246 observations were from the pre-PSLRA period and 119 in the post PSLRA period. Observation indicated that (59%) received a going concern opinion in the pre-PSLRA period, while only 45% received a going concern modification in the post PSLRA period.

Francis & Krishnan (2002) reach a similar conclusion. They provide a method to spearate auditor going concern decision behavior into two factors – changes in client characteristics and changes in auditor opinion strategy. They find that the primary determinant of granting significantly more going concern opinions is due to auditor factors rather than client characteristics.

Geiger and Raghunandan (2002) subsequently argue that the time period of PSLRA can be viewed in the broader context of certain legal decisions and act spanning the period from 1994 through 1998. They conclude that there is evidence that changes in the litigation environment of public accounting. Their conclusions indicate that auditors were more likely to issue going concern modifications in a baseline period of 1992-1993, prior to both the Central Bank of Denver case (which eliminated private litigation against auditors for abetting securities fraud), and the passage of the PSLRA, than in the period subsequent, 1996-1997. Further, they found even less likelihood of auditor going concern modifications in the period 1999-2000, after passage of the

Securities Litigation Uniform Standards Act of 1998 which eliminated certain state and common law actions against auditors.

Effect of Enron: The bankruptcy of Enron and the subsequent demise of Arthur Andersen reversed the trend in declining auditor going concern modifications. Geiger et al (2005), in a sample of 226 firms, found that in the period 2000-2001 (Pre-Enron), only 40% of firms filing bankruptcy had received a going concern opinion in the year prior to filing. By contrast, in the period 2002-2003, 70% of the firms filing bankruptcy received prior going concern opinions.

Nogler (2006) in a descriptive study involving 1,237 firms filing bankruptcy in the period November 1994 to June 2005 found evidence for both the increased likelihood that a financially distressed firm would not receive a going concern opinion after the passage of the PSLRA and that the likelihood increased significantly in the post-Enron period. Specifically, he finds that in the pre-PSLRA period the rate of going concern opinions preceding bankruptcy was 50.8%. After PSLRA, this rate drops to 44.3%. In the post-Enron period, the rate rises to 62.9%. These results, while not as dramatic, are consistent with Geiger et al (2005).

HYPOTHESES

Under the PSLRA of 1995, researchers found that auditors were less likely to issue going concern modifications (Geiger & Raghunandan 2001 Francis & Krishnan 2002). This study therefore posits that auditors were issuing going concern modifications only in the presence of more severe client financial distress, i.e. a greater likelihood of impending bankruptcy. In that circumstance, the incidence of bankruptcy

among firms receiving a going concern modification in this period would be significantly

greater than in the period reported by the baseline study (Nogler 1995) and also in the

period after Enron.

H_1 Among firms resolving their going concern modifications, the rate of bankruptcy in
the sample of opinions resolved during the period after the passage of the PSLRA and
prior to the Enron effect will be significantly greater, and consequently the rate of
successful resolution will be less, than either the baseline period or the post-Enron
period. This would suggest that auditors viewed the relative cost a Type 1 (going
concern modification, firm does not fail) versus Type 2 error (no going concern
modification, firm files bankruptcy) differently after the passage of the PSLRA, placing
relatively less weight on type 2 errors in their decision process.

In the period after Enron, researchers have found substantial support for the

concept that auditors became substantially more conservative (Geiger et al 2002,

Nogler 2006, 2008). This study therefore posits that auditors were issuing going

concern modifications in the presence of less severe client financial distress. In that

circumstance, the incidence of bankruptcy among firms receiving a going concern

modification in this period would be significantly less than in the period reported by the

baseline study (Nogler 1995) and also in the period after Enron.

H_2 Among firms resolving their going concern modifications, the rate of bankruptcy in
the sample of opinions resolved after Enron will be significantly less, and the rate of
successful resolution consequently greater, than either the baseline period or the
PSLRA period prior to Enron. This result would suggest that auditors placed
significantly more weight on type 2 errors after the failure of Enron and the demise of
Arthur Andersen.

DATA AND VARIABLE SELECTION:

Data Selection: Firms receiving going concern opinions in the period 1997 to

2009 were identified through MergentOnline using the search terms "substantial doubt"

and "going concern." The sample was limited to Standard Industrial Classification

Codes 2000-3999 (manufacturing) and 5000-5999 (wholesale/retail). The decision to

limit the sample to these firms was based on the fact that the environment of other firms

is intrinsically different. Nogler (2008) shows that the largest 18 bankruptcies in the

period 1997-2005, which amounted to nearly 50% of the total assets entering

bankruptcy, were outliers in terms of size. He argues these specific cases should be

treated separately, perhaps as case studies. Of these 18 firms, occurred primarily in

other sectors; 5 in telecommunications (SIC 4), 3 were airlines (SIC 4), 4 were energy

companies (SIC 1), and 3 were in financial services. This resulted in identifying 2,672

firms receiving going concern opinions in this period.   Consistent with Nogler (1995) we

did not include development stage companies in our sample. Nearly 1,000 firms were

excluded because they were development stage firms. The Securities and Exchange

Commission defines a development stage firm as (Regulation SX-Rule 1-02):

"A company shall be considered to be in the development stage if it is
devoting substantially all of its efforts to establishing a new business and
either of the following conditions exists: (1) Planned principal operations
have not commenced. (2) Planned principal operations have commenced,
but there has been no significant revenue therefrom."

The percentage of firms excluded for this reason is considerably greater than Nogler

(1995). That study relied upon the NAARS database which excludes many firms at this

stage. Additionally, nearly 600 firms were excluded as they either received an initial

going concern opinion outside the period of this study or they received a going concern

opinion on their initial public filing.

　　　After exclusions, a total of 483 firms which had resolved their going concern

opinions were analyzed. Table 1 describes the derivation of this sample in detail.

[Insert Table 1 about here.]

Variable Selection: Generally, variables selected were based on the model presented in Nogler (1995).   Variables retained from that model included:

Change in Total Liabilities/ Total Assets (CHSTLTA) – intended to measure solvency. The expected sign of this variable is negative as a relative increase in liability position would signal a weakening of the firms financial position.

Default (DEFAULT) – a dichotomous variable coded 1 if the firm had experienced an event of default, whether or not waived (consistent with Chen & Church 1992), and coded as 0 otherwise. The expected sign of this variable is negative as firms in default are unlikely to have their going concern opinions removed by the auditor.

Restructure (RESTRUCTURE) – a dichomotous variable coded as 1 for firms which restructured their debt, obtained new credit facilities, or issued stock, coded as 0 otherwise.   The expected sign of this variable is positive as any of these changes represent a perception that the firm is a viable entity.

Prior Going Concern Opinions (PRIORGC) – Prior research has indicated that the number of consecutive prior going concern opinions a firm has received may be a significant factor in the auditor's opinion decision. The sign of this variable is problematic. Mutchler (1984) suggest the sign should be negative as auditors may find it easier to repeat a going concern modification rather than initiate that modification. In a successful going concern resolution model, Nogler (1995) finds this variable to be positive and not significant. This result may be due to the lack of a size variable in Nogler's model.

One variable was changed from the Nogler model and four variables were added, as follows:

Change in Operating Cash Flow (CHSOPCF): The Nogler study (1995) included a variable change in operating income before depreciation to total assets. The period of this study (1986-1990) predated FASB #95 – Statement of Cash Flows which became effective for fiscal years ending after July 15, 1988. Since the cash flow statement provides a measure of non-cash income from operations (Cash from operations), this study uses the change in that number to total assets instead of the change in operating income before depreciation to total assets used in the prior study.

Risky (RISKY):   This is a dichotomous variable coded 1 for industries in SIC codes 2833-2836 (medical compounds), 3570-3577 (high tech equipment manufacturing), and 3600-3674 (electrical and telecommunications manufacturing) which are identified in Krasznik and Lev (1995) as high risk industries, coded 0 otherwise. The expected sign of this variable is negative.

Log of Total Assets (LOGTA): Size maybe a significant factor in a firm's ability to convince the auditor that is a viable enterprise. It may also be easier for larger firms to restructure debt or issue stock to resolve a going concern modification. The expected sign of this variable is positive.

 Net Income to Total Assets (NITA): This variable has intuitive appeal as it would be expected that firms resolving their going concern status might be expected to have turned profitable. Also, this variable was found significant in Nogler's use of the White

(1984) and Casey et al (1986) models which related to bankruptcy resolution. The expected sign would be positive.

Operating Cash Flow to Net Sales: This variable similarly has an intuitive appeal in that positive cash flow would be expected to be associated with improvement in operations.   The expected sign of this variable is positive.

MODEL SPECIFICATION

This study will first consider whether the relative proportions of firms resolving their going concern modifications whether through receiving an unmodified opinion, filing bankruptcy, or other (voluntary liquidation, merger, going private, or no longer filing with the SEC under rule 12(g) changed from the Nogler baseline data (1986-1990), the post PSLRA/pre-Enron period (1997-2001), and the post Enron period. This will be done using a test of proportions as follows for each period compared to all each other period. This will utilize a Fisher's exact test to obtain z values as follows:

$$z = \frac{\overline{}}{\overline{}}$$

Where $ - $.

This method will be used to test the hypotheses presented earlier in this study.

The study will then model the process of successful resolution (i.e. subsequent unmodified opinion) utilizing the variables cited above for the three periods described in the preceding paragraph. Comparison of the models will indicate if the variables or the

weights assigned the variables changed across the period under study which ranges from 1997 to 2009. The format of the proposed logistical models is:

The models will be viewed over the three periods to see if there are discernible patterns in the significance and weight of the independent variables.

RESULTS

The breakdown of the 483 firms in this study by resolution and time period is presented in Table 2. The results of the Nogler (1995) study are presented as a baseline for the period preceding the PSLRA. The PSLRA period of observations extends from January 1997 to October 31, 2001. The post-Enron period is dated, consistent with Geiger et al (2005) uses October 16, 2001 (the date Enron acknowledged accounting errors) as the beginning of the post-Enron period, and Nogler (2008) who cites December 31, 2001 as the beginning of the post Enron period. The post-Enron period continues to the end of the observation period December 31, 2009.

[Insert Table 2 here.]

Of specific interest, based on the hypotheses presented earlier, is the relative proportion of bankruptcies to successful resolutions in each time frame. In the baseline period, these are nearly equal with 33.12% bankruptcies and 35.03% successful resolutions. In the period after PSLRA, bankruptcies represent an increased portion of

resolution at 39.41%, while successful resolutions drop to 29.41%. The direction here is consistent with the hypothesis that firms receiving a going concern modification after the PSLRA were more financially distressed than previously. The Fisher's Exact Test, however, is unable to reject the hypothesis that these observations were drawn from different populations.

Bankruptcies in the post-Enron period dropped to 20.52% while successful resolutions rose to 44.54% of all resolutions. This is consistent with the second hypothesis which suggests that auditors became much more cautious in the post-Enron period. Comparing the base period to the post-Enron period, the Fisher Exact Test rejects the hypothesis that these observations were drawn from the same population. Comparing the PSLRA period to the post-Enron period, the Fisher Exact Test also rejects the hypothesis that these observations were drawn from the same population. This finding provides evidence that auditor behavior regarding the issuance of going concern modifications changed significantly in the post-Enron period, with the cost of a type 2 error (no going concern modification, firm files bankruptcy) having greater weight than in the previous environment.

Crossover firms are firms which received an initial going concern modification in the PSLRA period but resolved it in the post-Enron period. Interestingly, the proportion of resolutions in this group most closely resembles the post-Enron period.

While the hypotheses presented earlier dealt with the relative proportions of bankruptcies versus successful resolutions in the respective time periods, Table 2A presents descriptive details of the other resolutions (non-bankruptcy liquidation, firms no

longer filing with the SEC under Rule 12(g), merged or acquired firms, and firms taken private).

Using the year of the initial going concern opinion as the first observation and the year preceding bankruptcy or cessation of filing as the last observation (for firms which are categorized as bankrupt or other) and the year of resolution for resolving firms, produces a total of 1,067 observations.

Logistical models were created for the three periods, utilizing the variables described above, on these observations to describe the characteristics of the successful resolution process in each period. The results of these models are presented in Table 3.

[Insert Table 3 about here.]

In the period following the passage of the PSLRA, only four variables are significant at the 5% level in the model: Default, Restructure, the number of Prior Going Concerns Received, and Net Income to Total Assets. The model itself is significant at the 1% level and overall classification accuracy is 92.86% based on 280 observations. The model correctly classifies 82.2% of the 50 successful resolutions in this period.

In the post-Enron period, all of the variables with the exception of Risky and Net Income to Total Assets are significant. It is interesting to note that the variable Net Income to Total Assets was significant in the period preceding. The model itself is significant at the 1% level and classification accuracy is 91.62% based on 501 observations. The model correctly classifies 82.6% of the 102 successful resolutions in this period.

In the post-Enron period, all of the variables with the exception of Risky and Cash Flow from Operations to Net Assets are significant.   It is interesting to note that the variable Net Income to Total Assets is once again significant in this period.   The model itself is significant at the 1% level and classification accuracy is 91.96% based on 286 observations. The model correctly classifies 76.7% of the 40 successful resolutions in this period.

CONCLUSIONS

The two hypotheses presented earlier have somewhat mixed results. The first hypothesis posited that in the period when the Private Securities Litigation Reform Act was in force, and preceding the events surrounding the bankruptcy of Enron, that auditors would only issue a going concern modification in the presence of severe financial distress. This could be observed by a greater percentage of resolutions by bankruptcy than by receiving a subsequent un-modified opinion. The statistical results of this hypothesis lead to the rejection of the hypothesis at the 5% significance level. However, while the relationship is not statistically strong, the direction of observed is consistent with the hypothesis. In the base period, the rate was 33.12% bankruptcies and 35.03% successful resolutions, in the period after PSLRA, bankruptcies represent an increased portion of resolution at 39.41%, while successful resolutions drop to 29.41%.

The results of the second hypothesis are more definitive. That hypothesis posited that after Enron auditors would issue going concern modifications in the presence of less severe financial distress. This could be observed by a greater

percentage of successful resolutions compared to resolutions by bankruptcy. Statistical tests support the acceptance of this hypothesis.

Logistical Models for the three periods 1997-2001 (PSLRA before Enron), 2002-2010 (post Enron), and for observations in the period overlapping this period (received initial going concern opinion under PSLRA but resolved post Enron) suggest that the auditor decision models changed over this time period. Three variables, Default, Restructure, and Prior Going Concerns are significant across the three time periods. In the PSLRA before Enron period, Net Income to Total Assets is also significant and positive, suggesting that a turnaround in the profitability of the client is a significant factor.

In the post Enron period, in addition to the three consistent variables, the Change in Total Liabilities to Total Assets, Change in Operating Cash Flow, Size (Log of Total Assets), and Cash Flow to Net Sales become significant. This suggests auditors considered a more complex set of financial dimensions of the client in deciding to remove the going concern modification. The fact that Net Income to Total Assets is not significant in this model, but Cash Flow to Net Sales becomes significant may suggest that auditors were more interested in operating cash flow as a measure than net income in this period.

The model for the crossover period more closely resembles the post-Enron model, which is not surprising. Interestingly, for these firms, the variable Net Income to Total Assets once again is significant while the measure of operating Cash Flow to Net Sales is not significant in this model.

TABLE 1   Sample Selection Data

Firms initially Identified as Going Concern Opinions	**2,672**	**100.00%**
Going concern did not refer to this company	**129**	**4.83%**
Development stage firms	**989**	**37.01%**
Initial Going Concern opinion prior to 1996	**210**	**7.86%**
Going Concern Opinion on initial public filing	**388**	**14.52%**
Firm already bankrupt	**90**	**3.37%**
Continuing - still receiving Going Concern opinion	**145**	**5.43%**
Firm Changed Fiscal Year end	**48**	**1.80%**
Incomplete data, unaudited, etc.	**56**	**2.10%**
	617	**23.09%**
Not on Compustat	**134**	**5.01%**
Final Sample	**483**	**18.08%**

TABLE 2

Resolutions by Time period

	Baseline Nogler 1995	Pre-Enron	Post-Enron	Crossover	Total
Resolved	55	50	102	40	192
Bankrupt	52	67	47	22	136
Other	<u>50</u>	<u>53</u>	<u>80</u>	<u>22</u>	<u>155</u>
	157	170	229	84	483
Resolved	35.03%	29.41%	44.54%	47.62%	39.75%
Bankrupt	33.12%	39.41%	20.52%	26.19%	28.16%
Other	<u>31.85%</u>	<u>31.18%</u>	<u>34.93%</u>	<u>26.19%</u>	<u>32.09%</u>
	100.00%	100.00%	100.00%	100.00%	100.00%

TABLE 2 A

Breakdown of Other Resolutions by period

Periods	Pre-Enron	Post-Enron	Crossover	Total
Non-bankruptcy Liquidation	9	9	2	20
No Longer Files with SEC	19	37	12	68
Merged or Acquired	19	22	5	46
Taken Private	6	12	3	<u>21</u>
	53	80	22	155
Non-bankruptcy Liquidation	5.81%	5.81%	1.29%	12.90%
No Longer Files with SEC	12.26%	23.87%	7.74%	43.87%
Merged or Acquired	12.26%	14.19%	3.23%	29.68%
Taken Private	<u>3.87%</u>	<u>7.74%</u>	<u>1.94%</u>	<u>13.55%</u>
	34.19%	51.61%	14.19%	100.00%

TABLE 3 Coefficients, Chi Squares, and Probabilities

PSLRA Period(January1997 to October 31, 2001

Intercept/ Variables	Coefficient	Chi-Square Beta	Probability
Intercept	-2.602	8.26	.004*
	-0.005	0.01	.914
	0.499	0.46	.500
	1.200	4.10	.043*
	3.447	31.79	.000*
	.916	2.46	.117
	1.524	15.84	.000*
	-.004	0.07	.790
	1.719	8.60	.003*
	0.0002	0.09	.770
Pseudo R2 Model Chi-Square Probability Classification Accuracy			.364 154.620 .000* 92.86%

Post-Enron Period (November 1, 2001 to 2010)

Intercept/ Variables	Coefficient	Chi-Square Beta	Probability
Intercept	-3.836	53.34	.000*
	0.193	14.34	.000*
	1.380	14.49	.000*
	-3.236	15.76	.000*
	3.513	88.28	.000*
	-0.667	3.20	.074
	0.510	14.42	.000*
	0.454	14.06	.000*
	-0.0007	0.05	.830
	-002	5.47	.019*
Pseudo R2 Model Chi-Square Probability Classification Accuracy			.355 270.81 .000* 91.62%

Crossover Firms (January 1997 to 2010)

Intercept/ Variables	Coefficient	Chi-Square Beta	Probability
Intercept	-4.034	27.59	.000*
	-0.287	6.49	.011*
	-0.490	6.85	.009*
	-1.671	8.44	.004*
	3.237	39.24	.000*
	-0.088	0.03	.872
	0.369	7.62	.006*
	0.357	4.82	.028*
	0.416	8.87	.003*
	-.0004	0.14	.711
Pseudo R2 Model Chi-Square Probability Classification Accuracy			.276 105.12 .000* 91.96%

REFERENCES

American Institute of Certified Public Accountants. 1988. *Statement on Auditing Standards No. 59, The Auditor's Consideration of an Entity's Ability to Continue as a Going Concern.* New York: AICPA.

Financial Accounting Standards Board. 1987. *Statement of Financial Accounting Standards No. 95 – Statement of Cash Flows.*

Francis, J. R., and J Krishnan. 2002. Evidence on auditor risk-management strategies before and after the Private Securities Litigation Reform Act of 1995. *Asia Pacific Journal of Accounting and Economics.* 135-157.

Geiger, M. A., and K. Raghunandan. 2001. Bankruptcies audit reports and the Reform Act. *AUDITING: A Journal of Practice and Theory* (March). 187-196.

Geiger, M. A., and K. Raghunandan. 2002. Audit reports and the new litigation environment. *Accounting Horizons.* (January). 17-26.

Geiger, M. A., K. Raghunandan, and D. V. Rama. 2005. Recent Changes in the association between Bankruptcies and prior audit opinions. *AUDITING: A Journal of Practice and Theory* (May). 21-35.

Kasznik, R., and B. Lev. 1995. To warn or not to warn: management disclosures in the face of an earnings surprise. *The Accounting Review* 70 (1) (January): 113–134.

Kida, T. 1980. An investigation into auditor's continuity and related qualification judgments (Autumn). *Journal of Accounting Research.* 506-522.

Louwers, T. J., F. M. Messina, and M. D. Richard 1999. The auditor's going-concern disclosure as a self-fulfilling prophecy: a discrete-time survival analysis. *Decision Sciences* (June). 805-824.

Mutchler, J. 1984. Auditors' perceptions of the going-concern opinion. *AUDITING: A Journal of Practice and Theory* (Spring). 17-30.

Nogler, G. E. 1995. The resolution of auditor going concern opinions. *AUDITING: A Journal of Practice and Theory* (Fall). 54-73.

Nogler, G. E. 2006. The changing information content of auditor going concern opinions. Commercial Lending Review (January-February). 43-47.

Nogler, G.E. 2008. Going concern modifications, CPA firm size, and the Enron effect. *Managerial Auditing Journal.* 51-67.

Private Securities Litigation Reform Act, P.L. 104-67.1995. Washington D.C.: Government Printing Office.

Securities and Exchange Commission. 2000. *Revision of the Commission's Auditor Independence Requirements*. Release Nos. 33-7919; 34-43602; File No. S7-13-00. Washington D.C. SEC.

Securities and Exchange Commission. Regulation S-X, Rule 1-02. 17 § 210.1-02   Definitions of terms used in Regulation S-X. "Development Stage Company."

Zhau, S. 2009. The successful resolution of auditor going concern opinions: evidence from the U.S. and Australia. Working paper (November).

LIKELIHOOD OF U.S. INDIVIDUAL FILERS CHOOSING AN OFF-SHORE TAX PREPARER: AN EMPIRICAL STUDY

Deb Sledgianowski, Department of Accounting, Taxation, and Legal Studies,
Hofstra University, Hempstead, NY, (516) 463-4759, Deb.Sledgianowski@Hofstra.edu
Steven T. Petra, Department of Accounting, Taxation, and Legal Studies,
Hofstra University, Hempstead, NY, (516) 463-4178, Steven.T.Petra@Hofstra.edu
Suzanne P. Basilicato, Department of Accounting, Taxation, and Legal Studies, Hofstra
University, Hempstead, NY, (516) 463-5684, Suzanne.Basilicato@Hofstra.edu

ABSTRACT

This research-in-progress investigates the likelihood of individual tax filers choosing to outsource their income tax preparation to an offshore service provider. An internet panel of 254 respondents across the U.S. was used to assess the likelihood of their choosing an off-shore tax preparer over an American tax preparer as determined by 1) cost savings to prepare tax return, 2) time to complete tax return, and 3) quality of completed tax return. Initial univariate data analysis found that the U.S. respondents are more likely to choose an American tax preparer if there is no difference in cost to them, they are more likely to choose an American tax preparer if there is no difference in time to complete the return, and they are more likely to choose an American tax preparer if there is no difference in the quality of the completed tax return.

Keywords: Offshoring, outsourcing, tax preparation, service management

LITERATURE REVIEW

Offshore outsourcing of tax preparation occurs when a firm engaged by a client to prepare a tax return procures these services from a third-party located in a foreign country. Skills that are most likely to be off-shored are those that do not require close interaction between the client and offshore service supplier (Bullen et al., 2007). The tax preparation process is ideal for off-shoring since it does not typically involve interacting with the client after their tax-related documents are collected. The tax preparation process involves the tax preparation service provider scanning all client paper documents (such as W-2s, 1099s, and prior year tax return) into a digital format that is uploaded to the service providers network. Offshored tax preparers (usually chartered accountants or other equivalently trained workers) access the documents from the service providers secure network via a Web browser and download them to their local computer and prepare the tax return using the service provider's tax preparation software. The service provider electronically retrieves the completed return, reviews it and makes any necessary corrections, and prints the completed return to give to their client (e.g. see www.xpitax.com and www.sureprep.com).

Internal Revenue Service (IRS) statistics indicate that of the approximately 150 million individual U.S. tax returns filed in 2007, more than half were filed using paid preparers (IRS). Over the last several years CPA firms have begun shifting some of this tax preparation work overseas using outsourcing facilitators. Two such outsourcing facilitators are U.S.-based Xpitax

and Sureprep. These outsourcing facilitators use Chartered Accountants in India who are trained in U.S. tax laws. Bandyopadhyay and Hall (2009) surveyed 35 U.S. accounting firms and found that 13 (37%) utilize off-shoring for individual income tax preparation; of these firms, almost 20% of their returns were prepared off-shore. They also found that 38% of the responding firms reported that cost effectiveness was a factor in their decision to employ off-shoring and 46% employed off-shoring because they wanted to provide faster service to the client.

A review of the outsourcing literature relative to service-providing indicates that the service providing process is influenced by cultural and geographic differences between the service provider and the client seeking the service. Warden et al. (20003) found that the greater the cultural distance, the wider the gap in service expectations. Positive customer experiences with outsourcing increases when there is cultural similarity between the provider of the service and consumer of the service, but the end recipient of the service does not usually have a say in choosing the actual provider of the service (Hopkins et al., 2005). Thelen et al (2007) suggest that the offshore service is affected by lack of face-to-face contact and sensitivity of information exchanged.

Brody et al (2006) found that 72% of individuals who paid to have their return prepared indicated that it would make a difference to them if their accountants were to outsource their income tax return and 88% said it would make a difference if their return were off-shored (e.g. to India). Since 2005 the American Institute of Certified Public Accountants (AICPA) has required its members to inform their clients if they are using third-party service providers when providing them with professional services. As of 2009, the IRS has required paid preparers to obtain written consent from their clients before disclosing any of their tax information to another prepare located outside of the U.S. (Jamouneau, 2009). In 2010 the Treasury Department issued regulations under Code Sec. 6109 intended to provide some assurance to taxpayers that a tax return was prepared by an individual who is knowledgeable, skilled, and ethical and to improve the accuracy of tax returns and increase tax compliance. These regulations became effective on September 30, 2010. Unenrolled tax preparers are now required to become Registered Tax Return Preparers (RTRP) by registering for a permanent tax identification number (PTIN). This includes unregistered foreign preparers who do not have a social security number and may not be able to obtain one. Unenrolled tax preparers must pass a competency test in wage and non-business Form 1040 and wage and small business Form 1040 Schedules, C, E, and F. The IRS has proposed that RTRPs be required to complete annual continuing education credits in federal tax law updates, federal tax law, and ethics.

Our current research-in-progress investigates the likelihood of individual tax filers to outsource their tax preparation to an off-shore service provider.

PURPOSE OF THIS STUDY

This study investigates the intention of U.S. individual filers to off-shore their income tax preparation. We are interested in determining whether Bandyopadhyay and Hall's (2009) aforementioned findings from accounting firms regarding cost effectiveness and faster service to the client were also important to individual filers in determining their likelihood in choosing an

American or overseas tax preparer. Specifically, we are investigating intent to offshore tax return preparation based on the following factors:

1. Period of time necessary for an overseas service provider to prepare the return versus the time required by an American service provider. Thus,
H1: The less perceived time it takes to prepare a tax return off-shore compared to on-shore, the greater the likelihood of individual tax filers to outsource their tax preparation to an off-shore service provider.

2. Quality of the work done by an overseas service provider as compared to an American service provider. Thus,
H2: The better the perceived quality of work performed by an off-shore service provider compared to an American service provider, the greater the likelihood of individual tax filers to outsource their tax preparation to an off-shore service provider.

3. Cost of having the return completed by an overseas service provider as compared to an American service provider. Thus,
H3: The less expensive it is to prepare a tax return off-shore compared to on-shore, the greater the likelihood of individual tax filers to outsource their tax preparation to an off-shore service provider.

METHODOLOGY

The data for this study was collected as part of a larger study examining consumer opinion about off-shoring of services. A national (U.S.) internet panel was conducted resulting in a sample of two-hundred and fifty-four respondents. Internet panels use a pre-recruited group of individuals who have agreed to participate in online market research studies on a forward-going basis in return for compensation. Internet panels have been found to positively address weaknesses associated with other forms of web surveys (James, 2000), are more reliable than telephone-based surveys (Braunsberger 2007), and overcome the respondent issue of social desirability bias that can arise in contact surveys (Dennis 2005; Duffy 2005). Internet panel members were invited to complete the survey on-line. In total, 4,532 invitations were sent out with 299 respondents starting and 254 finishing the survey for a completion response rate of 5.6%. In order to achieve a sample representative of the general population invitations were sent out to a representative sample in regard to age, gender, income, education, employment and geographic location. Survey requests were sent out in waves and if it was noticed that a greater percentage of a particular group (e.g., one geographic location) were responding then less invitations were sent to that group in the next wave. These procedures provided a balanced sample. Table 1 details the sample characteristics.

Table 1: Demographic Profile of Respondents					
Gender		**Education**		**Employment Status**	
Male	130	High school or less	68	Full-time	118
Female	119	Some college	129	Homemaker	18
No response	6	College	39	Disabled	17
		Graduate school	17	Part-time	35
Age		No response	1	Retired	56
18-34	62			Unemployed	8
35-54	96			No answer	2
55-74	74	**Income**			
75+	14	<$20k	41	**Geographic Location**	
		>$20k<$40k	59	Northeast U.S.	57
Union Status		>$40k<$60k	45	Southern U.S.	83
Union	40	>$60k<$80k	44	Midwest U.S.	65
Non-Union	208	>$80k<$100k	28	Western U.S.	44
No response	6	$100k or more	34		

Attitudes toward off-shoring various services (*tax preparation, customer service and technical help with products, reading of X-rays*) were assessed by the survey instrument. These services were selected based on a separate pilot study of 72 undergraduate and graduate students asking them to evaluate degree of product involvement required by the offshore provider and the perceived risk to the consumer; out of the three off-shored services, tax preparation services was deemed by the pilot study sample to have the highest degree of product involvement and the highest degree of perceived risk.

RESULTS

This data will be analyzed using analysis of variance to determine if differences exist between groups based on cost of having the return prepared, period of time to prepare return, and quality of the work. Initial univariate data analysis was performed with the following results:

Cost as a Determinant
Survey respondents were asked how likely they were to choose an American service provider versus an Overseas service provider based on cost.

Preparation Time as a Determinant
Survey respondents were asked how likely they were to choose an American service provider versus an Overseas service provider based on time to prepare the tax return.

Quality as a Determinant
Survey respondents were asked how likely they were to choose an American service provider versus an Overseas service provider to complete their tax return based on the quality of work performed.

Initial univariate data analysis found that the U.S. respondents are more likely to choose an American tax preparer if there is no difference in cost to them, they are more likely to choose an American tax preparer if there is no difference in time to complete the return, and they are more likely to choose an American tax preparer if there is no difference in the quality of the completed tax return.

ACKNOWLEDGEMENTS

REFERENCES

Bandyopadhyay, J. & Hall, L. (2009). "Off-shoring of tax preparation services by U.S. accounting firms: an empirical study". Advances in Competitiveness Research, 17(1&2), pp. 71-89.

Braunsberger, K., Wybenga H., and Gates R. 2007. "A Comparison of Reliability Between Telephone and Web-Based Surveys." *Journal of Business Research,* 60 (July), 758-764.

Brody, R. G., Coulter, J. M., and Jewell, J. (2006). Losing the public's trust: Third-party service providers and disclosure. The CPA Journal*,* 76(9), 66-70.

Bullen, C.V., Abraham, T., Gallagher, K., Kaiser, K.M. and Simon, J. (2007). "Changing IT skills: the impact of sourcing strategies on in-house capability requirements", Journal of Electronic Commerce in Organizations, 5(2), pp. 24-46.

Dennis, J. M., (2005), "Data Collection Mode Effects Controlling for Sample Origins in a Panel Survey: Telephone versus Internet." (Draft Copy), [available at http://scholar.google.com/scholar?hl=en&lr=&safe=off&q=cache:pu36blZjHyoJ:www.intersurey.com/ganp/papers/research0105.pdf+%22internet+panel%22+and+marketing+research].

Duffy, B. (2005), "Comparing Data from Online and Face-to-Face Surveys." *International Journal of Market Research,* 47 (6), 615-639.

Hopkins, S. A., Hopkins, W. E., Hoffman, K. D. (2005) "Domestic inter-cultural service encounters: an integrated model", Managing Service Quality, 15(4), pp. 329-343.

Internal Revenue Service (2007). "Tax stats at a glance", retrieved from http://www.irs.gov/taxstats/article/0,,id=102886,00.html on April 1, 2010.

James, D. 2000. "The Future of Online Research." *Marketing News* January 3, 1.

Jamouneau, G. (2009). "New regs govern overseas disclosure and use of taxpayer information", Journal of Accountancy.

Thelen, S. T., Thelen, T.K., Magnini, V.P., Honeycutt, E. D. (2009)." An introduction to the offshore service ethnocentrism construct", *Services Marketing Quarterly,* Vol. 30(1), pp. 1-17.

Warden, C. A., Huang, C.-T., Lee, C.-H., Liu, T.-C. (2003) "Service failures away from home: benefits in intercultural service encounters", *International Journal of Service Industry Management,* Vol. 14(4), pp. 436-457.

THE IMPACT OF BRAND EXPERIENCE ON CUSTOMER EQUITY

Arthur Cheng-hsui Chen, National Yunlin University of Science and Technology, Taiwan,
chencs@yuntech.edu.tw
Chien-Lin Ma, National Yunlin University of Science and Technology, Taiwan,
g9722702@yuntech.edu.tw

ABSTRACT

The objective of this research is to explore the relationship between brand experience and customer equity (value equity, brand equity, and relationship equity), and then examine the impact of different contact points experience (communication contact, physical environment contact, personnel contact and product usage contact), and different brand experience dimensions on customer equity. Further, the different brand positioning strategies-hedonic and utilitarian is also being used to examine its moderating effect on the above relationships. Total four hundreds ninety three valid samples, including 410 with brand experience and 83 without brand experience, were collected via survey website. The results indicate that the positive and strong brand experience is a key factor for building strong customer equity. Specifically, brand equity is mostly influenced by brand experience, followed by relationship equity and value equity. Although the four contact points' brand experiences all have significant impact on customer equity, product usage contact is the most important role on influencing customer equity and its individual drivers. The different brand position strategies have moderating effect. For utilitarian brand, only brand experience at product usage contact point has significant impact on customer equity and its three drivers. For hedonic brand, the four contact points' experiences do have significant relationship with customer equity.

Key words: brand experience, customer equity, brand experience contact point, hedonic and utilitarian

INTRODUCTION

Due to the technology advancement, the differentiation of product functions gets smaller and smaller, but the closely following service economy is also facing challenge owing to being easily imitated. It makes the contemporary marketing focus transfer from the judgment or perception before purchasing to the practical experience after purchasing. Therefore, now it is the so called "Experience Economy Era". Experiential marketing has become one of the key efficient instruments for a brand to build competitive advantages. For instance, Starbucks is not just selling coffee. The retailing atmosphere and interaction with its service provider are always an enjoyable experience. The key satisfaction factor of Apple iPhone and iPad is to experience its innovative technology. As a result, the future marketing will primarily focus on the trend of satisfying experience for the customer. And the challenge for marketer is how to change from service economy thoughts to experience economy thoughts. This points out how experiential marketing plays a vital role in marketing in the twenty-first century.

Based on literature review, many researches have paid much attention to the experience marketing in the past ten years (Morrison and Crane 2007；Pine and Gilmore 1998；Schmitt 1999). However, most of them put emphasis on elaborating experience concepts and characteristics (Phillips, Olson and Baumgartner 1995；Pine and Gilmore 1998；Schmitt 1999；Jensen 1999）or experience classification

(Pine and Gilmore 1998；Schmitt 1999 or experience measurement（Plutchik and Kellerman 1980；Russell and Mehrabian 1975；Brakus, Schmitt and Zarantonello 2009）. Document about the relationship between experiential marketing and management implication is still lack of attraction, not mention to the topic of brand experience (Brakus, Schmitt and Zarantonello 2009). Therefore, for the purpose of filling this research gap, it triggers the motivation to explore the topic of brand experience.

We realize the importance of brand experience, but how does it create the benefits for a brand? In terms of the concept of brand experience, its focus is on consumers' sense, feeling, thinking, and behavior. Customer equity, succeeding customer-based brand equity, is also one of the contemporary marketing streams focusing on consumer perspective strengthened by the academic research and marketing practice (Rust, Zeithaml, and Lemon 2000). Customer equity mainly figures out how to build the biggest value for a brand from consumer lifetime value. This demonstrates both customer equity and brand experience are on the same track of management philosophy. Specifically, higher customer equity is a long term objective a brand tries to achieve. Therefore, the question is will brand experience affects customer equity? In other words, will customer equity become a good consequence indicator for brand experience, particular for negative or no experience? This is the first main issue in this research.

Second, several scholar pointed out that experience takes place in searching for products, in the process of choosing products in retail stores, in contact with servers, and in using products (Arnould, Price and Zinkhan 2002；Brakus, Schmitt and Zhang 2008；Holbrook 2000). However, before consumers haven't formed actual purchasing motivation, it's possible for them to contact with brand communication media. In fact, this is also a part of brand experience for consumers. This research collectively calls these as brand experience contact points, where the experience occurs. Brand experience contact points include communication contact, physical environment contact, personnel contact, and product usage contact. Apparently these experiences in different contact points play different roles in marketing. And for these brand contact experiences, do they have different impact on customer equity?

Brakus, Schmitt and Zarantonello (2009) proposed that there are four dimensions in brand experience, including sensory, affective, intellectual, and behavioral experience. The four brand experience dimensions primarily come from the reactions to different brand related stimulation. On the other hand, brand marketing offer consumers with different stimulation in different experience contact points, which may cause different experience among these four dimensions. Therefore the third research question is to explore if the four experience dimensions have different influences on customer equity in each experience contact point?

Consumer choices are driven by utilitarian and hedonic considerations, which may happen on attribute level (Batra and Ahtola 1990), product level (Dhar and Wetenbroch 2000), or shopping value (Babin, Darden and Griffin 1994). Research suggests that these different considerations map onto independent components of product evaluations and attitudes and enable people to distinguish between goods according to their relative hedonic or utilitarian nature (Batra and Ahtola 1990; Mano and Oliver 1993). Voss, Spangenberg, and Grohman (2003) developed ten items with reliability and validity to measure the hedonic and utilitarian dimensions of consumer attitudes toward product categories and different brands within categories. This indicates that marketer can also differentiate its brands as utilitarian or hedonic positioning strategy, which offers different experience value for customer. Therefore, the last research question we would like to examine is if the relationship between brand experience and customer equity will be moderated by the different brand positioning strategies?

To sum up, the objective of this research is to explore the relationship between brand experience and customer equity, and then examine the impact of different contact points experience, and different brand experience dimensions on customer equity. Further, the moderating effect by the different brand positioning strategies will also be examined on the above relationships. We look forward to offering some contribution to the field of brand experience marketing and customer equity for academicians and practitioners.

LITERATURE REVIEW

Experiential Marketing

Norris（1941）is the first scholar that proposed consumption experience. He emphasized that the importance of consumption experience is product service rather than product itself. And the following is that the marketer and consumer researcher are aware of the significance of hedonic consumption and consumer experience. Abbott（1955） thought that what people really want is not product itself, but is satisfying experience. Pine and Gilmore（1998）pointed out that product is tangible, while service is intangible. And the product and service experiences created for consumers are unforgettable. Besides, consumption experience is a kind economic product. Difference from the past, both product and service are external, while experience is internal and lives in personal mind. Breckler and Wiggins (1984) argued when consumers have more direct experience, it's easy for them to produce higher cognitive belief, emotional reaction, and future behavior intention for products. As a result, consumers would cultivate a consistency between attitude and behavior altitude. And, it isn't easy for consumers to change their attitudes produced by direct experience. In addition, Experience would strengthen the connection between the attitudes of consumers and products.

Schmitt（1999）invented the term, "Experiential Marketing". He thinks what nowadays consumers want is to glamour sensory, stir up spirits, trigger thoughts and make brand blended in their daily lives by products and marketing activities. For example, iPhone, Singapore Airlines, and Starbucks Coffee are all practitioners of brand experiential marketing. Petromilli and Michalczyk（1999）thought that each consumer cultivates brand experience by contacting with different kinds of behavior, and brand relationship that consumer accumulate by consumption experience. Experience can make consumers produce higher familiarity, stronger brand association and deeper product impression. As a result, experience could strengthen the relationship between people and brand (Alba and Hutchinson 1987）. Biel (1992) explained that brand image is a collection of brand attributes and the connection that consumers produce for brand names. After consumers join in Experiential Marketing, they would have positive evaluation of thoughts and activities, and then their emotions would be affected. As a result, it would help consumers strengthen the brand image. Brakus, Schmitt and Zarantonello (2009) firstly introduced the concept of brand experience and defined it as subjective, internal consumer responses (sensations, feelings, and cognitions) and behavioral responses evoked by brand-related stimuli that are part of a brand's elements (e.g., logo, character and slogan), marketing mixes efforts (e.g., product packaging, media communications, and store's environments), and secondary associations (e.g., country of origin, endorser, and third party's support).

Customer Equity

Blattberg, Getz, and Thomas（2000）proposed that customer equity thinks of consumers as a measurable, managerial, reinforceable asset just like other assets. And, management of customer equity

is a dynamic and integrated market system by manipulating financial evaluation techniques and purchase behavior data of consumer dynamics. With the three indexes of customer acquisition, customer retention, and add-on selling, market systems can work out an optimizing ratio choice. Rust, Zeithaml, and Lemon (2000) and Rust, Lemon, and Zeithaml (2004) propose a comprehensive model that identifies different drivers of a company's outcomes. They suggest a customer value model, stating that three equity drivers-value equity, brand equity, and relationship equity-influence a customer's switching matrix, which in turn has an impact on customer lifetime value (CLV) and customer equity. According to a number of researchers, three intercorrelated drivers – customer attraction, customer retention and add-on selling – are equally important in customer equity creation (Blattberg et al., 2001; Hansotia, 2004; Rust et al., 2004; Malthouse & Blattberg, 2005; Villanueva & Hanssens, 2007; Kumar & Rajan, 2009). It is emphasized that organizations should focus their marketing efforts on the value equity, brand equity and relationship equity improvement and in each stage of customer relationship development process choose the most relevant customer equity drivers (Rust et al., 2004).

As a matter of fact, customer equity is not only a measure of consumer relationship assets, but also a marketing system, an innovative method of restructuring distributing resources and maintaining relationship of consumers in enterprises. Therefore, Rust, Zeithaml and Lemon（2004）jumped off consumer lifetime value from financial view, and proposed three dimensions of customer equity from marketing view. They are value equity, brand equity, and relationship of marketing. All the three dimensions are the main resources of lifetime customer equity.

Brand Experience Contact Point

Though we have not found out there are related documents officially using the concept of brand experience contact point in experiential marketing and consumer behavior, we could detect several critical points when consumers contact with brand during the consumer purchase processes（Arnould, Price, and Zinkhan 2002；Brakus, Schmitt, and Zhang 2008；Holbrook 2000）. For instance, it's the first brand experience contact point when consumers are searching for information by contacting and communicating with media or virtual platforms. Sometimes even if consumers do not cultivate their purchase motivation, brand experience has started. What we call this phenomenon is indirect contact（Hoch and Ha 1986；Kempf and Smith 1998）. The following is when consumers begin picking out and buying, what they contact is physical environments of brand distribution channels such as the decoration and design, atmosphere and styles. At the same time, consumers also have interaction with service personnel（Hui and Bateson 1991；Kerin, Jain and Howard 2002）.After consumers buy the brand products, the following contact points are usage and consumption. It is consumers that have more direct contacts. Direct contacts may include utilitarian product attributes and category experience or hedonic dimensions, such as feelings, fantasies, and fun (Holbrook and Hirschman 1982). Therefore, this study defines brand experience contact points into four including communication contact, physical environment contact, people contact and product contact.

Framework and Hypotheses
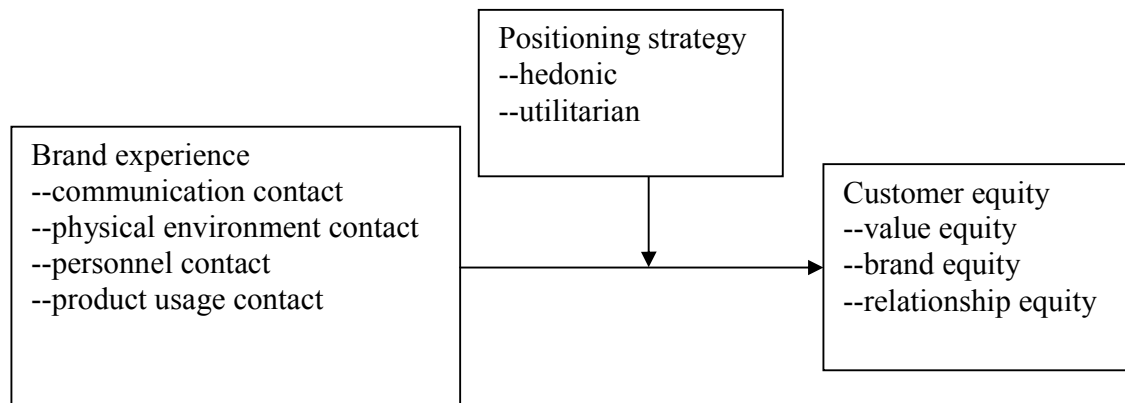
H1：Brand experience affects customer equity (value equity, brand equity, and relationship equity) positively.

H2：The different brand contact points' experience has different effects on customer equity (value equity, brand equity, and relationship equity).

H3：The relationship between brand contact points' experience and customer equity (value equity, brand equity, and relationship equity) will be moderated by brand positioning strategy.

H4：At communication contact point, different experience dimension has different effects on customer equity (value equity, brand equity, and relationship equity).

H5：At physical environment contact point, different experience dimension has different effects on customer equity (value equity, brand equity, and relationship equity).

H6：At personnel contact point, different experience dimension has different effects on customer equity (value equity, brand equity, and relationship equity).

H7：At product usage contact point, different experience dimension has different effects on customer equity (value equity, brand equity, and relationship equity).

METHODOLOGY

Measures

We developed the items for measuring the constructs of the study, drawing on prior research in the literature. We used multi-item seven-point Likert scales anchored by 1 = "strongly disagree" and 7 = "strongly agree".

Adapting from Brakus, Schmitt, and Zarantonello(2009), we defined brand experience as subjective, internal consumer responses (sensations, feelings, and cognitions) and behavioral responses evoked by brand-related stimuli that are part of a brand's elements (e.g., logo, character and slogan), marketing mixes efforts (e.g., product packaging, media communications, and store's environments), and secondary associations (e.g., country of origin, endorser, and third party's support). The brand experience dimensions included sensory, affective, intellectual, and behavior experience that will happen in the four brand experience contact points- communication contact, physical environment

contact, personnel contact, and product usage contact. The followings are the operational definition of the four experience dimensions.

Sensory：consumer response to brand-related stimuli via visual, aural, olfactory, taste, and tactile aspects.
Affective：consumer internal perception of feeling, mood, and emotional in response to brand-related stimuli.
Intellectual：consumer's convergent/analytical and divergent/imaginative thinking due to the brand-related stimuli.
Behavior：consumer response in life behavior toward the brand-related stimuli.

We developed three measurement items for each brand experience dimension in the four contact points, which resulted to total 48 items.

Value equity was defined as the perceived ratio of what is received (e.g., a product) to what must be sacrificed (e.g., the price paid for the product)(Rust, Zeithaml, and Lemon 2001; Vogal, Evanschitzky and Ramaseshan 2008). Four items were used to measure including value/price, quality/price, receive/pay, and overall perceived value toward the brand purchasing experience.

We adapted the brand equity concept from Aaker (1996)、Keller(1993) and defined it as the subjective appraisal of a customer's brand choice and the added value of the brand name. We measured it with four items like brand recall, brand value, brand likability, and brand uniqueness. We also added another open question to ask respondents to answer "how is the percentage of the current price you are willing to pay if taking out the brand name and logo?"

We defined relationship equity as the tendency of the customer stick with the brand, above and beyond the customer's objective and subjective assessment of the brand (Lemon, Rust, and Zeithaml 2001). We measured it with four items including brand relationship, brand switch, always think the brand in life, and the brand like good friend in life.

Research Design

In the first step, a pretest was run to identify the object for both utilitarian and hedonic brands. Two major retailing coffee brands- Starbucks and City Café (set inside at 7-11 convenience store, providing economic cup coffee) were chosen and tested its brand positioning strategy, using the scales that Voss et al. (2003) developed. Thirty samples were recruited via website to answer 10 questions (five for utilitarian and hedonic respectively) with seven-point Likert scales anchored by 1 = "strongly disagree" and 7 = "strongly agree". The results indicated that hedonic rating for Starbucks was 5.73 vs. 3.60 for City Café (p=0.000), but the utilitarian rating for City Café was significantly higher than that of Starbucks (4.45 vs. 3.21, p=0.001). This result confirms that Starbucks and City Café can be used as research object to represent hedonic and utilitarian brands respectively.

In the next step, two questionnaires for two coffee brands were put separately on the survey website (http://survey.youthwant.com.tw/) to recruit participants. Participants were incentivized by being entered into a free lucky drawing to win one of ten $30 gift certificates for the USB. Each participant agreed to be called if they won the drawing or for personal validation. Total four hundreds ninety three valid samples, including 410 with brand experience and 83 without brand experience, were collected within two weeks. The samples include 56% female; the major age is 20-29 (79%), following by 30-39 (15%).

Following Gerbing and Anderson (1988), we conducted a confirmatory factor analysis to assess the reliability and validity of the multi-item scales. The coefficient alphas for all variables were between 0.82-0.95, exceeded the typical threshold .7 proposed in the literature (Hair et al. 2006; Nunnally 1978).

RESULTS AND ANALYSIS

We used regression analysis to test our hypotheses. We found that the brand experience has a strong impact on customer equity (β=0.782, t=25.35, p<.000). Among three customer equity drivers, brand equity is mostly influenced by brand experience (β=0.762, t=23.76, p<.000), followed by relationship equity (β=0.720, t=20.95, p<.000). Although value equity has least impact by brand experience, the relationship is still significant (β=0.548, t=13.22, p<.000). In addition, we also compare the customer equity differences between with and without brand experience using ANOVA. The results show very significant difference for overall customer equity (4.68 vs. 3.61, t=9.236, p=.000) and for value equity (4.78 vs. 4.25, t=4.299, p=.000), brand equity (4.87 vs. 3.73, t=8.063, p=.000), relationship equity (4.37 vs. 2.87, t=10.377, p=.000). These results indicate that the positive and strong brand experience is a key factor for building strong customer equity. Therefore, H1 is supported.

Next we test the impact of brand experience on customer equity in the four contact points. Regression analysis results show that the four contact points' brand experiences all have significant impact on customer equity(communication contact,β=.110, t=2.623, p<.001; physical environment contact, β=.118, t=2.323, p<.01; personnel contact,β=.186, t=3.984, p<.001; product usage contact, β=.477, t=9.548, p<.001). In specifically, on the value equity, the other three contact points' brand experiences (except physical environment contact) all have significant influences. However, the brand experience at communication contact has less impact on brand equity, but the other three all show significant effect. For relationship equity, all four points' experiences have significant impact. It is worth to note that product usage contact always play the most important role on influencing customer equity and its individual drivers.

In order to test the moderating effect of brand concept on the relationship between brand experience and customer equity, we divided the sample into hedonic and utilitarian groups and ran the regression analysis individually. Interestingly, we found that for utilitarian brand, only brand experience at product usage contact point has significant impact on customer equity (β=.610, t=9.255, p<.001) and its three drivers (value equity,β=.493, t=6.005, p<.001; brand equity, β=.578, t=8.262, p<.001; relationship equity,β=.544, t=7.256, p<.001), the other three contact points' experiences do not have any relationship with customer equity. However, for hedonic brand, the experience at communication contact, personnel contact, and product usage contact do have significant effect on customer equity and its three drivers except experience at physical environment contact point. On the other hand, we also can not find the relationship between brand experience and value equity at product usage contact point. Based on these results, we can conclude that the relationship between brand experience and customer equity will be moderated by different brand concepts. Therefore, H3 is supported.

Finally, we examine the impact of individual brand experience dimension on customer equity at four contact points. At the communication contact point, the intellectual experience has significant influence on customer equity (β=.340, t=5.200, p<.001) and its drivers- value equity (β=.279, t=3.789, p<.001), brand equity (β=.290, t=4.253, p<.001), and relationship equity (β=.320, t=4.726, p<.001). And the affective and behavioral experience only has significant effect on relationship equity (β=.185, t=2.162, p<.05 and β=.136, t=2.102, p<.05). There has different finding for the brand experience at physical

environment contact point. The significant phenomenon happen on the behavioral experience dimension and customer equity (β=.341, t=6.670, p<.001) and its drivers- value equity (β=.256, t=3.184, p<.01), brand equity (β=.395, t=6.156, p<.001), and relationship equity (β=.456, t=6.709, p<.001). At the personnel contact point, the intellectual experience again plays an important role on influencing customer equity (β=.409, t=6.080, p<.001) and its drivers- value equity (β=.281, t=3.536, p<.001), brand equity (β=.355, t=5.062, p<.001), and relationship equity (β=.421, t=5.882, p<.001). The behavioral experience also has slightly significant impact on customer equity (β=.144, t=2.520, p<.05), mainly coming from relationship equity (β=.158, t=2.602, p<.01). At the product usage contact point, both sensory and behavioral experience dimensions have strong impact on customer equity. In specifically, sensory experience not only affects customer equity significantly (β=.319, t=5.744, p<.001), but also happens across all three drivers- value equity (β=.372, t=5.030, p<.001), brand equity (β=.302, t=5.331, p<.001), and relationship equity (β=.187, t=3.026, p<.01). However, the behavioral experience has significant impact on customer equity (β=.322, t=5.959, p<.001), particularly from brand equity (β=.278, t=5.062, p<.001) and relationship equity (β=.360, t=6.006, p<. 001), but not value equity.

CONCLUSIONS AND IMPLICATIONS

Customer centrism is essential for a firm to flourish. Brand experience and customer equity are two major contemporary streams focusing on customer in the academic and practical marketing field. The objective of this research is to explore the relationship between brand experience and customer equity (value equity, brand equity, and relationship equity), and further to examine the impact of different contact points experience (communication contact, physical environment contact, personnel contact and product usage contact), and different brand experience dimensions on customer equity. The different brand positioning strategies-hedonic and utilitarian is also being used to examine its moderating effect on the above relationships. This model offer a theoretical framework for making the firm customer centered.

This result shows that the positive and strong brand experience is a key factor for building strong customer equity. The brand experience has a strong impact on customer equity. Specifically, brand equity is mostly influenced by brand experience, followed by relationship equity and value equity. The four contact points' brand experiences-communication contact, physical environment contact, personnel contact, product usage contact all have significant impact on customer equity. Among that, product usage contact is the most important role on influencing customer equity and its individual drivers. For relationship equity, all four points' experiences have significant impact. However, physical environment contact and communication contact experience has less impact on value equity and brand equity respectively.

We also find that the relationship between brand experience and customer equity will be moderated by different brand position strategies. For utilitarian brand, only brand experience at product usage contact point has significant impact on customer equity and its three drivers, the other three contact points' experiences do not have any relationship with customer equity. However, for hedonic brand, the experience at communication contact, personnel contact, and product usage contact do have significant effect on customer equity and its three drivers except experience at physical environment contact point.

Finally, the four experience dimensions developed by Brakus, Schmitt, and Zarantonello(2009) have different impact on customer equity and its three drivers at different experience contact point. At

communication contact point, the intellectual experience has significant influence on customer equity and its drivers, but the affective and behavioral experience only has significant effect on relationship equity. At physical environment contact point, the behavioral experience dimension is a key indicator with customer equity. At the personnel contact point, the intellectual experience again plays an important role on influencing customer equity and its drivers. The behavioral experience also has slightly significant impact on customer equity, mainly coming from relationship equity. At the product usage contact point, both sensory and behavioral experience dimensions have strong impact on customer equity. Specifically, sensory experience not only affects customer equity significantly, but also happens across all three drivers. However, the behavioral experience has significant impact on customer equity, particularly from brand equity and relationship equity, but not value equity.

Table 1 Regression Analysis Results of Brand Experience at Contact points: Standardized Coefficients

Independent Variables	Customer equity		Value equity		Brand equity		Relationship equity	
	β	t value	β	t value	β	t value	β	t value
Brand experience	0.782	25.35***	0.548	13.22***	0.762	23.76***	0.720	20.95***
Communication contact	0.110	2.623***	0.149	2.613***	0.036	0.842	0.111	2.352**
Physical environment contact	0.118	2.323**	-0.075	-1.072	0.236	4.475***	0.118	2.063**
Personnel contact	0.186	3.984***	0.243	3.800***	0.121	2.502**	0.141	2.678***
Product usage contact	0.477	9.548***	0.316	4.622***	0.467	9.037***	0.451	8.028***
Hedonic								
Media contact	0.284	4.150***	0.264	2.954**	0.225	2.928**	0.234	3.143**
Physical environment contact	-0.032	-0.408	-0.072	-0.713	-.021	-.238	0.010	0.118
People contact	0.360	5.154***	0.394	4.311***	0.226	2.871**	0.302	3.985***
Product contact	0.269	3.708***	0.022	0.232	0.363	4.455***	0.283	3.599***
Utilitarian								
Media contact	0.067	1.139	0.041	0.561	0.032	0.517	0.094	1.414
Physical environment contact	0.082	1.222	0.028	0.334	0.125	1.760	0.065	0.855
People contact	0.089	1.446	0.109	1.425	0.073	1.113	0.060	0.851
Product contact	0.610	9.255***	0.493	6.005***	0.578	8.262***	0.544	7.256***

***p < .001, ** p < .01, * p < .05

Table 2 Regression Analysis Result of experience dimensions: Standardized Coefficients

Independent Variables	Customer equity		Value equity		Brand equity		Relationship equity	
	β	t value	β	t value	β	t value	β	t value
Media contact								
Sensory	0.079	1.092	0.132	1.615	0.109	1.440	-.016	-0.216
Affective	0.134	1.621	0.029	0.309	0.118	1.376	0.185	2.162*
Intellectual	0.340	5.200***	0.279	3.789***	0.290	4.253***	0.320	4.726***
Behavioral	0.121	1.855	0.062	0.851	0.110	1.618	0.136	2.012*
Physical environment contact								
Sensory	-0.047	-0.683	-0.074	-0.856	0.071	1.029	-.121	-1.666
Affective	0.204	2.441	0.106	1.015	0.197	2.381*	0.218	2.479*
Intellectual	0.149	2.315	0.182	2.263*	0.095	1.489	0.125	1.837
Behavioral	0.341	6.670***	0.256	3.184**	0.395	6.156***	0.456	6.709***
People contact								
Sensory	0.100	1.497	0.118	1.496	0.133	1.915	0.021	0.294
Affective	0.097	1.224	0.035	0.368	0.134	1.623	0.078	0.927
Intellectual	0.409	6.080***	0.281	3.536***	0.355	5.062***	0.421	5.882***
Behavioral	0.144	2.520*	0.125	1.848	0.093	1.560	0.158	2.602**
Product contact								
Sensory	0.319	5.744***	0.372	5.030***	0.302	5.331***	0.187	3.026**
Affective	0.142	2.417	-0.037	-0.474	0.256	4.267***	0.127	1.941
Intellectual	0.08	1.651	0.081	1.265	0.005	0.095	0.122	2.275
Behavioral	0.322	5.959***	0.185	2.569	0.278	5.062***	0.360	6.006***

***$p < .001$, ** $p < .01$, * $p < .05$

References is available on request

The ORDER EFFECT WITHIN PRODUCT ATTRIBUTES ON CONSUMERS' FOOD DECISION MAKING

Jun-Fang Liao, Wenzao Ursuline College of Languages, 900 Mintsu 1st Road, Kaohsiung, Taiwan, (886)7342-6031, melodyliao2009@gmail.com

Chun-Min Yang, Min-Chuan University, 5 De-Ming Road, Gui-Shan, Taoyuan County, Taiwan, (886)33507001ext.3574, cmyang@mcu.edu.tw

I-Ling Ling, National Chiayi University, 580 Shin-Min Road, Chiayi City, Taiwan, (886) 5273-2843, yiling@mail.ncyu.edu.tw

ABSTRACT

This article focuses on how attribute orders influence consumers' calorie estimation. Two experiments were conducted with 142 participants. The results indicate that consumers estimate calorie content are more influenced by the high/low attribute order than the low/high attribute order. More importantly, combining the cognitive load, the attribute order and cognitive load create an interaction effect. Cognitive load will moderate the effect of attribute order on calorie estimation. Theoretical and managerial implications of the findings are provided.
Keywords: order effect; attribute order; cognitive load; calorie estimation

INTRODUCTION

The 2007-2008 National Health and Nutrition Examination Survey (NHANES) data for adults aged 20 and over suggest an increase in obesity between the late 1980s and today in the United States, with the estimated age-adjusted prevalence moving upward from a previous level of 23% in NHANES III (1988–1994) to approximately 34% in 2007–2008 [1]. Even though a public policy effort to promote the consumption of healthy foods has been implemented, there is no significant change in obesity trends. The proportion of overweight individuals continues to increase even though many people are concerned about weight management and try to avail themselves of the healthier options that are available.

Assessment and regulation of calorie intake continues to be documented to play a central role in the prevention and treatment of many diseases. Clearly, not all consumers judge their food consumption decisions on health or nutrition consideration [2] [3]. The existence of a bias in estimating a meal calorie content naturally raises the question of how to correct it. Chernev and Gal (2011a) showed that when faced with a meal comprised of both healthy and indulgent items, consumers tend to systematically underestimate its calorie content, to the extent that they may perceive the combined meal not only as having fewer calories than the sum of its individual components but also as having fewer calories than in the indulgent item alone [4]. More interestingly, Chernev demonstrated that weight-conscious individuals, including those on diets, are *more* likely to believe that a meals tendency to lead to weight gain can be decreased by simply adding a healthy item. He referred to this phenomenon as the "*dieters' paradox*" [5]. Individuals most concerned with managing their weight will be more likely to believe that the combination of a healthy item and an unhealthy item is likely to have fewer calories than the unhealthy item alone.

Numeric judgments made with uncertainty are easily influenced by readily available anchors [6]. It has been shown that the sequential evaluations typically result in the assimilation of the numeric estimate toward the initially generated value, so that that smaller initial estimates are likely to produce smaller subsequent estimates, and larger initial estimates are likely to lead to larger subsequent estimates [7]. However, this is not always the case and that sequential evaluations depend not only on the numeric value of the previously evaluated item, but also on the semantic relationship between the evaluated items, including similar and opposing categories. For example, the author showed that items classified into opposing categories (vice and virtues) lower initial estimates and can lead to higher (rather than lower) subsequent estimates [4].

Will order effects within product attributes bias calorie estimation? Building on the existing literature [2] [3] [4] [5] [8], this research posits that the anchoring effects occur in estimating the calorie content of different sequentially presented items. In an attempt to better understand the influence of the order effect, this research seeks to explore whether attribute orders will bias a consumer's calorie estimation. We argue that order effect, especially a primacy effect, will work on the combined food attributes regardless of whether the sequences of attributes is classified into opposing categories or not. The rationale for this standpoint is discussed in more detail in the following section.

THEIRETUCAL BACKGROUND

Determinants of Visual Sensory Attention: Top-Down and Bottom-Up

There are two broad determinants of selective attention: a more deliberate top- down process versus a more automatic bottom-up process. Top-down processes describe knowledge-driven mechanisms designed to enhance the neuronal processing of relevant sensory input, to facilitate the discrimination between signal and 'noise' or distractors, and to bias the subject toward particular locations in which signals may appear [9]. Bottom-up perspectives attempt to explain a subject's ability to detect targets and target-triggered attentional processing largely by the sensory salience of the targets, and their ability to trigger attentional processing by recruiting 'higher' cortical areas in a bottom-up manner (e.g., from the processing of a visual target in the primary visual cortex to temporal regions for object identification and to arousparietal regions for location). Importantly, top-down and bottom-up processes represent overlapping organizational principles rather than dichotomous constructs, and in most situations, top-down and bottom-up processes interact to optimize attentional performance[10].

Based on their food research, Smith and Kosslyn emphasize that top-down processing holds that external information provided about the food is processed more deliberately and that it affects taste perception in a cognitive manner [11]. On the other hand, bottom-up processing suggests that information about the food is processed more automatically and heuristically, when driven by stimulus such as intrinsic cues. Sometimes it is questionable which process is working. We look at outside influences both in bottom-up and top-down processing contexts. For example, Raghunathan, Naylor, and Hoyer's research examines more bottom-up processing and the effect of labeling a food item as either healthy or unhealthy and provides support for an unhealthy-equals-tasty intuition that consumers hold, whereby food categorized as unhealthy results in higher taste perceptions than food categorized as healthy[12]. Krishna and Morrin (2008) demonstrate the automatic effect of extrinsic cues such as product haptics on taste perception [13]. They show that the haptic quality of glasses from which water and other drinks are consumed can affect taste perception. They argue that the haptic effect on taste is automatic and that more deliberate processing would make people realize that the containers are non-diagnostic for taste and should not affect their perception. Thus, both processes may operate simultaneously and interact with one another. Within the present context, the ambiguity of attribute orders would then lead to more susceptibility to, and increased utilization of, external influences in forming overall calorie estimation.

Order Effects in Sequential Estimates

There is a long history of research into response anchoring effect in surveys. The observed effect is generally that when items are presented to respondents in a list, they are more likely to choose those at the top of the list – called a 'primacy effect'. In contrast, when options are read out to respondents, they are more likely to choose the last items in the list – a recency effect'. A large number of experiments have examined these effects, and a number of theories have been put forward as to why they might occur. Early theories focused on the impact of memory. This can help to explain the recency effects seen in questions where a large number of complex responses are read out to respondents, as recall of early items from short term memory is likely to be more difficult [14]. However, this theory is much less satisfactory in explaining those cases where the same effect is observed with a small number of simple items, and where respondents have a written list of items to which they can refer.

The proposition that people's decisions are independent of the context in which the decision is made has been challenged by mounting evidence suggesting that numeric judgments made under uncertainty are easily influenced by readily available anchors[6][15][16]. Thus, it has been shown that sequential evaluations typically result in assimilation of numeric estimates toward initially generated values, so that smaller initial estimates are likely to produce smaller subsequent estimates and larger initial estimates are likely to lead to larger subsequent estimates[17][18].

Alternative theories have therefore focused on the cognitive processes that respondents go through when choosing items [19]. Here, primacy effects for visual material are the result of two key factors. Firstly, options presented early in any list may help establish a cognitive framework or standard of comparison that influences interpretation of later options. These early options may therefore assume special significance with respondents. Secondly, early items are also likely to be subject to greater cognitive processing so that by the time respondents consider later items their minds may be cluttered with thoughts about previous items, which may, in turn, prevent full consideration of these later items.

> H_1: Consumers will perceive a higher calorie count in the high/low attribute order than in the low/high attribute order.
> H2: Compared to a low cognitive load, consumers with a high cognitive load will perceive a higher calorie sweetness in the sweet/sour attribute.
> H3: Compared to a low cognitive load, consumers with a high cognitive load will perceive a lower calorie sweetness in the sour/sweet attribute.

EXPERIMENT 1

Pretest

A pretest was conducted to recognize participants' knowledge regarding calorie estimation. We adopted beef as the high calorie attribute and vegetables as the low calorie attribute. Fifty-five graduate students were recruited from the business school at a university located in south Taiwan. The manipulation check of calories indicated that 100 g beef is higher than 100g vegetables (M_{beef} = 321.25, $M_{vegetables}$ = 84.78, 1, 53) =31.79, p < .01). The manipulation check is effective. We adopted beef to represent the high calorie attribute and vegetables to represent the low calorie attribute.

Participants and Design

Fifty-two undergraduate students participated in this experiment in exchange for credit. Participants were randomly assigned to one of two (low/high or high/low) conditions. Participants then read one of the two ads and were told to ask the researcher for the MOS hamburger when done reading.

Procedure

In the low/high attribute order of the MOS hamburger, the ad showed the following: Organic Vegetables+ New Zealand Beef = MOS Energy Hamburger. In the high attribute orders of the MOS hamburger, the ad showed the following: New Zealand Beef + Organic Vegetables = MOS Energy Hamburger

Measures

In order to avoid participants having no idea regarding calorie estimation, we provided a reference (sandwich: 215 calories). After reviewing the experimental advertisement, the participants were asked to answer one question: "What do you think about this hamburger's calorie count?"

Results and Discussion

It was found that a significant difference existed among two levels of attribute order ($M_{high/low}$ = 533.40, $M_{low/high}$ = 408.78; $t(50)$ = 3.54, $p < .01$). The results indicated a primacy effect that consumers in the high/low order condition will estimate calories to be higher than those who were in the low/high order condition. Thus, H_1 was confirmed. Furthermore, there is no difference in gender($t(50)$ = 0.58, $p > 0.1$), frequency of consumption($t(50)$ = 1.49, $p > 0.1$) , diet experience($t(50)$ = 0.33, $p > 0.1$). The result shows that there is a significant difference in BMI (Body Mass Index). Splitting participants into two groups by median score, it was found that a significant difference existed among the two levels of BMI ($M_{low-BMI}$ = 520.57 $M_{high-BMI}$ = 416.80; $t(50)$ = 2.84, $p < .05$). The finding was consistent with Wansink (2006). Participants who are overweight will underestimate food's calorie.

The objective of this experiment was to test the effect of attribute order on calorie estimation. The results show that order effect on hamburger's attributes is the primacy effect. When consumers perceive the attribute order as high/low, they overestimate the calorie count of the hamburger. In contrast, when consumers perceive the attribute order as low/high, they underestimate the calorie count of the hamburger. Due to the impact of anchoring, people are likely to be subject to greater cognitive processing and as a result will give a higher weight to early items because by the time the consumer considers later items, their minds may be cluttered with thoughts about previous items, which may in turn prevent full consideration of these later items[19].

Elder and Krishna(2010) indicate that the processing of ads is deliberate and cognitive, so that the enhancing effect of multiple-sense ads is reduced when cognitive resources are constrained. Study 2 will investigate the interaction effect of inconsistent multiple-attribute order and cognitive load on sweetness estimation. In Study 2, we will use drinks as the experimental stimuli to test the robustness of the primacy effect obtained in Study 1.

EXPERIMENT 2

Participants and Design
Ninety-eight undergraduate students participated in this experiment in exchange for course credit. Participants were randomly assigned to one of four conditions.

Procedure
In the manipulation for attribute orders, we use two opposite attributes (sweet/sour vs.

sour/sweet). Two kinds of mixed drinks (Honey Lemon Tea vs. Lemon Honey Tea) were shown separately in the ads. According to the manipulation for cognitive load, the authors used music and a pencil test. In the high cognitive load condition, participants needed to memorize five slogans about drinks while a loud   music was playing in the background. Following the cognitive load task, participants involved in the experiment were asked to estimate the sweetness of the inconsistent-attribute tea after reviewing the experimental information (a drink ad).

Measures

Coca-cola was provided to give the participants a baseline regarding sweetness estimation. After reviewing the experimental advertisement, the participants were asked to answer the question: "What do you think about this drink's sweetness?" using the Likert scale ranging from 1 to 6.

Results and Discussion

A 2(attribute orders: sweet/sour vs. sour/sweet) $\times$ 2(cognitive load: yes vs. no) between-subjects ANOVA was conducted. The means values for sweetness estimation in different situations were shown as in Figure 1. In terms of the effect of attribute order, the results reveal main effect on sweetness estimation ($F(1,94) = 12.57$, $p < .01$, respectively). Not surprisingly, sweetness estimation is higher when the attribute order is sweet/sour than when it is sour/sweet. A non-significant main effect of cognitive load on sweetness estimation ($F(1,94) = 0.44$, $p > .05$) is shown. Furthermore, the attribute order $\times$ cognitive load interaction reached significance on sweetness estimation ($F(1,94) = 2.67$, $p < .1$). As predicted, the cognitive load effect emerged as significant for the sweet/sour attribute order but not for the sour/sweet attribute order condition. For the low cognitive load, there is no significant difference between the high/low attribute order and low/high attribute order in sweetness estimation.

Under cognitive load and time pressure, consumers estimate sweetness in the high/low attribute order condition ($M = 6.09$), higher compared to the low/high attribute order condition. For low cognition load, there is no significant difference between the high/low attribute order and low/high attribute order in calorie estimation. Thus, H_2 are supported by these findings. Cognitive load will moderate the effect of attribute order on sweetness estimation. The findings show that the anchoring effect was more significant when consumers perceived cognitive loading and time pressure, increasing the attribute order effect.
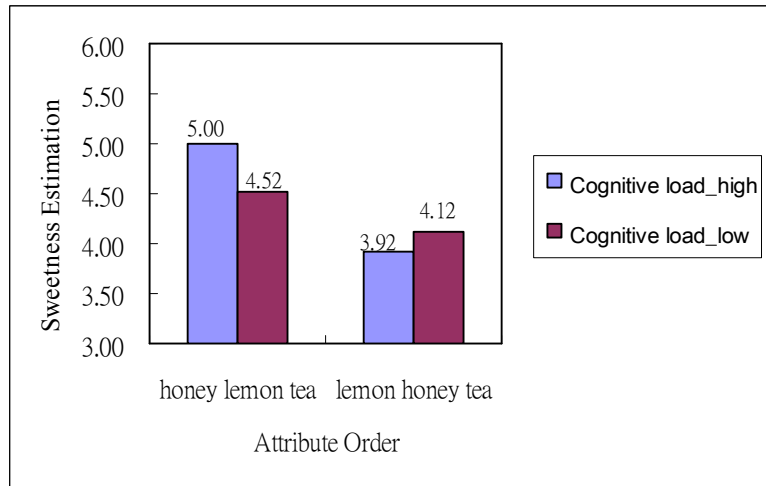
Figure 1 The Effects of Attribute Order and Cognitive Load on Sweetness Estimation

GENERAL DISCUSSION

In the past, most research discussed food decisions based on labels, advertising, taste and package size. This research focused on a new perspective to investigate whether order effect within product attributes will bias consumers' calorie estimation. This article also investigates the moderating role of cognitive load on calorie estimation. There were two experiments conducted. Experiment 1 is a two-level experimental design to discuss the order effect with product attributes. Experiment 2 is a 2(attribute order：high-low/low-high）× 2(cognitive load：yes/ no) between-subject design and the purpose is to understand the interaction effect of the cognitive load and order effect on calorie estimation.

Results show that order effect within product attributes will bias calorie estimation. This bias from the anchor, which consumers give the product, affects the priority attribute. Furthermore, Experiment 2 demonstrates that the order effect will be moderated by the cognitive load. When consumers are under the condition of cognitive load and time pressure, they do not have enough cognitive resource to handle the information. This finding shows that the order effect will be much stronger and result in wrong calorie estimation. Theoretical and managerial implications of the findings are provided.

The findings in this article have important public policy implications. The results show that in the attribute order of high/low sequence consumers tend to overestimate a product's calorie content. In particular, we also show that consumers tend to underestimate the calorie content

of attribute order of low/high sequence. In this context, an important implication of the findings reported is that providing calorie information at the time of food selection could help minimize the over-consumption resulting from the reported averaging bias. Yet the findings suggest that this approach can sometimes yield the opposite results when it comes to monitoring calorie intake, so that health-based categorization can lead to underestimation of the calorie content of combinations of healthy and indulgent items. In turn, this can lead to counterproductive behaviors because, though people think they are eating a healthier, less caloric meal, they are actually consuming more calories than they realize.

Thus, whereas the underestimation of a meal's calorie content in simultaneous evaluations of virtue/vice combinations is likely to promote overconsumption, the overestimation of a meal's calorie content in sequential evaluations of virtue/vice combinations is likely to promote greater self-regulation of the consumption behavior[20]. In this context, sequential valuations can be strategically used to manage a meal's perceived calorie content and individuals' consumption behavior.

REFERENCES

[1] CDC. http://www.cdc.gov/nchs/fastats/overwt.htm, 2011

[2] Chandon, P. & Wansink, B. *The Biasing Health Halos of Fast Food Restaurant Health Claims: Lower Calorie Estimates and Higher Side-Dish Consumption Intentions*. Journal of Consumer Research, 2007, 34(3), 301−314.

[3] Chandon, P., & Wansink, B. *Is Obesity Caused by Calorie Underestimation? A Psychophysical Model of Meal Size Estimation.* Journal of Marketing Research, 2007, 44(1), 84−99.

[4] Chernev, A. (in press). *Semantic Anchoring in Sequential Evaluations of Vices and Virtues.* Journal of Consumer Research.

[5] Chernev, A. *The Dieter's Paradox*. Journal of Consumer Psychology, 21 (April), electronically published September 14, 2010.

[6] Tversky, A. & Kahneman, D. *Judgment under Uncertainty: Heuristics and Biases. Science*, 1974, 185 (4157),1124–31.

[7] Epley, N., & Gilovich, T. *The Anchoring-and-Adjustment Heuristic*. Psychological Science, 2006,17, 311–318.

[8] Scott, M.L., Nowlis, S.M., Mandel, N. & Morales, A.C. *The Effects of Reduced Food Size and Package Size on the Consumption Behavior of Restrained and Unrestrained Eaters. Journal of Consumer Research*, 2008, 35 (3),391–405.

[9] Kastner, S.& Ungerleider, L.G. *Mechanisms of Visual Attention in the Human Cortex.* Annual Review of Neuroscience, 2000, 23:315–41

[10] Egeth, H. E. & Yantis, S. *Visual Attention: Control, Representation, and Time Course.*

Annual Review of Psychology, 1997, 48, 269-297.

[11] Smith, E. E. & Kosslyn, S.M. Cognitive Psychology: Mind and Brain, Upper Saddle River, NJ: Pearson Prentice Hall. 2007

[12] Raghunathan, R., Naylor, R.W. & Hoyer,W.D. *The Unhealthy = Tasty Intuition and Its Effects on Taste Inferences, Enjoyment, and Choice of Food Products*. Journal of Marketing, 2006, 70 (October), 170–84.

[13] Krishna, A. & Morrin, M *Does Touch Affect Taste? The Perceptual Transfer of Product Container Haptic Cues*. Journal of Consumer Research, 2008, 34 (2), 807–18.

[14] Schwarz, N.& Bless, H. *Constructing Reality and Its Alternatives: Assimilation and Contrast Effects in Social Judgment*. in The Construction of Social Judgments, ed. Leonard L. Martin and Abraham Tesser, Hillsdale, NJ: Erlbaum, 1992, 217–45.

[15] Chapman, G. B. & Johnson, E. J. *Anchoring, Activation, and The Construction of Values*. Organizational Behavior and Human Decision Processes , 1999,79, 115{153.

[16] Green, Donald, Jacowitz, K.E. Kahneman, D. McFadden, D. *Referendum Contingent Valuation, Anchoring, and Willingness to Pay for Public Goods*. Resources and Egry Economic, 1998, 20(2),85-116

[17]Ariely, D., Loewenstein, G. & Prelec, D. *Coherent Arbitrariness: Stable Demand Curves withoutStable Preferences*. Quarterly Journal of Economics, 2003, 118(1), 73–105.

[18] Ariely, D.*Tom Sawyer and the Construction of Value*. Journal of Economic Behavior and Organization, 2006, 60 (1), 1–10.

[19] Krosnick J.A., Alwin, D.F. *An Evaluation of A Cognitive Theory of Response-Order Effects in Survey Measurement*. Public Opinion Quarterly, 1987, 51:201-219

[20] Chernev, A., & Gal, D. *Categorization Effects in Value Judgments: Averaging Bias in Evaluating Combinations of Vices and Virtues*. Journal of Marketing Research, 2010, 47(4), 738−747.

THE EXPLORATION OF TEAM LEARNING PERFORMANCE: THE EMPIRICAL STUDY ON COMMUNITY OF PRACTICE OF HOSPITALS IN TAIWAN

Shih-Wang Wu

Department of Hospital and Health Care Administration,

Chia Nan University of Pharmacy & Science, Taiwan

P.O.BOX 62-13, Tainan, Tainan City 70899, Taiwan

+886-6-2664911 ext.5225, scottwu101@mail.chna.edu.tw

Yafang Tsai

Department of Health Policy and Management,

Chung-Shan Medical University, Taiwan

No.82, Sec. 2, Chongsing Rd. Taichung City 404, Taiwan

+886-4-24730022 ext.17176, avon611@yahoo.com.tw

ABSTRACT

In a knowledge-based economy, an excellent team for learning can review their own performance and encourage participation in order to reach the goal of organization. The basic unit of learning is team in modern organization, so that the purpose of team learning is to motivate team and promote the progress of individual and organization. The way of team learning in common use is the community of practice in order to increase the problem-saving capability, like Quality Control Circle (QCC), grant round or seminar of medical record. This research took QCC in hospitals as the research object, and issued the structural questionnaires to members belonging to each QCC. 493 hospitals passed the hospital accreditation survey and contracted with Bureau of National Health Insurance in 2007 were the research subjects. The study expected to know if team learning can bring synergy, and to explore the relationship among active team knowledge acquisition and team learning performance in hospitals.

Keywords：Team Learning, Team Knowledge Acquisition, Community of Practice

INTRODUCTION

When directors of different departments have to make a collective decision, they usually

make different decisions because of their lack of other information. In order to eliminate the discordant opinions, the managers usually call the directors together, try to find the common view and increase the validity of decision. The directors discuss and learn to each other and exchange their information and then can make perfect strategies [1].

The learning way in common use of health care administration is Community of Practice (COP). Lave &Wenger (1990) proposed the concept of COP, the COP is to find some persons together to share their interest and professional opinions for common activities of their team, and the team is non-formal type [2]. Fox, Duffy & Conner thought that the organization is consisted of many different COP, and every community operates by their own pattern, for example, learning by doing, and mentoring [3]. It is the way close to Problem Based Learning (PBL) and can increase the problem saving capability, for example, Quality Control Circle (QCC), clinical seminar, etc. QCC is one of the most common activities, so that we take QCC team to be our samples in order to explore the influence of learning.

LITERATURE REVIEW

Edmondson defined team learning is kind of process that can help the team to understand their environment and customers, and can push to reach the goal of the team [4]. Edmondson emphasized that the team members should take communication and cooperation as the key-point in team learning, and establish the new paradigm [5]. Edmondson redefined team learning as a dynamic process including information sharing with members, performance feedback, problems discussion and going for new insight [6]. The team could not only increase their ability to execute their mission, but also could innovate new knowledge.

As regards the learning performance is the change in knowledge, skill and attitude after learning. The self-performance evaluation is to know if the learner have self-confidence and capability to present though their external behavior [7]. Ingram also proposed that the desire to learn of information receiver in learning process will influence the learning performance a lot, that is the more active the learner, the better the performance [8].

Chan, Lim & Keasberry defined the active knowledge acquisition to be internal learning and external learning, and expected to measure the initiative of team members [9]. And they wished to understand if there is positive change in team learning with appropriate learning process. In the perspective of the measurement of learning performance, Rosamaria Valle, et al. proposed four dimensions to evaluate if learning does happen [10]. The four

dimensions are independent analysis, team member interaction, reasoning ability and active participation. It can be expected to understand the behavior influence after team learning, and the results showed that the more active the team members, the better the performance.

METHODOLOGY

493 hospitals passed the hospital accreditation survey and contracted with Bureau of National Health Insurance in 2007 were the research subjects. We chose the QCC to stand for team in hospitals, and took their members to be our sample by convenience sampling. We issued the structural questionnaires to members belonging to each QCC. The valid questionnaires were 94, and that meant the rate of valid retrieving was 19%. Because there are almost 166 hospitals implement QCC according to the data provided by Taiwan Joint Commission on Hospital Accreditation and Quality Improvement (TJCHA), and we can get the retrieving rate rising to 57%. Each of Cronbach α of "internal learning" and "external learning" is 0.864 and 0.881. Each of Cronbach α of "independent analysis", "team member interaction", "reasoning ability" and "active participation" is 0.975, 0.956, 0.971 and 0.961.

RESULTS

The measurement of active knowledge acquisition

In the perspective of active knowledge acquisition, the sort of top five items are as follows. 1. We will learn from mistakes by discussion in our team. 2. We tend to openly discuss when we have different opinions in our team. 3. We encourage proposing questions and mistakes to communicate appropriately. 4. There are always members to check and review if the working process is appropriate. 5. We will discuss how to improve our working process regularly. The results show us that the interviewers perceived more internal learning, and score internal learning higher.

The measurement of team learning performance

In the perspective of team learning performance, the sort of top five items are as follows. 1. The team members can respect their colleague each other. 2. The members can follow the collective decision. 3. The members can listen to other members' opinions. 4. The members can accept colleagues' suggestion about their work. 5. The members can clarify the fact and professional terms. The highest four items are belong to team member interaction, and the result show us that interviewers perceived more member interaction of learning performance.

The analysis of team characteristics' influence on active knowledge acquisition

The score of interviewers who participate actively (M = 3.99) is significantly higher than who join with other reasons (M = 3.60). The score of the team who have internal and external assistants (M = 4.03) is significantly higher than the team without any assistant (M = 3.81). The score of the team who receive internal and external awards (M = 4.09) is significantly higher than the team without any award (M = 3.89). We also find the longer the QCC operated, the higher the active knowledge acquisition. But the more times that the team execute QCC, the lower the active knowledge acquisition.

The influence that active knowledge acquisition on team learning performance

The result (Table 1) showed that active knowledge acquisition can influence team learning performance positively. The result was also fit our hypothesis, and it showed the same finding with previous study, and proved that the model can work in health care industry..

Table 1　The examination of the model that active knowledge acquisition influence team learning performance positively

The relationship of latent variables	Prediction	Coefficient	t-value	Result
active knowledge acquisition → team learning performance	+	0.61*	9.62*	Supported

*表 *p*<0.05

<center>**DISCUSSION**</center>

The influence that team characteristics on active knowledge acquisition

In the attribute of COP that responded the questionnaire, there were 72.97% interviewers express that their bosses ask them to participate the QCC. The result showed that most people were not joining their team actively, and it would affect their initiative of learning. So interviewers who participate actively will show significantly higher initiative. There are 85% teams who have internal or external assistants or both. The score of the team who have internal and external assistants will show significantly higher initiative.

The score of the team who received internal and external awards was significantly higher than the team without any award. It might be occurred because the team receive awards

could introduce excellent capability and experience to discuss and analyze more precisely. We also found the longer the QCC operated, the higher the active knowledge acquisition. We thought that if the team members have more time to operate, they can discuss more detailed, and thus can show the positive influence.

The influence that active knowledge acquisition on team learning performance

The result of active knowledge acquisition showed that most members scored the internal learning higher. It showed that the perception of working process discussion, operative way and problem seeking is obvious. In the perspective of team learning performance, members scored the team interaction highest, but scores of other dimensions were still higher than 4 degree. It showed the members perceived excellent performance after participating their team.

We also can confirm that initiative of knowledge acquisition significantly affect the team learning performance. That is if the stronger the initiative of knowledge acquisition, the better the learning performance. It also showed that team members will present more spirit of independent analysis, and who can interact with their colleagues more, and can improve the reasoning ability, and strengthen the motive and induce the behavior of active participation.

CONCLUSIONS

We proposed three conclusions according to the discussion as follows.

1. In the competitive environment, the higher level of hospital accreditation the hospitals were, the better team learning performance their teams showed. That is there will be positive changing via external monitoring, ex: the hospital accreditation system and customers.

2. The administrators should strengthen the active team knowledge acquisition first in order to induce better team learning performance. If administrators can provide the incentive or set up the internal instructor at least, the staffs could be motivated the will to learn actively.

3. The active team knowledge acquisition does positively influence team learning performance significantly. The administrators can improve the team learning performance via inducing active team knowledge acquisition.

REFERENCE

[1] Edmondson, A.C. (2003). Speaking up in the operating room: How team leaders promote learning in interdisciplinary action teams. *Journal of Management Studies*, 40(6), 1419-1453.

[2] Wenger, E. C., & Snyder. W. M. (2000). Communities of practice: The organizational frontier. *Harvard Business Review*, 78(1), 139-146.

[3] Fox, V. J., Duffy, W. J & Conner, R. (2000). Applying the health care industry representative statement to your setting. *Association of Operating Room Nurses*, 72(2), 266-273.

[4] Edmondson, A. C. (1999). Psychological safety and learning behavior in work teams. *Administrative Science Quarterly,* 44(2), 350-384.

[5] Edmondson, A.C., Bohmer, R. M. & Pisano, G. P. (2001). Disrupted routine: Team learning and new technology implementation in hospital. *Administrative Science Quarterly*, 46(4), 685-717.

[6] Edmondson, A. C. (2002). The local and variegated natured nature of learning in organizations perspective. *Organization Science*, 13(2), 128-146.

[7] Piccoli, G., Ahmad, R. & Ives, B. (2001). Web-based virtual learning environments: A research framework and a prim nary assessment of effectiveness in basic skills IT training. *MIS. Quarterly*, 25(4), 401-426.

[8] Ingram, P. (2002). Interorganizational learning. In Joel A. C. Baum (Ed.), *Companion to Organizations* (pp. 642-663). MA: Blackwell.

[9] Chan, C. C., Lim, L., & Keasberry, S. K. (2003). Examining the linkages between team learning behaviors and team performance. *The Learning Organization*, 10(4), 228-236.

[10] Rosamaria, V., lleana, P., Adrian, Martinez-Gonzalez., Jose Antonio, Rojas-Ramirez., Sara, Morales-Lopez. & Beatriz, Pina-Garza. (1999). Assessment of student performance in problem-based learning tutorial sessions. *Medical Education*, 33(11), 818-822.

MAINTENANCE VERSUS TRANSFORMATION:
A PARADOX OF THE SLACK IN DYNAMIC ORGANIZATIONAL PROCESS

Ching-Chou Chen

Department of Business Administration, National Dong Hwa University
No. 1, Sec. 2, Da Hsueh Rd., Shoufeng, Hualien 97401, Taiwan, R.O.C.
ching@mail.ndhu.edu.tw, (886) 3863-3017

Hsin-Hua Hsiung

Department of Business Administration, National Dong Hwa University
No. 1, Sec. 2, Da Hsueh Rd., Shoufeng, Hualien 97401, Taiwan, R.O.C.
hsiung@mail.ndhu.edu.tw, (886) 3863-3027

Nien-Tai Tsai

Department of Business Administration, National Dong Hwa University
No. 1, Sec. 2, Da Hsueh Rd., Shoufeng, Hualien 97401, Taiwan, R.O.C.
lothar49@ms21.hinet.net, (886) 3863-3017

ABSTRACT

Using insights from strategy management and organization behavior literature, we explore how internal and/or external organizational conditions influence slack formation, transformation/maintenance and feedback. Proposing a comprehensive framework, this paper addresses a current paradox of organizational slack by systemically linking its antecedents and subsequences to dynamic organizational procedures. We argue that organizational anticipations motivate them to choose different influence paths (maintenance or transformation) to keep the levels and the types of their slack. This study also develops a set of propositions to aid the future research on the contingencies that produce high quality strategic decision making for variable slack portfolio.

Keywords: slack portfolio, dynamic organizational process, maintenance, transformation

INTRODUCTION

Although organizational slack refers to the stock of excess resources available to the organization during a given planning cycle [31] [36] [43], on the other hand, an organizational slack can be understood as an expression of what that an organization intends to change or maintain current situations according to its circumstances. Past literature tended to take a statical view and regarded the organizational slack as a contextual control variable when theorists investigate strategic actions and organizational phenomena [12] [33]. However, slack in organizational process are dynamic resources between characteristics, quantity, and quality,

especially in planned changes [20], environmental shifts [9] and competitive responses [24] [40]. Indeed, recent advances in organizational slack studies have begun to provide a finer-grained analysis of the ways different types of slack affect organizational outcomes through circumstance conditions [43].

The main purpose of this study was to integrate the current conceptualizations of internal/external conditions, strategic choices, and performance feedback for organizational slack as they pertain to the dynamic organizational process. Existing scholarship lacks careful theorizing about how various external/internal organizational conditions jointly have an impact on organizational slack. Further, the ambiguous relationship between slack and performance has not been completely solved yet [11]. Consequently, up to date we do not have sufficient ability to support or refute what slack aids or harms its organization [5] [21]; the question needs more unequivocal evidences [41] and theoretical explanations for priori but post hoc arguments [18]. Despite these controversies, a fruitful integration of the differing conceptualizations of organizational slack is possible. The key to integration is to recognize temporal differences among internal/external organizational conditions, strategic choice, and responses for organizational slack. Moreover, longitudinal, integrative theoretical analyses that are based on stimulus-response, organizational capability, as well as motivation and the covering of a feedback loop are relatively scarce. Our theorizing focuses on four key establishment stages: Firstly, slack presence was triggered by internal/external organizational conditions under varied structural constrains. Secondly, strategic choices were based on slack portfolio, motivation, and capability. Thirdly, slack transformation or maintenance affects organizational performance depending on the extent of fitting market expectations. Finally, the real performance would carry out feedback adjustments for organizational indigenous controls.

Following stability and change as a duality [15], we propose a comprehensive framework to solve the paradox of the slack in dynamic organizational process. The efficiency perspective of organizational slack, grew out of an economic research tradition, and tended to focus on operational functions, but not being used to resources as strategic management [2] [32]. In the line of strategic management, Mishina, Pollock, and Porac (2004) identified and specified the very roles that specific forms of slack play in different behavior and outcomes. Further, different forms of slack resources provide firms with a greater or lesser degree of flexibility in their approach to be efficient or adaptive [37]. From slack cycle and its varied types in dynamic organizational process, we can clearly explain why slack sometimes aids performance but other times do not. Thus, the main contribution of this study was to model antecedents and subsequences for organizational slack, to solve the questions that different perspectives on organizational slack could not coexist.

THE SLACK IN DYNAMIC ORGANIZATIONAL PROCESS

Organizational theory researchers have proposed antecedents of organizational slack which gives the firm latitude in managing changes in response to internal or external changing environments [38]; these antecedents have generally focused on predictors that industry, firm, and coalition level separately described as environmental contingencies, organizational characteristics, and the values and beliefs of dominant coalition [36]. We further categorized these factors into two types, internal and external organizational conditions, which might be provided to the strategic decision-making process.

Being resource limitations, organizations produce varied slack forming the portfolio when slack presence is inevitable. Different types of organizational slack give resource allocators varied degrees of discretion and flexibility in particular ways that can be used to reduce or buffer internal or external pressures [36]. Organizational expectations may make slack changing its content and form under certain scope of slack portfolios. Although organizational slack can ease pressures from exogenous contexts and indigenous controls, slack itself also exist pressures to the organization until those strategic decisions have being made.

Strategic choice is a long embraced notion in the field of strategic management [10]. Organizations are under pressure to select the best choice for resource allocations. For organizations with finite resources, it is imperative that they actively evaluate current resources and then divest resources of less value to generate the slack and flexibility needed to acquire or accumulate higher ones [42]. Sirmon, Hitt, and Ireland (2007) have suggested that resources that are not likely to contribute to developing or maintaining a competitive advantage or excess resources that cannot be bundled and leveraged profitably are viable candidates for strategic divestment [39]. Different slack that may be exchanged form strategic resource divestitures, for example, which are layoffs of human capital, divestitures of noncore businesses, sell-offs of specific assets, spin-offs of businesses, and outsourcing of functions. However, organizations investing in real options are often unaware of the future value for these sources [25]. Slack portfolio within organizations will keep continually dynamic adaptations depending on their motivation and capability. Organizations can divest valuable resources and harm their ability to build capabilities that can be leveraged successfully when they hasten to reduce costs in response to changes in competitive or survival conditions [30] [39]. Therefore, having the pressure to strategic decision, organizations with slack portfolio will strategically choose to transform or to maintain existing slack according to their motivation and capability.

Organizational indigenous controls may change deriving from the levels of performance. From the behavioral theory of the firm, Greve (2003) considered that an organization takes a problemistic search when the performance is low and makes a slack search when resources

are in excess [17]. If the organization has higher levels of performance, its indigenous controls may maintain or release the conditions.

As mentioned previously, four general sets of steps (internal/external conditions and slack presence, strategic response, outcomes of implementation, and reaction from feedback) illustrate the variations of organizational slack in dynamic organizational process. Our theorizing therefore synthesizes prior research and in doing so is able to develop a clearer elaboration of four stages of the organizational slack process. From these steps we propose a conceptual framework FIGURE 1 shows in detail, our conceptual model positions slack as a responsive outcome in the dynamic organizational process. Each of these steps is described in further detail next.

The slack presence from internal/external organizational conditions

Although organizational slack commonly produce within organizations, some specific internal and external conditions may easily make slack presence. We compiled, supplemented, and extended Sharfman's et al. (1988) work [36] into two categories: exogenous contexts and indigenous controls.

Firstly, exogenous contexts include external organizational conditions and factors which organizations are autonomously unable to change. Organizational slack have functions as that cushion of actual or potential resources which allows an organization to adapt successfully to external pressures for change in policy, as well as to initiate changes in strategy with respect the external environment [5]. Organizations need to respond to environmental opportunities and threats they perceived [7]. Specific external conditions can affect the type and the magnitude of organizational slack. Sharfman et al. (1988) argue that three factors (the interaction of the rate and magnitude of general environmental change, the availability of resources in market, and the structure of that market) in environmental characteristics and two factors (basic nature of the output and stage in the industry life cycle) in industry structure, are external conditions that lead to the development of organizational slack [36]. We included examples of such factors as outlined by Sharfman et al. and suggested five adaptive conditions; organizational age, stage in industry life cycle, environmental turbulence, munificence, and technology readiness.

FIGURE 1

A concept framework underlying the slack in dynamic organizational process
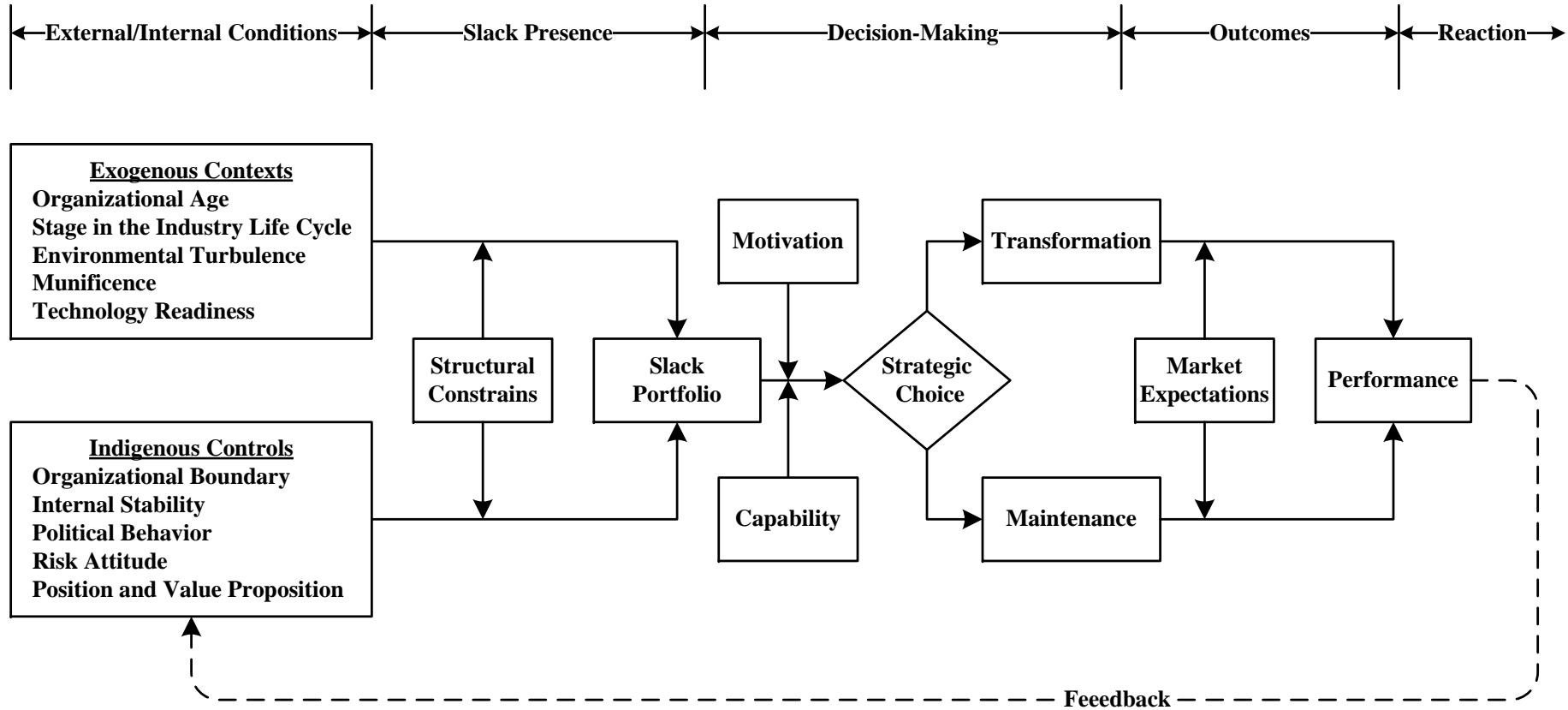
Secondly, organizations owned some artificial factors influencing change, accumulation, and reallocations of the slack resources. These factors come from indigenous controls which organizations can autonomously change. Sharfman et al. (1988) noted five factors in organizational characteristics such as size, performance, age, technology, and internal stability and two factors in values and beliefs of the dominant coalition such as political behavior and risk attitude [36]. Larger organizations mostly with market power and financial resources were better able to undertake risky long-term investments unrelated directly with operational efficiency [29] [35]. In this paper, we identified five specific internal controls that trigger to the creation of organizational slack. They are including organizational boundary, internal stability, political behavior, risk attitude, and position and value proposition. With resource limitation, the organizations formed vary slack portfolios under different internal and external conditions.

Different exogenous contexts and indigenous controls may cause accumulations for distinct slack under structural constrains. Recent work in the field of strategy has begun to acknowledge that organizations may actively or passively produce or develop certain types of slack resources [27]. Specifically, slack resource absorption and potentiality may lead to structural constrains on its deployment in different organizational activities. By considering the rarity and absorption of a slack resource, Voss et al. (2008) propose four types of slack resources that result from combining these two resource characteristics: financial, customer relational, operational, and human resources [43].

Organizational slack can generally be categorized by the nature or by the functions. Following the nature of the slack, Bourgeois and Singh (1983) suggested three types including available, recoverable, and potential slack [6]. Similarly, Singh (1986) suggested two types including absorbed and unabsorbed slack [38]. In contrast, following the functions of the slack, Mishina et al. (2004) suggested human resource slack [27]. Further, Voss et al. (2008) proposed four types including financial, customer relational, operational, and human resource slack [43]. In terms of relational slack connecting with interpersonal relationships, customer, social capital, and network capital can be included. In this sense, especially, the authors also extended both the notion of social capital [1] [28] and the resource-based view of relational resources [4] [45], and suggested that research should consider the effects of relational slack resulting from other organizational stakeholders including social capital among work groups and network capital with key suppliers. Adopting the slack functions, we can gain more insights for practitioners and theorists from empirical specific operationalization.

Proposition 1a: Apart from effects of indigenous controls, the likelihood of portfolio formation for organizational slack is positively affected by organizational age, stage in the industry life

cycle, and environmental turbulence, but is negatively affected by munificence, and technology readiness.

Proposition 1b: Apart from effects of exogenous contexts, the likelihood of portfolio formation for organizational slack is positively affected by the extent of organizational boundary, internal stability, and political behavior, but is negatively affected by risk attitude, and position and value proposition.

Structural constrains for forming slack portfolios

The nature of resources, for slack portfolio, has two types of structural constrains: absorption and potentiality. From inside organizations, Voss et al. (2008) suggested the absorption is one of critical structural constrains [43]. Unabsorbed slack refers to some resources that exist currently not to be committed and can be redeployed easily within organizations [36] [41] [43]. From outside organizations, we extended that Bourgeois and Singh (1983) noted the concept of potential slack available from the external environment [6], suggested that the potentiality is another important structural constrain.

Without an organization, three types of slack with higher potentiality include customer relation slack, social capital slack, and network capital slack. Customer relation slack refers to excess resources committed to an organization by specific relational customers, who are valued resources providing tangible benefits to an organization [43]. Social capital slack refers to excess resources committed to an organization by specific relational work groups, who are valued resources providing tangible benefits to an organization [13]. Network capital slack refers to excess resources committed to an organization by specific relational key suppliers, who are valued resources providing tangible benefits to an organization [14].

Within an organization, three types of slack with lower potentiality include financial, human resource, and operational slack. Financial slack refers the level of liquid assets, such as cash on hand, that is available to an organization [22]. Human resource slack refers to specialized and skilled human resources that are rare and absorbed [27]. Operational slack derives from unused or underutilized operational resources, such as excessive production capacity as it is absorbed and generally tied to a specific purpose within an organization [43].

Higher potential slack, relatively low cost of maintaining other resources, usually is accumulated form social relations resulting from organizational stakeholders and along unique historical trajectories [4], while higher excess absorbed resources may be viewed as costs [43]. Structural constrains limit the recovery of excess levels of organizational resources and the redeployment for other activities advancing synergy [27].

Proposition 2a: The positive effect of external and internal conditions on the likelihood of particular slack formation in the slack portfolio will be stronger at higher levels of the absorption and lower levels of the potentiality.

Proposition 2b: The negative effect of external and internal conditions on the likelihood of particular slack formation in the slack portfolio will be stronger at higher levels of the absorption and lower levels of the potentiality.

Strategic choice for organizational slack between transformation and maintenance

Theoretical and empirical evidence supported the effects of slack resources on organizational decision making and organizational responses to environmental threats and opportunities [7] [43], specific pressure to change for organizations. For example, Smith et al. (1991) argued that organizations with slack resources can afford sophisticated search activities and are more able to implement competitive responses [40]. According to slack resource theory [44], less profitable organizations, lacking financial slack, have fewer resources to spare for socially responsible activities, a kind of strategic actions, than more profitable ones. Similarly, Voss et al. (2008) emphasized that the associations between customer relation, operational, and human resource slack and production exploration are negative but financial slack, the associations between customer relations, operational, and human resource slack and production exploitation are positive but lack financial slack [43].

Although, organizations undergo internal and/or external conditions to change slack arrangements, still, a number of critical considerations exist in implementing this decision for them. Chen (1996) proposes three underpinning drivers of predicting competitive behavior: awareness, motivation, and capability [8]. Awareness may result from internal or/and external conditions organizations perceived. Voss et al. (2008) also suggested that organizations face the motivational constrain which determine how to interpret slack and its subsequent influence on organizational actions [43]. Therefore, if organizations have intent on changing slack from one type to another, they need both the capability and motivation to adjust those changes.

Proposition 3a: Organizations with both higher capability and higher motivation will have a propensity to transform one type of slack into other types.

Proposition 3b: Organizations with either lower capability or lower motivation will have a propensity to keep exiting different types of slack.

The performance decided by the gap between strategic choices and market expectations

Through expectations, organizations decide on their strategic actions and choose to keep the status quo or to change. Strategic factor market theory suggests that superior expectations are

necessary for organizations to appropriate gains from valuable resources excluding luck. Theorists hold that organizations cannot appropriate gains from the deployment of valuable resources unless they have superior expectations about their future value or unless they are beneficiaries of pure luck [3] [23]. Organizations allocate their resources according future expectations. Existing absorbed resources have particular use to efficiency; organizational slack can be used for change. Indeed, organizations may maintain a certain level of slack to wait a good opportunity as valuable resources. Voss et al. (2008) argue that slack can accrue as a result of organizational performance in periods, as a planned buffer, or as a result of poor planning [43]. In this sense, Rindova and Kotha (2001) introduced the concept of continuous morphing to describe the comprehensive ongoing transformations through which the focal firms sought to regenerate their transient competitive advantage [34].

Proposition 4: Whether the organization chooses maintenance or transformation for slack, high levels of organizational slack will result in high levels of performance when the decisions approach market expectations and low levels of performance when the decisions were far from market expectations.

Performance feedback to indigenous controls

Organizations will receive particular performance feedback under prior decision making. According the levels of the performance, organizations, including their internal coalitions, can adjust indigenous controls to fit market expectations in the next period, while exogenous contexts cannot be changed. For example, if the organization has acquired better performance, they perceived opportunities and intend to enlarge on the organizational boundary and maintain existing position and value proposition. Organizations that have successfully navigated some specific changes in the past are more likely to do so in the future; the accumulated slack derived from prior dominance may help them to successfully navigate the new one [19]. Differentiated between organizational performance and aspiration levels, organizations rearrange their varied discretionary resource allocation [16] [17].

Proposition 5: An organizational performance will affect the indigenous controls. Organizational performance in the prior period increases the likelihood that the organization will make its boundary broader, its internal system more stable, political behavior decreasing, risk attitude increasing, and keeping existing product position and value proposition in the present period.

CONCLUSIONS

In summary, the current study makes several contributions to the literature: First, we enhanced research on organizational slack by integrating contending theories of slack resources. Second,

this paper extended the organizational slack and strategic action literatures by linking external/internal conditions, strategic choice, and feedback loop into a comprehensive relationship structure. Third, the proposed framework completely explains the slack paradox in the dynamic organizational process by introducing slack system view.

Several issues remain. First, it is important to recognize that different types of slack within which organizational operations are not always statical. Responding to Miller and Tsang's (2011) fragmented theorizing [26], one of major obstacles to testing theories, dynamic slack theory is needed for advanced development. A second potential extension of the current research is to consider the issue of whether to focus on the existing slack per se or its strategic functions for the future. Finally, we note that slack resources can be dynamic in the sense that organizational slack can be changed from one form to another or maintain status quo over time. Thus, building on the methodology and measures in these advanced studies, future research could empirically test the propositions purpose herein via survey- or simulate-based approaches. In term of empirically testing the propositions advanced, an integrated approach drawing on quantitative and qualitative methods such as mixed methodology may allow researchers to test these propositions. The theory development for organizational slack contributes to the understanding of why some particular organizations continue to perform well following environmental change.

Overall, we hope future research will empirically test the relationships proposed in this paper as well as build on the different conditions presented in order to further our understanding of the nature of slack resources in organizational dynamic process. Additionally, empirically examining the value decision makers assign to usage will advance our understanding of what slack does with regarding to aiding or harming organizations. Our framework thus provides both a theory of decision making for organizational slack and a theory of potential sustained competitive advantage in conditions of environmental change.

REFERENCES

[1]Adler, P.S.& Kwon, S.-W. *Social capital: Prospects for a new concept.* **The Academy of Management Review,** 2002, 27(1), 17-40.

[2]Alvarez-Gil, M.J.,Berrone, P.,Husillos, F.J.& Lado, N. *Reverse logistics, stakeholders' influence, organizational slack, and managers' posture.* **Journal of Business Research,** 2007, 60(5), 463-473.

[3]Barney, J.B. *Strategic factor markets: Expectations, luck, and business strategy.* **Management Science,** 1986, 32(10), 1231-1241.

[4]Barney, J.B.& Hansen, M.H. *Trustworthiness as a source of competitive advantage.* **Strategic Management Journal,** 1994, 15(s), 175-190.

[5]Bourgeois, L.J. *On the measurement of organizational slack*. **Academy of Management Review,** 1981, 6(1), 29-39.

[6]Bourgeois, L.J.& Singh, J.V. 1983. *Organizational slack and political behavior among top management teams*. Paper presented at the Academy of Management Proceedings.

[7]Chattopadhyay, P.,Glick, W.H.& Huber, G.P. *Organizational actions in response to threats and opportunities*. **Academy of Management Journal,** 2001, 44(5), 937-955.

[8]Chen, M.-J. *Competitor analysis and interfirm rivalry: Toward a theoretical integration*. **The Academy of Management Review,** 1996, 21(1), 100-134.

[9]Cheng, J.L.C.& Kesner, I.F. *Organizational slack and response to environmental shifts: The impact of resource allocation patterns*. **Journal of Management,** 1997, 23(1), 1-18.

[10]Child, J. *Organizational structure, environment and performance: The role of strategic choice*. **Sociology,** 1972, 6(1), 1-22.

[11]Daniel, F.,Lohrke, F.T.,Fornaciari, C.J.& Turner, R.A. *Slack resources and firm performance: A meta-analysis*. **Journal of Business Research,** 2004, 57(6), 565-574.

[12]Dooley, R.S.& Fryxell, G.E. *Attaining decision quality and commitment from dissent: The moderating effects of loyalty and competence in strategic decision-making teams*. **Academy of Management Journal,** 1999, 42(4), 389-402.

[13]Dyer, J.H.& Chu, W. *The role of trustworthiness in reducing transaction costs and improving performance: Empirical evidence from the united states, japan, and korea*. **Organization Science,** 2003, 14(1), 57-68.

[14]Dyer, J.H.& Singh, H. *The relational view: Cooperative strategy and sources of interorganizational competitive advantage*. **Academy of Management Review,** 1998, 23(4), 660-679.

[15]Farjoun, M. *Beyond dualism: Stability and change as a duality*. **Academy of Management Review,** 2010, 35(2), 202-225.

[16]George, G. *Slack resources and the performance of privately held firms*. **Academy of Management Journal,** 2005, 48(4), 661-676.

[17]Greve, H.R. *A behavioral theory of R&D expenditures and innovations: Evidence from shipbuilding*. **Academy of Management Journal,** 2003, 46(6), 685-702.

[18]Greve, H.R. Organizational learning from performance feedback: A behavioral perspective on innovation and change. Cambridge, UK: Cambridge University Press, 2003.

[19]Hill, C.W.L.& Rothaermel, F.T. *The performance of incumbent firms in the face of radical technological innovation*. **Academy of Management Review,** 2003, 28(2), 257-274.

[20]Huy, Q.N. *Time, temporal capability, and planned change*. **Academy of Management Review,** 2001, 26(4), 601-623.

[21]Jensen, M.C.& Meckling, W.H. *Theory of the firm: Managerial behavior, agency costs and ownership structure*. **Journal of Financial Economics,** 1976, 3(4), 305-360.

[22]Kraatz, M.S.& Zajac, E.J. *How organizational resources affect strategic change and*

performance in turbulent environments: Theory and evidence. **Organization Science,** 2001, 12(5), 632-657.

[23]Makadok, R. *Toward a synthesis of the resource-based and dynamic-capability views of rent creation*. **Strategic Management Journal,** 2001, 22(5), 378-401.

[24]Miller, D.& Chen, M.J. *Sources and consequences of competitive inertia: A study of the U.S. Airline industry*. **Administrative Science Quarterly,** 1994, 39(1), 1-23.

[25]Miller, K.D.& Arikan, A.T. *Technology search investments: Evolutionary, option reasoning, and option pricing approaches*. **Strategic Management Journal,** 2004, 25(5), 473-485.

[26]Miller, K.D.& Tsang, E.W.K. *Testing management theories: Critical realist philosophy and research methods*. **Strategic Management Journal,** 2011, 32(2), 139-158.

[27]Mishina, Y.,Pollock, T.G.& Porac, J.F. *Are more resources always better for growth? Resource stickiness in market and product expansion*. **Strategic Management Journal,** 2004, 25(12), 1179-1197.

[28]Nahapiet, J.& Ghoshal, S. *Social capital, intellectual capital, and the organizational advantage*. **Academy of Management Review,** 1998, 23(2), 242-266.

[29]Nelson, R.R.& Winter, S.G. An evolutionary theory of economic change. Cambridge, MA: Harvard University Press, 1982.

[30]Nixon, R.D.,Hitt, M.A.,Lee, H.U.& Jeong, E. *Market reactions to announcements of corporate downsizing actions and implementation strategies*. **Strategic Management Journal,** 2004, 25(11), 1121-1129.

[31]Nohria, N.& Gulati, R. *Is slack good or bad for innovation?* **Academy of Management Journal,** 1996, 39(5), 1245-1264.

[32]Phan, P.H.& Hill, C.W.L. *Organizational restructuring and economic performance in leveraged buyouts: An ex post study*. **Academy of Management Journal,** 1995, 38(3), 704-739.

[33]Reuer, J.J.& Leiblein, M.T. *Downside risk implications of multinationality and international joint ventures*. **Academy of Management Journal,** 2000, 43(2), 203-214.

[34]Rindova, V.P.& Kotha, S. *Continuous "morphing": Competing through dynamic capabilities, form, and function*. **Academy of Management Journal,** 2001, 44(6), 1263-1280.

[35]Schumpeter, J.A. Capitalism, socialism and democracy. New York: Harper & Row, 1942.

[36]Sharfman, M.P.,Wolf, G.,Chase, R.B.& Tansik, D.A. *Antecedents of organizational slack*. **Academy of Management Review,** 1988, 13(4), 601-614.

[37]Simsek, Z.,Veiga, J.F.& Lubatkin, M.H. *The impact of managerial environmental perceptions on corporate entrepreneurship: Towards understanding discretionary slack's pivotal role*. **Journal of Management Studies,** 2007, 44(8), 1398-1424.

[38]Singh, J.V. *Performance, slack, and risk taking in organizational decision making*.

Academy of Management Journal, 1986, 29(3), 562-585.

[39]Sirmon, D.G.,Hitt, M.A.& Ireland, R.D. *Managing firm resources in dynamic environments to create value: Looking inside the black box.* **Academy of Management Review,** 2007, 32(1), 273-292.

[40]Smith, K.G.,Grimm, C.M.,Gannon, M.J.& Chen, M.-J. *Organizational information processing, competitive responses, and performance in the U.S. Domestic airline industry.* **Academy of Management Journal,** 1991, 34(1), 60-85.

[41]Tan, J.& Peng, M.W. *Organizational slack and firm performance during economic transitions: Two studies from an emerging economy.* **Strategic Management Journal,** 2003, 24(13), 1249-1263.

[42]Uhlenbruck, K.,Meyer, K.E.& Hitt, M.A. *Organizational transformation in transition economies: Resource-based and organizational learning perspectives.* **Journal of Management Studies,** 2003, 40(2), 257-282.

[43]Voss, G.B.,Sirdeshmukh, D.& Voss, Z.G. *The effects of slack resources and environmental threat on product exploration and exploitation.* **Academy of Management Journal,** 2008, 51(1), 147-164.

[44]Waddock, S.A.& Graves, S.B. *The corporate social performance-financial performance link.* **Strategic Management Journal,** 1997, 18(4), 303-319.

[45]Wernerfelt, B. *A resource-based view of the firm.* **Strategic Management Journal,** 1984, 5(2), 171-180.

DISSENTERS' IMAGES ON SUPERVISORS AND COLLEAGUES: THE CONTIGENCY PERSPECTIVE

Hsin-Hua Hsiung

Department of Business Administration, National Dong Hwa University

No. 1, Sec. 2, Da Hsueh Rd., Shoufeng, Hualien 97401, Taiwan, R.O.C.

hsiung@mail.ndhu.edu.tw, (886) 3863-3027

Nien-Tai Tsai

Department of Business Administration, National Dong Hwa University

No. 1, Sec. 2, Da Hsueh Rd., Shoufeng, Hualien 97401, Taiwan, R.O.C.

lothar49@ms21.hinet.net, (886) 3863-3017

Ching-Chou Chen

Department of Business Administration, National Dong Hwa University

No. 1, Sec. 2, Da Hsueh Rd., Shoufeng, Hualien 97401, Taiwan, R.O.C.

ching@mail.ndhu.edu.tw, (886) 3863-3017

ABSTRACT

Employees' dissent behaviors may play a constructive role for organizations. However, organizational members may not always appreciate these behaviors. The purpose of this study is to explore the impact of principled organizational dissent on personal image. We propose that the relationship between principled organizational dissent and personal image will be moderated by two contextual factors: interpersonal position and interest orientation. We contend that when an employee's informal social network is high, and when the raised issue would benefit the organization or all employees, principled organizational dissent is less likely to result in a negative image.

Keywords: Principled organizational dissent, Personal image, Informal social network

INTRODUCTION

Dissent is an important communication activity, and often appears in organizations. Organizational dissent takes place when employees disagree with organizational path about critical issues or when employees espouse significantly different positions on issues they feel

are important to themselves or the whole [39]. Comparing with silence and voice, dissent can be viewed as a kind of constructive aggressive communication[34].

The expression of dissent derives from an employee's dissatisfaction with the employing organization. Graham (1986) proposed a theoretical conception of principled organizational dissent (POD) and, attempted to explain the causes of dissent and explore the intrinsic psychological process of this behavior[14]. According to his definition, "Principled organizational dissent is the effort by individuals in the workplace to protest and/or to change the organization's status quo because of their conscientious objection to current policy or practice." [14, p.2] Similarly, Kassing (1997) conceptualized dissent as expressing disagreement or contradictory opinions about organizational practices, policies, and operations, but not necessarily violating legitimacy or societal norms [17].

Previous literature has diversified views on the consequences of dissent behavior. Dissent behavior is usually seen as having a negative influence in traditional organizations. The predominant view is that dissenters tend to argue in a verbally aggressive manner and are seen as troublemakers, and not accepted by peers [33]. However, dissenters can yield substantial benefits for an organization. Dissent has a constructive function in the organization as a barrier to "Groupthink" [17]; dissent may facilitate and enhance decision quality [10] [15]; dissent enables employees to increase their creativity and promotes organizational innovation [8] [30]. A number of studies have shown that minority dissent may foster team innovation [8], and enhance or facilitate decision-making quality and team effectiveness [9] [32] [36]. Therefore, Dissenters should be recognized and appreciated, as well as a protected and nurtured [38]. Although dissent has many benefits in organizations, why don't managers welcome dissent? One of important cause could be attributed to negative impression by people perceived.

Image is an important personal factor by which one influences others. If an individual has a good, positive image, then he/she will be appreciated. Similarly, if an individual has a professional image, then he/she can often exert influence on others with certain skills. These influences of power on others may be viewed as referent or expert power. Individuals, who have a good image, are more likely to have successful careers and power. In contrast, an individual will readily lose both of these advantages if he/she has a poor image. A main purpose of dissent may be to change the organizational status quo and ideas of decision makers. However, to achieve the intended purpose, may depend on the managers or colleagues who perceive the dissenters' behavior in the organization. By expressing disagreement and opposing opinions, dissenters' images may be damaged and their opinions may not be accepted, even though they might benefit the organization in the long run. Accordingly, image is very important to dissenters.

However, little research has been done on the effects of dissent on individual image and related contextual factors. We propose a theoretical model (Figure 1), which shows the moderating roles of contextual factors in the relationship between POD and personal images. The individual images we mentioned are based on the perceptions from supervisors and colleagues in workplaces. Through a literature review, we identified two moderators in the dissent-image relationship: the centrality of one's informal network (i.e., interpersonal position) and the raised issue would benefit the organization or all employees (i.e., beneficiary of the issue).

This paper has two main contributions: (1) To shed light on the relationship between the POD and people's personal images of the dissenter; and (2) To offer more insight into the moderators that affect the relationship between the POD and the personal images. Furthermore, this study promises to have practical implications for dissenters. According to our proposed model, it could enable dissenters to see the value of image management as well as suggest dissenters how to select a proper strategy to express their dissent—a strategy that could be accepted and/or supported by supervisors and colleagues.

FIGURE 1
Theoretical model
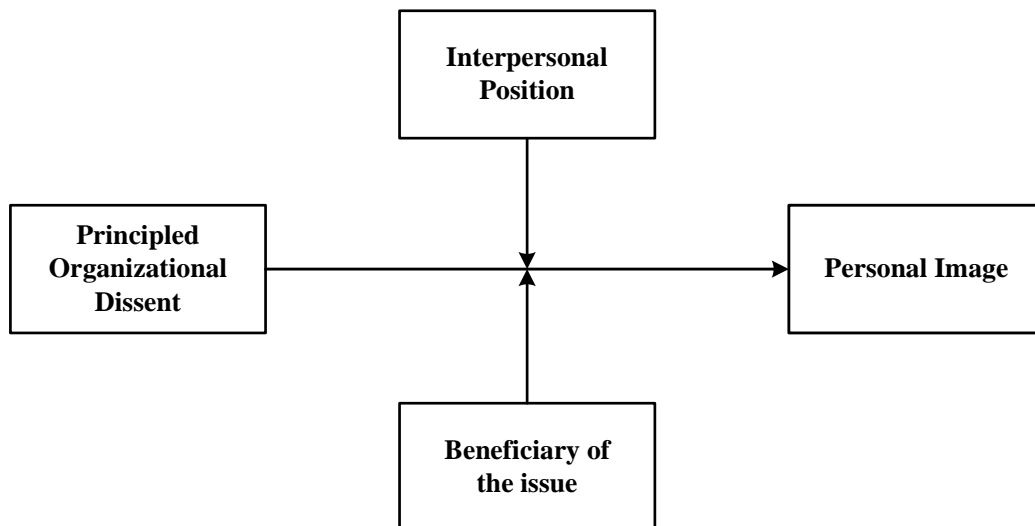
THEORY AND PROPOSITIONS

Principled organizational dissent and personal image

The definition of POD indicates that individuals in the workplace strive to protest and/or to change the organizational status quo because of their conscientious objection to current

policy or practices that might violate laws or ethics [14]. Dooley and Fryxell (1999) defined dissent as when employees openly express contradictory opinions about organizational practices, or express disagreement with the strategic decision-making process [10]. Expressing dissent can be viewed as adversarial behavior [18] and that the dissenters are often highly ethically motivated and want to contribute to the organization's wellbeing [38].

Personal image refers to the aggregate of the public's perceptions of the person. Dowling (1986) defined image as "specific viewpoints toward a certain matter through description, memory, or other ways of association with such matter [11]. This definition is an externally oriented public persona that is based on reflected appraisal rather than on one's self-image [16] [35]. It results from the interactions among people's impressions, existing beliefs, thoughts, and feelings on such a thing.

Image is an important factor in successful interpersonal relationships. Image will affect follow up action and decision on a target [11]. A positive image can give a person confidence in one's thoughts and actions; a negative image, in contrast, may make a person doubtful of one's capabilities and ideas. Additionally, a positive public image helps people achieve desirable social outcomes such as approval, friendship, and power [25]. Also, someone with a positive image often may readily access the resources of others [2]. Poor image can negatively affect appraisals from others and, thereby, impede the development of an individual careers [1]. Roberts (2005) suggests that image plays a major role in interactions with others, e.g., supervisors and colleagues, and can impact the quality of interpersonal relationships, as well as performance ratings [35].

Dissenters express disagreement or contradictory opinions, which can be perceived as negative feedback [3].Consequently, supervisors and other colleagues often recognize dissent as a threat [38]. The act of expressing dissent can be carried out in various ways, including (1) questioning management; (2) questioning workplace policies; (3) raising criticisms about organizational changes; (4) speaking with supervisors regarding disagreement with workplace decisions; and (6) telling management about perceived unfair treatment of employees [18]. Organizational leaders typically suppress free speech and impede employees' ability to offer criticisms and propose constructive alternative opinions, in order to show the harmony and unity within the organization and to demonstrate the power of leadership [38]. King (1999) points out: "Within a bureaucratic or centralized structure, top management may feel threatened or challenged by individuals who dislike an established policy or action within the organization" and may, therefore, prevent opposition [22]. Similarly, members are reluctant to change organizational practices and the status quo, because of the organization is operating successfully. When employees express dissent, members may feel uncomfortable and although it could actually benefit the organization, just the act of dissenting might be viewed as undesirable behavior. Compared to coworkers who do not actively dissent,

dissenters are more likely to be excluded by peers [31]. Therefore, dissent tends to be characterized as a threat or offensive behavior, and is negatively rated by most of their supervisors. This general reaction by supervisors tends to squelch much potential workplace dissent. Many employees say that they are reluctant to speak up about problems and concerns because they feel psychologically unsafe [29]. They fear that raising an issue may damage their image and lead to negative consequences, such as being labeled a "troublemaker" or having their job performance evaluations adversely affected [1] [12]. Klaas and DeNisi (1989) note that an employee's grievance activity may cause him/her to be labeled as a troublemaker and to receive lower performance ratings [23].

Interpersonal position and beneficiary of the issue

Indeed, many scholars have emphasized the importance of dissent in organizations. For example, the proper recognition and appreciation of dissent promotes constructive transformation within organizations [17] [38]. It may also break the tendency toward "Groupthink" and enhance the quality of decision making [10], as well as foster team performance and promote team innovation [8] [9]. Therefore, given the right organizational attitude and atmosphere, dissenters may not be excluded by others, but instead may be highly valued and respected by their supervisors and colleagues. In this section, we will explain our perspective in the aspects interpersonal position and beneficiary of the issue.

The first component of dissent is the person, the dissenter. Employees can express their dissent in various ways and toward multiple audiences. Possible audiences for a dissenting action include managers, supervisors, colleagues, and people outside the organization. Individual, relational, and organizational influences provide the basis for selecting a dissent strategy [17]. Employees considering dissent must assess the interactions of these factors in their selection process. It is critically important for them to consider how their expression of dissent will be perceived and how their working relationships will likely be affected [19] , such as, employees perceive that they have high-quality relationships with supervisors and who believe that their management is supportive of freedom of speech, since the dissent may influence decision-makers in the organization. [20] [21].

In terms of social network perspective, there are two structures within organization, formal structures and informal structures. Both structures play important roles in organizations. Workplace practices rely on task–oriented relationships that are formed through the formal structure of the organization, and on social relationships that are formed by informal structures and associations within the group. Social relationships focus on the individual's interpersonal relationships that foster feelings of personal obligation, gratitude, and trust [4]. These relationships may be characterized as "communal" [6], and the pattern of an employee's social network indicates that individual's social capital in the workplace [27]. In

our research we focus on informal social networks because these could influence formal reporting procedures, fomenting opposition to organizational change, driving the collective thought processes, actions, and reactions of its organizations [24].

Friendship networks and advice networks are important informal social networks [24]. Friendship networks are sets of linked friends who share information reciprocally based on individual affective needs and interpersonal trust. Advice networks relate to individuals who possess relevant expertise and who make themselves available to provide work-related advice for others in organizations. An individual's status within a social network is measured by network centrality [5]. Network centrality refers to the individual's degree of access to others within organizational networks. Individuals in central network positions (i.e., those with high network centrality) have greater potential control over resources and information, more opportunities to affect others, and greater capabilities of interaction with others [24] [26] [37]. Goodwin, Bowler, and Whittington (2009) suggest that centrality in an advice networks adds a relational component beyond the expert power base, and this aspect may provide individuals with more opportunities to help others resolve problems [13]. An individual's centrality in the advice network will enhance the relationship between the supervisor and that employee, from the supervisor's perspective [7] [13]. Similarly, individuals in a central position within friendship network will be trusted more by members within organizations. Lee et al. (2010) found that Individuals with higher friendship network centrality have more effective interaction with colleagues [26]. As noted by Kassing and Avtgis (1999), The higher the centrality of the employee, who expresses dissent, the less likely they are to suffer negative consequences. [21]. This reasoning forms the basis of our propositions:

Proposition 1a: From a supervisor's perspective, employee dissent will be less likely to be associated with a negative image when the dissenter's informal network centrality is high.

Proposition 1b: From a colleague's perspective, the act of dissenting will be less likely to be associated with a negative image when the dissenter's informal network centrality is high.

The second component of dissent is the principled issues. The value of dissent is that it can enhance the ability of organizations to address problems and benefit the entire group. The beliefs, attitudes, ideas, procedures, and perspectives of the all the members of a group tend to consistency when the organization is in a long-term stable state. Although this consistency may contribute to efficient communication and organizational harmony, it may also lead to rigidity. When an organization faces serious issues and the radical changes, if it does not have the means or inclination to respond appropriately, the final result could be the failure of the organization. Here dissent can play an important role in an organization by arousing the group's leadership to value the issues [28], elicit divergent thinking and enhance creativity

[8], and disrupt group-think [30] [38]. Therefore, when the dissenter expresses dissent, if the results of issues are a benefit for the organization, the dissenter is less likely to have a negative image on the supervisors.

Proposition 2a: From a supervisor's perspective, employee dissent will be associated with a less-negative image when the raised issue might benefit the organization.

Dissent is sometimes viewed as prohibitive and challenging behaviors [40]. Prohibitive behaviors are both protective and preventative. For instance, an employee may protest and raise objections in order to protect those with less power and to prevent improper and unethical behavior. However, the decision to raise dissent is a difficult one to make. Most Generally, employees keep silent because they weigh possible payoffs against likely costs, and factor in any positive feelings for their current supervisor [41]. Again, the most common "likely costs" that employees fear are gaining a negative label [1], and offending colleagues or supervisors and experiencing some sort of retaliation [29]. Despite these nagging fears, when organizational policy is unfair, or when the issue involves wrongdoing or the violation of a principle that could negatively affect the interests of employees, dissenters will often speak up and strive for the rights of employees. Nemeth and Goncalo (2010) note that dissenters who maintained their position are seen as courageous, may be viewed as heroes [30]. Consequently, when a dissenter expresses dissent, if the results of the issues promise to benefit the employees, then the dissenter is less likely to gain a negative image from colleagues. This can be summed up in our final proposition.

Proposition 2b: From a colleague's perspective, an employee's dissent will be associated with a less-negative image when the raised issue might benefit all employees.

DISCUSSION

Scholars tend to believe that dissenters may foster organizational prosperity, and encourage leaders to accept dissent; while practitioners find that dissent could undermine organizational harmony, and therefore act to suppress dissent. Previous research regarding the concept of dissent as a whole might lead to a different perspective. We argue that dissent encompasses the person and the issues, dissenter and dissent's issue, and should be viewed from an interactional perspective. In line with this premise, we propose four propositions; we suggest that the relationship of dissent and personal image may be moderated by an informal network of dissenters and results of issues. The theoretical contribution of the present study is that it provides important insights into the relationship between the dissent and the dissenting employee's image; that is; dissent might not cultivate a negative image on members in

organization, if the dissenter's informal network centrality is higher, or if the results of issues benefit the organization/ employees.

The managerial implication of our work is that (1) employees should deal with image management and should cultivate their communication skills; (2) in addition to professional competence, employees must give to help colleagues in the workplace, and maintain friendly relations with colleagues; (3) and where there are inconsistent viewpoints between leaders and employees, organizational leaders can create open environment that allows employees to express novel ideas and tolerate opposite opinions.

REFERENCES

[1]Ashford, S.J.,Rothbard, N.P.,Piderit, S.K.& Dutton, J.E. *Out on a limb: The role of context and impression management in selling gender-equity issues.* **Administrative Science Quarterly,** 1998, 43(1), 23-57.

[2]Ashford, S.J.& Tsui, A.S. *Self-regulation for managerial effectiveness: The role of active feedback seeking.* **Academy of Management Journal,** 1991, 34(2), 251-280.

[3]Bailey, K.D. *Towards unifying science: Applying concepts across disciplinary boundaries.* **Systems Research and Behavioral Science,** 2001, 18(1), 41-62.

[4]Blau, P. Exchange and power in social life. New York: John Wiley, 1964.

[5]Brass, D.J. *Being in the right place: A structural analysis of individual influence in an organization.* **Administrative Science Quarterly,** 1984, 29(4), 518-539.

[6]Clark, M.S.& Mills, J. *Interpersonal attraction in exchange and communal relationships.* **Journal of Personality and Social Psychology,** 1979, 37(1), 12-24.

[7]Claudia, C.C.& Chester, A.S. *Exploring work unit context and leader-member exchange: A multi-level perspective.* **Journal of Organizational Behavior,** 2000, 21(5), 487-511.

[8]De Dreu, C.K.W. *Team innovation and team effectiveness: The importance of minority dissent and reflexivity.* **European Journal of Work & Organizational Psychology,** 2002, 11(3), 285-298.

[9]De Dreu, C.K.W.& West, M.A. *Minority dissent and team innovation: The importance of participation in decision making.* **Journal of Applied Psychology,** 2001, 86(6), 1191-1201.

[10]Dooley, R.S.& Fryxell, G.E. *Attaining decision quality and commitment from dissent: The moderating effects of loyalty and competence in strategic decision-making teams.* **Academy of Management Journal,** 1999, 42(4), 389-402.

[11]Dowling, G.R. *Managing your corporate image.* **Industrial Marketing Management,** 1986, 15(2), 109-115.

[12]Dutton, J.E.,Ashford, S.J.,O'Neill, R.M.,Hayes, E.& Wierba, E.E. *Reading the wind: How middle managers assess the context for selling issues to top managers.* **Strategic Management Journal,** 1997, 18(5), 407–425.

[13]Goodwin, V.L.,Bowler, W.M.& Whittington, J.L. *A social network perspective on lmx relationships: Accounting for the instrumental value of leader and follower networks.* **Journal of Management,** 2009, 35(4), 954-980.

[14]Graham, J.W. *Principled organizational dissent: A theoretical essay.* **Research in Organizational Behavior,** 1986, 8, 1-52.

[15]Herbert, T.T.& Estes, R.W. *Improving executive decisions by formalizing dissent: The corporate devil's advocate.* **Academy of Management Review,** 1977, 2(4), 662-667.

[16]Ibarra, H. *Provisional selves: Experimenting with image and identity in professional adaptation.* **Administrative Science Quarterly,** 1999, 44(4), 764-791.

[17]Kassing, J.W. *Articulating, antagonizing, and displacing: A model of employee dissent.* **Communication Studies,** 1997, 48(4), 311-332.

[18]Kassing, J.W. *Development and validation of the organizational dissent scale.* **Management Communication Quarterly,** 1998, 12(2), 183-229.

[19]Kassing, J.W. *Investigating the relationship between superior-subordinate relationship quality and employee dissent.* **Communication Research Reports,** 2000, 17(1), 58-69.

[20]Kassing, J.W.& Armstrong, T.A. *Examining the association of job tenure, employment history, and organizational status with employee dissent.* **Communication Research Reports,** 2001, 18(3), 264- 273.

[21]Kassing, J.W.& Avtgis, T.A. *Examining the relationship between organizational dissent and aggressive communication.* **Management Communication Quarterly,** 1999, 13(1), 100-115.

[22]King, P.I. *The implications of an organization's structure on whistlblowing.* **Journal of Business Ethics,** 1999, 20(4), 315-326.

[23]Klaas, B.S.& DeNisi, A.S. *Managerial reactions to employee dissent: The impact of grievance activity on performance rating.* **Academy of Management Journal,** 1989, 32(4), 705-717.

[24]Krackhardt, D.& Hanson, J.R. *Informal networks: The company behind the charts.* **Harvard Business Review,** 1993, 71(4), 104-111.

[25]Leary, M.R.& Kowalski, R.M. *Impression management: A literature review and two-component model.* **Psychological Bulletin,** 1990, 107(1), 34-47.

[26]Lee, Y.H.,Yang, L.S.,Wan, K.M.& Chen, G.H. *Interactive effects of personality and friendship networks on contextual performance.* **Social Behavior & Personality: An International Journal,** 2010, 38(2), 197-208.

[27]Lin, N. Social capital: A theory of social structure and action. Cambridge, UK.: Cambridge University Press, 2001.

[28]McGuire, W.J. *The nature of attitudes and attitude change*. In G. Lindzey & E. Aronson (Eds.), The handbook of social psychology, 2 ed., Vol. 3: 136-314. Reading, MA: Addison-Wesley, 1969.

[29]Milliken, F.J.,Morrison, E.W.& Hewlin, P.F. *An exploratory study of employee silence: Issues that employees don't communicate upward and why*. **Journal of Management Studies,** 2003, 40(6), 1453-1476.

[30]Nemeth, C.J.& Goncalo, J.A. *Rogues and heroes: Finding value in dissent*. In J. Jetten & M. Hornsey (Eds.), Rebels in groups: Dissent, deviance, difference and defiance: 17-30. Oxford: Wiley-Blackwell Publishing Co, 2010.

[31]Nemeth, C.J.& Staw, B.M. *The tradeoffs of social control and innovation in small groups and organizations*. In L. Berkowitz (Ed.), Advances in experimental social psychology, Vol. 22: 175-210. New York: Academic Press, 1989.

[32]Park, G.& DeShon, R.P. *A multilevel model of minority opinion expression and team decision-making effectiveness*. **Journal of Applied Psychology,** 2010, 95(5), 824-833.

[33]Phillips, K.W.& Loyd, D.L. *When surface and deep-level diversity collide: The effects on dissenting group members*. **Organizational Behavior and Human Decision Processes,** 2006, 99, 143-160.

[34]Rancer, A.S.& Avtgis, T.A. Argumentative and aggressive communication: Theory, research, and application. Thousand Oaks, CA: SAGE, 2006.

[35]Roberts, L.M. *Changing faces: Professional image construction in diverse organizational settings*. **Academy of Management Review,** 2005, 30(4), 685-711.

[36]Schulz-Hardt, S.,Brodbeck, F.C.,Mojzisch, A.,Kerschreiter, R.& Frey, D. *Group decision making in hidden profile situations: Dissent as a facilitator for decision quality*. **Journal of Personality and Social Psychology,** 2006, 91(6), 1080-1093.

[37]Settoon, R.P.& Mossholder, K.W. *Relationship quality and relationship context as antecedents of person- and task-focused interpersonal citizenship behavior*. **Journal of Applied Psychology,** 2002, 87(2), 255-267.

[38]Shahinpoor, N.& Matt, B.F. *The power of one: Dissent and organizational life*. **Journal of Business Ethics,** 2007, 74(1), 37-48.

[39]Sydow, J.,Schreyogg, G.& Koch, J. *Organizational path dependence: Opening the black box*. **Academy of Management Review,** 2009, 34(4), 689-709.

[40]Van Dyne, L.& LePine, J.A. *Helping and voice extra-role behaviors: Evidence of construct and predictive validity*. **Academy of Management Journal,** 1998, 41(1), 108-119.

[41]Withey, M.J.& Cooper, W.H. *Predicting exit, voice, loyalty, and neglect*. **Administrative Science Quarterly,** 1989, 34(4), 521-539.

THE HIDDEN COST OF VERTICAL INTEGRATION: EVIDENCE IN DISPLAY INDUSTRY

Dr. Yuh-Yuan Tsai, Professor, National Dong Hwa University, Taiwan, 03-8633023
email: yytsai@mail.ndhu.edu.tw

ABSTRACT

Vertical integration, defined as a high degree of internal transfers of goods or services, can reduce costs because of the economies firms achieve from avoiding transaction costs and market exchanges, exploiting opportunities for coordinating internal activities, and creating power over buyers and suppliers (D'Aveni and Ravenscraft, 1994). The theoretical foundations for vertical integration are based primarily on transaction cost. Mahoney (1992) points out six key factors influencing governance choice such as frequency, asset specificity, demand uncertainty, technological uncertainty, task programmability, and non-separability and emphasizes that vertical integration can reduce asymmetric uncertainty, shirking and appropriate R&D spillover. Although there is a lot of vertical integration in the supply chain from panel makers to LCD components, in key areas of the Display industry, there are still several major firms adopting different strategies other than vertical integration. It is interesting to find out why they did the other way.

INTRODUCTION

In Display industry, five panel firms now dominate revenues of worldwide TFT-LCD. Concentration has been increasing. With the Chimei-Innolux merger, the top five panel firms such as Samsung, LG Display, AUO, CMO, Innolux account for more than 90% of revenues. However, only the top three firms make profits according to "the rule of three" of industry characteristics. There is a great deal of vertical integration in the LCD-TV supply chain from panel makers vertically integrating LCD-module assembly. LCD-TV set makers have started to invest in LCD-module assembly lines, mostly working with panel makers, and in many cases have formed joint ventures. For Example: AUO and Changhong jointly invested in an LCD-module line; LG Display formed an alliance with AmTRAN to set up an LCD-module and LCD-TV-set   assembly line and set up a joint venture with TPV for LCD-module lines; AUO and TPV are jointly establishing an LCD-module line in Eastern Europe. For CPT or even AUO, it is now in a situation of turning point about continuing business or make a turnaround due to the difficulties in profiting. This article focuses on the specific issues about whether to vertically integrate in their supply chain management.

LITERATURE REVIEW

Theory of vertical integration: The earliest and most common applications of TCE focus on the vertical integration decision. These studies typically focus on a manufacturing firm's decision to backward integrate into the supply of materials or components or forward integrate into distribution and sales. Balakrishinan and Wernerfelt (1986) summarize six theories for vertical integration, such as economics of integration, competitive considerations, production economies of integration, transactional economies in advantages of hierarchies, transactional economies in limits to corporate span, and investments in specialized assets and technological change. An exhaustive review of the economic and strategy literature suggests that the motives for vertical integration may be classified into four major categories: (1) transaction costs considerations; (2) strategic considerations : (3) output and/or input price advantages; and (4) uncertainties in costs and/or prices (Mahoney, 1992). Vertical integration strategies are really a combination of decisions regarding weather the firm should provide goods and services in-house through its own business unites, or purchase them from

outsiders, instead. These decisions include how much of a particular good or service in-house or sell to outsiders, and how far backward or forward in a vertical chin of activities to integrate. For each production stage in the chain of transformations that they engage in－from ultra-raw materials to the ultimate customer－firms also decide on their breath of activities and their ownership form. That is, how much equity to hold in each vertically related business unit. Firms choose combinations of these dimensions that they believe are the most appropriate means to achieve their strategies needs (Harrigan, 1986).

The boundary of firm: Regarding the boundary and theory of the firm, Coase (1937) maintains the cost of using price mechanism. Williamson (1975) defines this construct in operational level using transaction cost as comparative assessment to choose the governance structure (e.g., market, hybrid and hierarchy). In particular, discrete structural forms need to be identified that have differential efficiencies, and the observable dimensions of transaction costs need to line up with various governance structures in discriminating way (Williamson, 1991).
Teece (2007) indicates 4 key elements to set enterprise boundaries ensuring to gain the benefit of innovation: (1) the appropriability regime; (2) the nature of the complementary assets that an innovating enterprise possessed; (3) the relative positioning of innovator and potential imitators with respect to complementary assets; and (4) the phase of industry development.

Economies of Integration: Economic benefits from vertical integration arise when internalization overcomes three transaction difficulties associated with market exchange. First, internalization allows the firm to invest in specialized assets which result in the production of goods and services at a lower economic cost then when non-specialized assets are used. In market exchange the prospect of opportunism in small numbers trading relationships, or the existence of information impactedness, deter firms from making such specific investments with other firms; internalization overcomes this difficulty. Second, internalization leads to economic benefits because it removes the possibility of resource misallocation due to information impactedness. Third, internalization produces economic benefits because it obviates the need to write complex contracts between the various parts of the business. These benefits arise because the resources which would have had to be expended in the essentially non-productive activity of negotiating, monitoring and enforcing exchange can be used to produce goods and services.

Incentive to vertical integration: D'Aveni and Ravenscraft (1994) indicates the incentive to vertical integration depends on the type of production involves, the extent of transaction costs, the amount of specialized assets, the degree of market power at each stage of production, the separability of activities, the amount of uncertainty concerning prices and costs.

Cost reduction by vertical integration: Costs may be decreased by avoiding market costs, by eliminating the distortion in input costs caused by imperfect competition in the upstream market, by reducing transaction costs, by decreasing uncertainty or asymmetric information, resulting in a more efficient use of inputs, and by protecting proprietary. Vertical integration can also increase profits through higher prices by creating barriers to entry, allowing price discrimination, reducing service and advertising externalities, or providing a firm with power over buyers or suppliers (D'Aveni and Ravenscraft,1994).

Creating power over buyers and suppliers: A high degree of backward vertical integration through internal transfers can create economies by ensuring adequate suppliers, reducing cost distortions from monopolized inputs, and providing access to information about input costs and manufacturing processes. Ensuring supplies reduces costs by eliminating shortages that temporarily increase input costs and by avoiding unused capacity in the event of a shortage (D'Aveni and Ravenscraft, 1994).

Environment control: Firms vertically integrate to gain greater control over their economic environment. Backward integration can be used to endure the supply of scarce raw materials, while

forward integration ensures market access through limited distribution channels. By assuming greater control over suppliers or market access, the firm reduces the probability of suppliers or other channel members exerting monopolistic power and charging excessive prices for their products or services. Thus vertical integration can reduce both risk and cost when channel members or suppliers can become powerful.

Strategic advantage: The control gained through vertical integration can be used to achieve a strategic advantage over competition. By controlling the supply of key raw materials or a distribution channel to a market, a firm creates a strategic advantage over competition. This advantage enables the firm to either differentiate its offerings or charge lower prices than the competition and achieve superior performance.

Forces affecting the use of vertical integration: Harrigan (1986) points out four major categories of forces used to predict vertical integration strategies: demand and infrastructure uncertainty, industry volatility, bargaining power, and corporate strategy needs.

By selecting a particular governance structure, management aims top minimize the sum of production and transaction costs. Based on organizational economics theories of vertical integration, Lajili, Madunic and Mahoney (2007) develop 10 propositions for vertical integration. In the cases of high frequencies of transacting, high level of asset specificity, higher demand uncertainty, high task programmability, high non-separability problem, vertical integration is a more likely governance choice. For instance:

- Vertical integration is a more likely governance choice when there is a high frequency of transaction
- Vertical integration is more likely governance choice when there is a high level of asset specificity which locks trading partners into a small numbers trading situation that may make contracting hazardous due to potential haggling costs and hold-up problems.
- Vertical integration is more likely governance choice when there is higher demand (volume) uncertainty, when makes contracting more hazardous (under conditions of asset specificity).
- Vertical integration is more likely governance choice when there is low uncertainty about the timing of the obsolescence of specific assets since this condition will allow greater investment in specific assets, which increase the likelihood of vertical integration.
- Vertical integration is more likely governance choice when there is increased complexity, which necessitates a higher degree of complex firm-specific language and routines.

- Vertical integration is more likely governance choice when there is higher degree of difficulty in ascertaining quality of a differentiated product by inspection, which suggests that monitoring of inputs is required.

THE DISPLAY INDUSTRY

The Display Market has been growing over the past decades due to the dramatically expanding notebook-PC market, desktop-PC monitors as well as televisions. Accordingly, the advantages of TFT-LCD technology over CRT technology are with improvements in volume, weight, screen sizes, resolutions and image quality.

The challenge was in bringing the price down to an acceptable level, which was achieved through massive investments in successive generations of TFT-LCD manufacturing and in the material and equipment that round out the supply chain. These investments led to significant declines in price, as well as dramatic increases in screen sizes. With notebook PCs and desktop monitors completely penetrated by TFT-LCDs, much of the growth in recent years has come from the TV market. This high penetration, combined with increasing pressure on prices, will lead to slower revenue growth, particularly for TFT-LCDs.

In key areas of the display industry, there is an ongoing trend toward increase consolidation. In the large-area TFT-LCD market, five panel makers now dominate revenues. This has come about through aggressive investments and mergers. There are now two large TFT-LCD manufacturers in Korea (Samsung and LG Display), two in Taiwan (Chimei Innolux and AUO), and two in Japan (Sharp and Panasonic, which owns IPS Alpha). Japanese panel makers, traditionally the leaders in small and medium displays, are scaling back, as exemplified by Epson, Hitachi, NEC, and Toshiba.

Historically, growth on the large-sized TFT-LCD-panel market has been driven more by "supply push" than by "demand pull". The investments of suppliers in new-generation fabs have been a much greater factor in fueling expansion than end-user demand.

Samsung Electronics: Samsung Electronics is an electronics Samson. One of the world's largest semiconductor manufacturers, Samsung Electronics is also South Korea's top electronics company. It makes many kinds of consumer devices, including DVD players, big-screen television sets, and digital still cameras; computers, color monitors, LCD panels, and printers; semiconductors such as DRAMs, static RAMs, flash memory, display drivers, and MP3 player chips; and communications devices ranging from wireless phones to networking switches. The company, which is the flagship member of Samsung Group, also makes microwave ovens, refrigerators, air conditioners, and washing machines.

While many Asian-based companies faltered, Samsung continued to forge ahead. The company established U.S.-based subsidiary Alpha Processor Inc. to oversee sales and marketing for its 64-bit Alpha processor product line. The firm also secured the top position in the TFT-LCD global market by capturing 18 percent of the market. In August of 1998, the firm developed a flat-screen television. Despite the trying economic times, Samsung recorded a greater than eight percent increase in gross sales. In January 1999, *Forbes Global Business & Finance* recognized the firm as the world's premiere consumer goods and services company.

Chimei Innolux: Chimei Innolux was formed in the 2010 merger of Chi Mei Optoelectronics (CMO), Innolux Display, and TPO Displays. The company makes LCD panels for a variety of applications, including LCD television sets, notebook computers, desktop monitors for PCs, and displays for mobile devices and medical equipment, among other uses. Chimei Innolux also makes full LCD modules for easier integration into finished products. The LCD panel market is dominated by Asian manufacturers. The formation of Chimei Innolux creates the biggest LCD panel maker in Taiwan, surpassing AU Optronics; the company is expected to provide greater competition to the Korean leaders of the market, Samsung Electronics and LG Display.

LG Display: The world is truly flat for LG Display -- as in flat-panel displays. The company is one of the world's top producers of TFT-LCDs (thin-film transistor liquid-crystal displays), the svelte screens that go into laptop and notebook computers, desktop PC monitors, TV sets, wireless handsets, and a variety of applications in automotive navigation, avionics, consumer electronics, instrumentation, and medical equipment. LG Electronics and Philips merged their LCD businesses in 1999. Philips no longer holds any equity in LG Display. The company gets about 70% of its sales from customers in the Asia/Pacific region.

AUO: If you use a laptop or a wireless handset, you could be looking at AU Optronics (AUO). The company is a leading manufacturer of flat-panel displays. TFT-LCD (thin-film transistor liquid-crystal display) panels provide a clear, sharp image -- even in poor lighting -- and are used in car navigation systems, digital still cameras, laptop and notebook computers, television sets, and wireless phones. AUO sells to contract manufacturers, such as Compal, and to OEMs, including Acer, Audiovox, Panasonic, and Sony. The company sells primarily to customers in Asia; 70% of sales goes to customers in China and Taiwan. AU Optronics was formed by the merger of Acer Display Technology and Unipac Optoelectronics.

CPT: Chunghwa Picture Tubes (CPT) makes optoelectronic display components used in TFT-LCD (thin-film transistor liquid-crystal display) and CRT (cathode ray tube) applications. The company makes large, medium, and small display panels used in LCD televisions, monitors, notebooks, netbooks, and other products such as smartphones. Samsung Electronics and Sony are among its largest customers. CPT also partners with companies such as PC maker Compal Electronics. Chunghwa Picture Tubes, which was established in 1971, also makes color picture tubes and electron guns. The company has plants in China, Malaysia, and Taiwan (answers.com, 2011)

DILEMMA OF DISPLAY INDUSTRY

Take TV market for instance, the rewards for LCD-panel makers may be very great, but also are the risks. The major risk is financial, the minimum price of entry into the TV makers for LCD-panel makers ranging from $3 to $5 billion, which is what it costs to build a seventh-, eighth-, or ninth-generation fab. This compares to $1-2 billion for fourth- and fifth-generation fabs. The next highest risk is that of competition. previously, the competition was relatively straightforward for LCD-panel makers. In the notebook market, TFT-LCD-panel suppliers only had to compete with lower-priced and lower-performance STN-LCDs. In the monitor market, TFT-LCD makers had to contend with only lower-priced and more-well-established CRTs. While in the TV market, LCD s must compete with a variety of different technologies. The third risk involves dealing with the complexity and costs of television. In the notebook and monitor markets, panel suppliers had to deal with only two or three mainstream sizes. But in television, there are many display sizes, ranging from 10.4 in. to greater than 50 in. The proliferation of sizes, and the requirement to support them, translates into higher costs in maintaining a competitive product line. The final risk is profitability. Mobile-PC users are somewhat sensitive to price; however, monitor buyers are more sensitive. With price as the overriding concern, along with intense competition, the dream of TFT-LCD makers of attaining higher margins in the TV market may be difficult to achieve. Chances are that higher margin may not be realized until after the dust settles and the competition shakes out. The price trend and market share for major makers are shown as the follow Fig. 1 and Fig. 2(Information Display 5/10, 2010) :
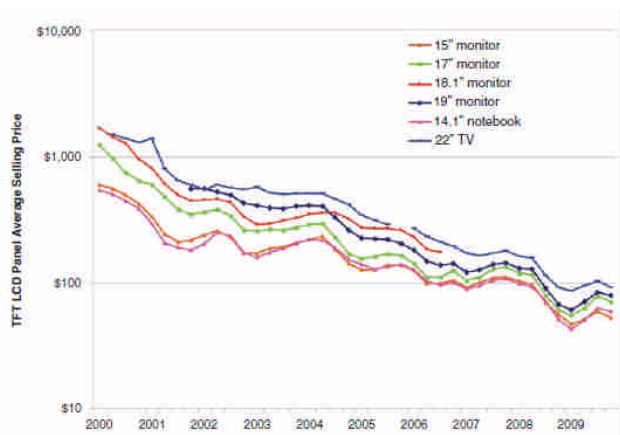
Fig. 1: The prices of TFT-LCD panels fell by a factor of 10 during 2000s
Source: Information Display 5/10, 2010
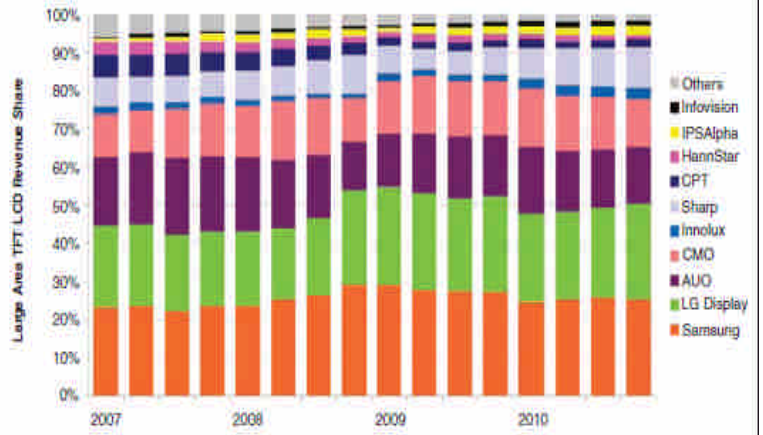
Fig. 2: Concentration in large-area TFT-LCD revenues has increasing. The top five accounts for 90% of revenues
Source: Information Display 5/10, 2010

THE CASES OF CPT

If the market share or the firm size is not in the rank of number one or two, the firm will be in a very vulnerable situation in profiting. Even the number 3 is probably in the margin of loss. The is a very difficult situation for firms whose ranking are 4 or 5, such as AUO or CPT. How can they survive in this industry will be a tough issue for they top management team. To build, to hold or to

divest will wonder CEO if they are already in loss for years. Unless they will exit, the following countermeasures would provide some directions for them to stay in this industry. In addition, in the case of vertical integration, there is still hidden cost which should be carefully considered.

The hidden cost of vertical integration: Take Drive IC or Control IC for Instance, It is a critical component for each CPT product and the demand is of large amount. According the theory of vertical integration, CPT is most likely to manufacture it internally. However, due to the following reasons, CYT eventually prefers joint venture or long-term contracting with the Drive IC or Control IC Design House: (1) organization culture is different between manufacturing and design house; (2) The core technology is totally different; (3) the mindset of employees between two firms is opposite; (4) most employees in design house are more likely geniuses and would not follow HR formality: (5) the traditional reward system is different; (6) large number of design houses and easy to find suppliers; (7) if internalized into value chain, compatibility will become an issue.

Joint venture instead of vertical integration: Although the benefits of vertical integration are clear as above, they often get eaten away by cost which managers concern more. Nevertheless, AUO has carried out a joint venture with Raydium Semiconductor which is a Design House specializing in the research, development, design, and manufacturer of Flat Panel Driver IC and controller IC. CPT tries to apply the similar model with Sitronix instead of merger. These cases are different from the theory of vertical integration which promote merger. The factors which challenge vertical integration will be studied by case study and propositions will be raised according to the categories of technology, competence focus, and strategic consideration.

Reducing manufacturing costs: Much of the cost reduction over the past decade has been driven by investment in TFT-LCD factories. Larger substrates, higher throughput equipment, and the learning that comes from cumulative production have driven down the cost of producing the LCD, particularly the backlight which can account for 25% of manufacturing cost in large panel.

Technology development: The plant facility and equipment of large size panel needs large investment and its production technology changes very fast generation by generation. It would not be the arena for number 4 in ranking as CPT. However, if CPT could develop new technology either in production or product using the existing facility for potential new product, it will create its new advantage,

New product development: CPT may search for new demands for even touch panel such as mobile phone in addition to PC, NB and TV. If CPT could not keep abreast with new generation of TFT-LCD factories, try to find new product for existing production line.

Searching niche market: Marketwise thinking, try to find a niche market which Samsung Electronics, Chimei Innolux and LG Display could not cover or ignore.

CONCLUSION

The rule of three reveals that only top three in industry can win. The Display Industry is just the case which only Samsung Electronics, Chimei Innolux, LG Display make profit and the others are all in a tough situation of loss. If these firms would stay in business, the alternatives are reducing manufacturing costs, technology development, new product development, searching niche market and avoid hidden cost of vertical integration. Most TFT-LCD firms try to integrate backward or forward their suppliers or buyers and build a comprehensive value chain. According to theory of vertical integration, Driver IC or Control IC is compliance with this hypothesis. However, CPT and AUO use the other strategies of joint venture or long-term contracting with their design houses. The reasons are culture and mindset difference in manufacturing and R&D, the employee and reward system are incongruent in both firms. If internalized into value chain, compatibility will become an issue and it will cause a lot of hidden cost for vertical integration.

THE INTERACTION BETWEEN RELATIONAL CAPITAL AND BOUNDARIES INTEGRATION IN STRATEGIC ALLIANCE

Hou, Jia-Jeng, National Chiayi University, (886) 5271-7160, jjhou@mail.ncyu.edu.tw

*Yung-Chuan Lin, National Chiayi University, (886) 5273-2824, chyuan17@gmail.com

Yue-Nan Lin, National Chiayi University, (886) 5273-2824, nlinyn1003@gmail.com

ABSTRACT

In this study, the author discussed the possible scenarios that would developed by the firms in a strategic alliance from the resource -base prospect. By extent the transaction cost theory into the concept of relational capital, the author inferred the modulation structure of relational capital in the boundaries integration, and embedded the relational capital concept into the strategic management processes frame model. As the value of the relational capital was highly respect in the field of boundary integration, It was defined as the close interactions and trust relation within the organization. It leads to a proper prospect of the colleagues and the mutual beneficial behaviors. Relationship is a strategic capital and the processes to rule the relationship constitutes the relationship asset management. In the market, it means that by strengthening the relational capital in the integration boundaries, it would enforce better sale conditions, and achieve the performance targets. In order to improve the performance and the maintenance of the quality of distributions, besides the dependence of resources from vendors, it was recommended to operate in different business strategies to cope with the high performance goals, the uncertain market demands, and to promote the competitive advantage. While enterprises of different business types had different resource in hands, the implementation of each strategy has unique difficulty to conquer. When choosing different strategies to execute, the difficulties each strategic plan faced and the cost paid should be weighted, and the benefits by strategy implementation and the consistency of the strategy and the corporate values should also be considered to approach the cooperate vision. By application of the concepts of the relational capital and integration boundaries, it enables the enterprises to access the resources more effectively. And through the inter-organizational resources integration, it enforces the competitive advantage.

In this study, by way of clarifying the structure of relational capital in the integration boundaries, the author tried to find out the best business model, inferred the position and rank of the relations model in the management programs which enhanced the competitiveness and developed some propositions.

Keywords: Strategic alliance, Relational capital, Integration boundaries, Propositions

*Corresponding Author

INTRODUCTION

The exploration of new techniques was a key factor to constitute the competitive advantage in most of the industries. In the past, most of the corporates preferred to develop new techniques by them self（Mowery, 1983）.They may get the origins of the techniques from outer resources（Hippel, 1982; Bozeman and Link, 1983; Pavitt,1986）, the competitors, the suppliers, or other organizations by contracts or franchising. Schumpeter（1975:83）argued that the induction of new techniques to an old industry frequently induced unexpected changes and he recognized the adaptation of techniques between industries a form of "Creative destruction". It means that after induction of new techniques, the original processes needs lots of transformation. In the traditional strategy management processes, Hitt（2001）had categorized two analytic models: "industrial organization model" which favored in highly potential industries, recommended to applied proper strategies to obtain sustainable competitive advantage; and "resource based mode", which explored the resources and ability from inside the corporate to set the foundation of the competitive advantage（Aaker, 1989）. In either way, when the corporate earned more economic value than her competitors, she got the competitive advantage. The economic value is based on the differences from perceived benefits of the customers and the cost to produce that products or services（Barney, 2006）.

While discussion the measurement of economic values or the possession of competitive advantage, one of the key points is the interaction between the resource elements and other corporates during the strategy formation. Pisano（1990）, in his documents-"The R&D Boundaries Of The Firm: An Empirical Analysis", had used the transaction cost theory to explain this claim and suggested to get part of the resources by contracts. The transaction cost theory had been applied in the practical analysis of products and supply chain（Monteverde and teece, 1982a, 1982b; Stuckey, 1983; Masten, 1984; Walker and Weber, 1984; Balakrishnan and Wernerfelt, 1986; Joskow, 1987）, marketing（Anderson and Schmittlein, 1984）, while in the field of boundary integration, it still remained some constrains（Williamson, 1975）, Teece and Armour（1977） and Teece（1988）had argued that the issues of boundary integration research were more important than that issues in marketing under the frameworks of the transaction cost theory in the bureaucratic organizations. Besides , most researches aimed at the decisions analysis between "voluntarily manufacture" and "outsourcing" were based on cases analysis and latent groups data analysis（Monteverde and Teece, 1982a; Masten, 1984;Walker and Weber, 1984）; only little discussed the inter-organization strategic relations.

In this study, the authors discussed the possible scenarios that would developed by the firms in a strategic alliance from the resource -base prospect. By extent the transaction cost theory(Tichy, Tushman, & Fombrun, 1979; Granovetter, 1985; Dyer and Singh, 1998) into the concept of relational capital, the author inferred the modulation structure of relational capital in the boundaries integration, and embedded the relational capital concept into the strategic management processes frame model published by Barney（2006）.

As the value of the relational capital was highly respect in the field of boundary integration （Dyer & Singh, 1988; Ahuja, 2000; Dyer & Nobeoka, 2000; Kale, Singh & Perlmutter, 2000）, It

2

was defined as the close interactions and trust relation within the organization. It leads to a proper prospect of the colleagues and the mutual beneficial behaviors（Thorelli, 1986; Kale, Singh & Perlmutter, 2000）. Richardson, Vidaurreta, Gorman（2002）had suggested that relationship is a strategic capital and the processes to rule the relationship constitutes the relationship asset management.

In the market, it means that by strengthening the relational capital in the integration boundaries, it would enforce better sale conditions, and achieve the performance targets. In order to improve the performance and the maintenance of the quality of distributions, besides the dependence of resources from vendors, it was recommended to operate in different business strategies to cope with the high performance goals, the uncertain market demands, and to promote the competitive advantage.

In this study, by way of clarifying the structure of relational capital in the integration boundaries, the authors tried to find out the best business model, inferred the position and rank of the relations model in the management programs which enhanced the competitiveness and developed some propositions.
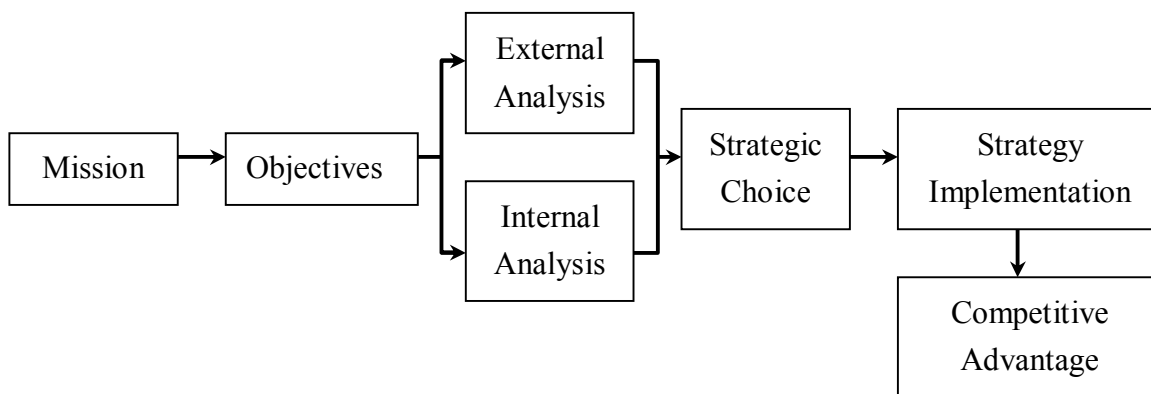
Fig 1：strategic management processes
Source:Jay B Barney & William S. Hesterly, "Strategic Management and Comoetitive Advantage", p.5, 2006。

MOTIVATION OF CREATING RELATIONSHIP CAPITAL

The concept of relationship capital is derived from the social capital. Nahapiet & Ghoshal (1998) categorized the social capital into structural, cognitive, and relational dimensions. Moreover, the dimension"relationship"includes four sub-dimensions—trust and trustworthiness, norms and sanctions, obligations and expectations, and identity and identification. On the one hand, the social network of relationship capital is composed of many links to form the "homophily;" on the other hand, it also can create complementarity to achieve synergy. Alliance of "diversification" emphasizes the relationship between a variety of strong and weak cooperations, heterogeneous contents, or degree of resource abundance. This alliance includes expectation, trust, and reciprocity in the social network（Tichy, Tushman, & Fombrun, 1979）. Especially, the trust can be used as an embedded mechanism in the relationship management（Uzzi, 1996）.

3

Granovetter (1985) argued that when this embedded relationship and social network link, it seems like an inter-organization form to become cross-boundary. However, the past related research showed that it is more useful for individuals on the relationship in an organization (Jacobs, 1961; Loury, 1977; Burt, 1992;1997; Tsai & Ghoshal, 1988). Relationship capital can enhance cooperative opportunities, help firms to highly obtain new competitive capability, and save related resources and diversify risk for organization. The relationship has played an important role on producing position and reputation in the supply chain.

Additionally, the relationship benefits in the information, technology, knowledge, and other resources for manufacturing flow path, and then generate the necessary power and control because the higher trust between both parties can be more confident for trading; thus, the necessary defensive design is relatively less (Williamson, 1991; Shapiro, Sheppard, and Chersakm, 1992).

Proposition 1: The higher degree the relationship capital, the higher degree will be the possibility of integration boundaries.

STRATEGIC ALLIANCES AND INTEGRATION BOUNDARIES

Strategic alliance is two or more independent organizations to have a cooperative relationship in the research & development, manufacturing, and selling products or services (Barney, 2006). For analyzing types of strategic alliance, the value chain is divided into vertical and horizontal alliances by some scholars (e.g., Porter & Fuller, 1986; Nueno & Oosterveld, 1988; Hitt, Ireland, & Hoskisson, 1997; Gulati, 1998;). Seetoo (1996) classified the strategic alliances of domestic SMEs into three categories; namely, vertical alliances, horizontal alliances, and asymmetric alliance. Vertical alliances focus on the combination of complementary resources, horizontal alliances pay attention to economies of scale, while the use of asymmetric alliances combine the power of large and small enterprises together to block access to their competitors. Not only combining the homogeneous resource can enhance value of resources, but also integrating heterogeneous resources can be a complementary strategy. Further, the value of complementary vertical, horizontal, and horizontal-strengthening alliances can be obtained when combining with the alliance activities such as united marketing（Huang & Yang, 2002）.

Barney (2006) classified strategic alliances into three categories: non-equity alliances, equity alliances, and joint venture (Figure 2).
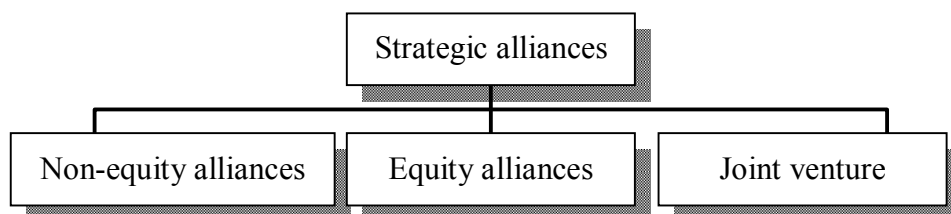

Figure 2: Types of Strategic Alliance
Source: Strategic Management and Competitive Advantage (Barney & Hesterly, 2006, p279)

Non-equity alliances do not hold shares in each other between the two sides (organizations),

also it is not an independent organization of co-management and cooperation but on a contract basis. Further, non-equity alliances are divided into licensing agreements, supply agreements, and distribution agreements. In terms of of alliance constraints and agreement, this kind of alliance in the integration boundaries may be used to reduce expenditure for trading, contract signing, and compliance. Specifically, these related expenditures would include money, time, spirit, and so on. Through analyzing strategic alliances based on the theory of transaction cost, the alliance can easily reduce transaction costs for both trading parties (Williamson, 1985; Hennart, 1988). Therefore, based on the analyses of non-equity strategic alliance boundaries, this study suggests that the theory of transaction cost should be considered for the basis of the strategic alliances.

Badaracco and Hasegawa (1988) analyzed Ford and Mazda for their shareholding relationship with each other and found that the two sides complement the existing technology-based resources; so, each can obtain critical technologies from the other side.

In other words, when an organization can not obtain benefits independently, and then may wish to take advantage of strategic alliances with other companies to achieve the purpose of complementary resources (Lynch, 1989).

In terms of Resource-Based View (RBV), cross-boundary equity alliance firms primarily are motivated by a desire to obtain resources from outside. However, in order to absorb external information and technology, organization itself must first have the necessary in-house technological skills (Mowery, Oxley, & Silverman, 1998). Stafford (1994) argued that this strategic alliance is as a "pooled strategy", which emphasizes the cooperative relationships for the concentration of resources. Mowery, Oxley, and Silverman (1998) and Gulati and Singh (1998) proposed strategic alliances between the parties combined their technical resources would have some close integration of boundaries resources to complete the joint activities. Therefore, this study suggests the integration boundaries of strategic equity alliances should be based on the theory of RBV.

A joint venture is to create a legitimate and independent company and shares the business benefits (Barney, 2006). Its network relationships usually involve a large range of considerations. Gulati (1998) argued that when the enterprises enter the joint venture model, the relationship will be key inter-organization basis of trust. Mowery et. al. (1996) emphasized that the joint venture alliance has a better effect on transferability of knowledge and related competence than any other type of alliance because of the dependence of the performance share.

The knowledge and related competence based on integration boundaries to business performance are important for the explanatory power. Simultaneously, due to share ownership, responsibility, risk, and reward, the existence of relationship between strategic alliances becomes one of the most important factors of inter-dependence (Lynch, 1989). Thus, in view of the integration boundaries of the joint venture, this study considered that the basis of strategic alliance should be constructed on the views of relational capital. Propositions 2, 3, 4, 5 are below:

Proposition 2: The higher the intensity of strategic alliances, the higher will be the possibility of integration boundaries.

Proposition 3: There are high relationships between the integration boundaries of non-equity

strategic alliances and the theory of transaction cost.

Proposition 4: There are high relationships between the integration boundaries of equity strategic alliances and the perspectives of the theory of RBV.

Proposition 5: There are high relationships between the integration boundaries of joint venture strategic alliances and the perspectives of the relationship capital.

The comprehensive strategic alliance and the integration boundaries are integrated on perspectives of transaction cost, resource based view, and relationship capital, and as shown in Figure 3.
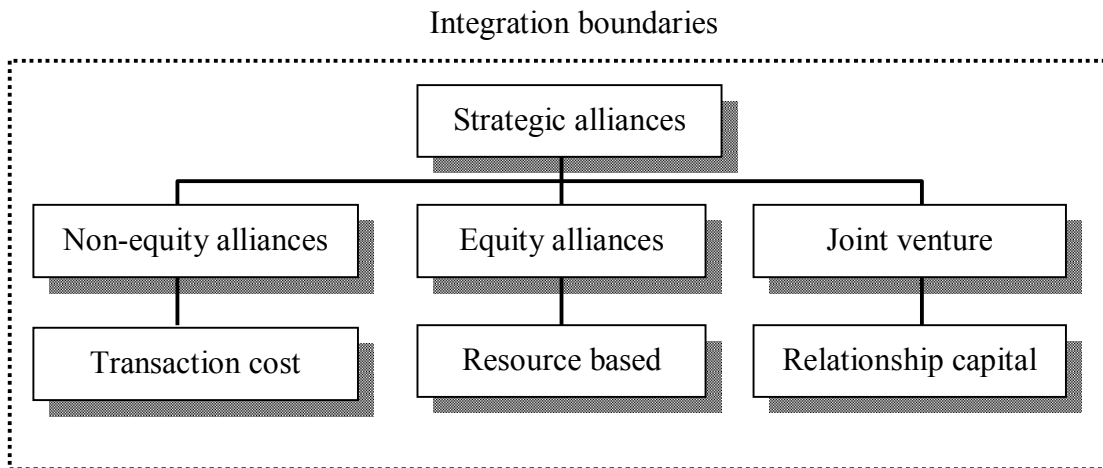
Integration boundaries

Figure 3: The relationship between strategic alliance and integration boundaries based on the perspectives of transaction cost, resource based view, and the relationship capital.
Source:Modified from Strategic Management and Competitive Advantage (Barney & Hesterly, 2006, p. 279）

THE STRATEGIC ELEMENTS IN BOUNDARIES INTEGRATION

The concept of integrated boundaries originates from the value chain activities between the materials to end users. The degree of integration was judged by the steps between corporate boundaries used in the value chains（Barney & Hesterly, 2008）, The values earned by the integration came from the reducing the risks of speculation（Williamson,1975, 1985; Crawford and Alchian, 1978） and most discussiones in boundary integration based from transaction cost theory（Shelanski and Klein, 1995; Jacobides and Hitt, 2005）. Williamson（1975）had provided a "Transaction Government structure" which recognized that by analysis of the frequency and investment specificity of transaction pattern, the managers could adapt proper decisions in either "Hierarchy Choice" or "Market Choice" models, and no matter in market dominant, dual dominant or triparties dominant situation, the integration spontaneously happened in the industry. To survive in the processes of boundaries integration, the corporate should emphasie corporate capabilities and abilities those would keep the sustained competitive advantages（Barney,1991,1999;Conner and Prahalad, 1996）.

To discuss the boundaries integration in the frame of resource base theory, Tsang（1998）

6

suggested that the power of the corporate resource influences the competitive advantage. These resources should be valuable, scarce, and non-imitated. Ramaswamy（1997）argued that the strategic similarity in boundaries integration do influence the performance. Other researcher had explained the practical strategic behaviors empirically with RBV(Hou,1998).Collis（1991）suggested that RBV complementally explained the growth strategies that could not be fully discussed in traditional theories. The values of the boundaries integration should include stabilizing the supply chain, and maintaining long term relationship operations. And the goals of the resource integration should be to using the non-replaceable resources to purchase the long term competitive advantage（Afuah, 2001;Schilling & Steensma, 2002）. The core resource should be integrated with the complementary ones to produced sustained competitive advantage（Lin & Yang,2007）. Porter（1985） and Papows （1999）suggested in the processes of resources integration management, the goal is to direct the corporate to operate with her core resources and to outsource the non-core ones. The corporate should maximize her benefit by comparing the costs and performances the internal resources and external resources operated（Rodrigues & Child,2003）.

Contrarily, to discuss the resource base prospect in the frame of boundaries integration, the corporate should evaluate the situation the corporate exist and all the value activity the corporate operated and coordinated. From this point, to build up the sustainable competitive advantage would be the first priority the corporate should pay attention when integrating boundaries.

While considering the role of boundaries integration in strategic relationship capital, Standifird & Marshall（2000）believed that trust played only as a "assurance" in the relation capital. Hou(1999), in a cross national study, explored that when the resource was highly inexpressible and non tradable, the firms tend to have the possession; in the other way, when the core resource was highly expressible and tradable, the firm tend not to have the possession. From the view of investment, the possessions right may be "Transaction/relation-specific investment"(Chao, 1993）.

Gulati（1995）suggested that there were some rules in the resources exchange behavior. When the core resources was involved in the exchange relation, the deterrent force sometimes used in practical situations. Tsai and Ghoshal（1998）argued that the relation capital not only embedded in the social structural resources, but also the norm and values between the partners（Nahapiet and Ghoshal,1998）.That's to say, the relation capital was not totally based on trust, but a dependent relationship in the norm constructed between dual interaction. And under the norm, one could get the summed resources in social embeddedness（Nahapiet and Ghoshal,1998）.Only by the closed interactions with the society, the corporate enforced the width, scope and frequency in the resources transaction.（Yi-renko, Autio and Sapienza, 2001）.

Fuchs（2000）had used the concept of tuning in integrated strategies practices to described that the strategies integration was like a orchestra. Each player did his best to constitute a integrated, complemented chapter with a integrated rules . Fuchs（2000）inferred that the most competitive corporate was not excellent in every field of competition, but the one who efficiently organized his market positioning and integrated his practice(Hwong, 2006).

The authors concluded that in the boundaries integration, by the constrained power from the

7

economic transaction, the resources complementation from resources base theory and the strengthening of the interdependency in relation capital, strengthened the relation and the interdependency between the firms as the integration level elevated. And it extended the traditional linear model of value chain into a value space（Papows, 1999）. The corporate could exist in multiple value chain within the same industry or cross the industries to form the sustained competitive advantage.

That's to say, the boundaries integration implied the transaction cost theory, resources base theory and relation capital theory. And there was some similarity among these three theories. The authors inferred the proposition 6 as:

Proposition 6:The boundaries integration was influenced by the transaction cost theory, the resources base theory, and the relation capital theory.

The authors proposed a strategic structure for the boundaries integration as the figure 4:
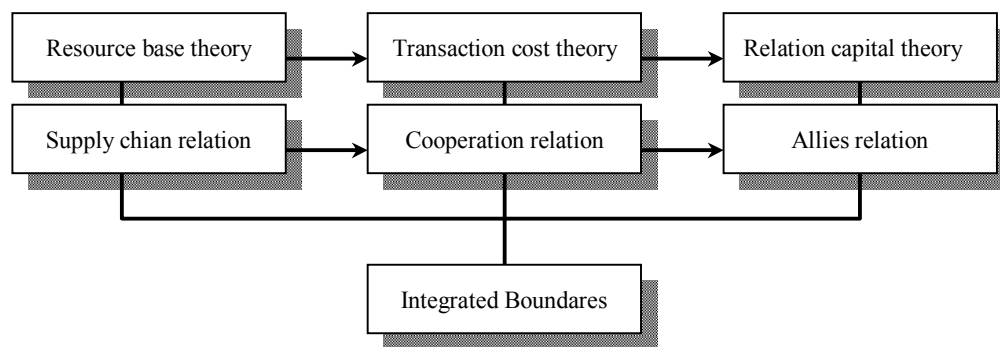
Figure 4: The structure of resources base theory, transaction cost theory and relation capital theory
（1）The resources base theory and the supply chain relation

Penrose（1959）had argued in "The theory of the growth of the firm"that to earn the benefits, the corporate not only need to possess excellent resources, but also the distinctive competence to use the resources（Montgomery, 1996）. In resouces base theory, the "continued deepening and repeating of the strategic resources" was emphasized（Prahalad & Hamel, 1990）. Amit and Schoemaker（1993）had recognized that the strategic resources have 8 valuable characteristics. Peteraf（1993）at the mean times devided the competitive advantages of the resources into four categories: heterogeneity, imperfect mobility, ex ante limits to competition, ez post limits to competition(figure 5). Heterogeneity means monopoly; imperfect mobility means rent resource; ex ante limits to competition means enlargement of competitive competitors' risk; and ez post limits to competition means sustainability. The organization should face the problems of resources using and the difficulties to obtain the resources（Jeffrey Pfeffer, 2007）.

8
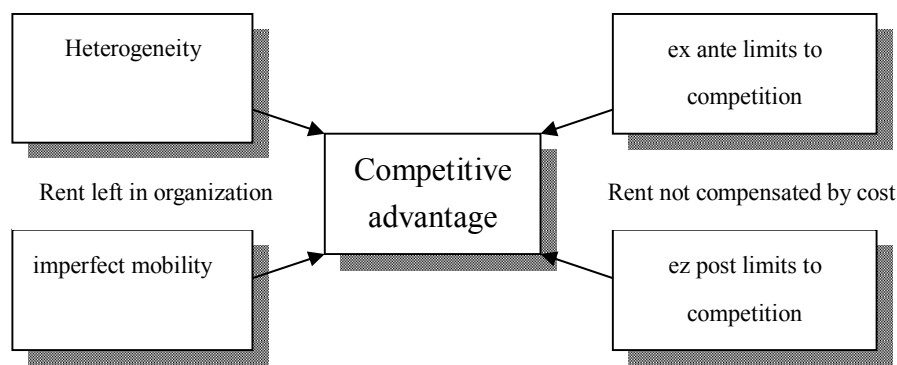
Figure 5: the foundation of the competitive advantage

Source:Peteraf, M.A.（1993）,"The Cornerstone of Competitive Advantage: A Resouce-Based View", Strategic Management Journal, Vol.14,pp.179-191.

Grant (1991) thought of the resources as the foundation of the profitability of the organization, and the base to form the capability of the organization. When thinking of the problems of resources consumptions, it was usual proposed that the organization had had the resources recommended. However, it is hard to keep the resources supply, the production, and the sales stable when the corporate would run smoothly only if a stable transaction would promise the resources availability （Jeffrey Pfeffer, 2007）. The fluctuation of resources supply means short supply sometimes. And for some organizations, to keep the operation stabilized was more important than to earn enough profit or to keep the growth. Short of supply means hard to survive（Jeffrey Pfeffer, 2007）.Wernerfelt(1984) argued that the proper usage of the resources and the efficiency of resources management, enabled the organization to accumulate the resources advantage that the other competitors would not achieved and formed the sustainable competitive advantage.

It means that the unstable resources supply relationship brought to the organization more operational risks in the supply chain relationships（Simchi-Levi, Kaminsky, 2000）. That is why Leiblein & Miller（2003）recognized the vertical integration of the industrial resources as the trend of the global development.

Huang & Lee（2005）had categorized the supply chain management into four identities in resource base: 1. to the suppliers, the supply chain management provides definite and stable needs, reducing the cost of storage and keep the trust relation; 2 to the producers, supply chain management explored the resources gaps among the organizations upstream and downstream in the same industry, justified the time and categories of material supplies according to the resources inventorying principle; 3. to the distribution business, the supply chain management shortened the lead time for stocking, reduced stock cost and operated the resources efficiently; 4. to the retailers, the supply chain management means stabilized product supply, real time ordering, small amount of diversity, high frequency of replenishment, and quick return and barter. And to the end users, reducing prices, improved quality and service, and satisfaction owing to the full supply.

In the prospect of resources integration in cross boundaries, the supply chain management means efficient cost management. In the frame of resource base theory, the fluency of resources allocation processes implies operation profits, and the more profitability means the more competitive

9

advantage. In either prospects, the relationship management, resource flow management and the logistic management are the key successful factors（Carter and Narasimhan 1996, Brewer and Speh 2000）. Relation ship management implies mutual trust, commitment and benefit sharing. Stuart （1993）suggested that with the complementary resources bases, the supply relation and processes would trigger the formation of strategic allies and this allies usually mean powerful, long term and interdependent relations. That statements inferred our proposition 7.

Proposition 7: The higher the cooperation level is, the tighter the supply chain relationship will be in resources base prospect.

（2）The inter-organization cooperation relationship in transaction cost prospect

The transaction cost theory had been applied in many fields. For example,including engineering intensity and design specialization（Monteverde and Teece, 1982a, 1982b; Masten, 1984）, the technical uncertainty（Walker and Weber, 1984; Balakrishnan and Wernerfelt, 1986）,the co-localization of specialized assets in corporation（Joskow, 1987）.Most of the papers discussed the topics of inter-organization cooperation relation was based on the transaction cost theory（Shelanski and Klein, 1995; Williamson, 1990; Jacobides and Hitt, 2005）.

Classically, the transaction cost theory argued using of contracts and rights to reduce the speculation behaviors. Williamson（1975）had suggested that the market dysfunction, the transaction unable to carried on, resulted from the human and environment factors in the transaction processes. And it resulted in the difficulty in transaction and the production of transaction cost (figure 8). Williamson（1985;1991） further divided the assets by the specificity into site specificity, physical asset specificity, human asset specificity, brand name capital, dedicated asset and temporal specificity.
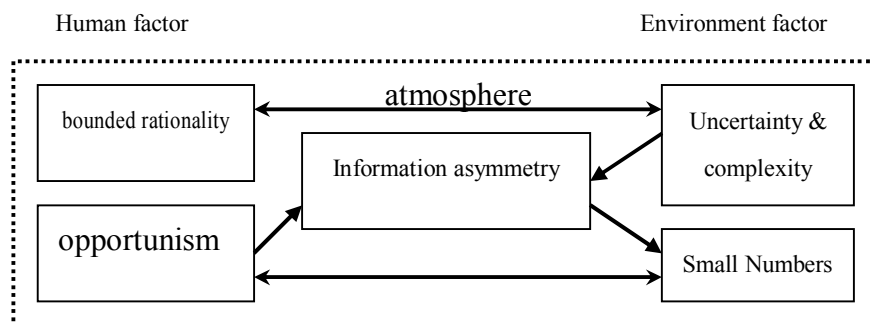
Figure 6：organization and market dysfunction
Source:Williamson, 1975, p.40.

Generally speaking, when the level of asset specificity becomes higher, the organization will tend to internalize the transaction to reduced the transaction cost. Koopmans（1965）had classified the uncertainty of resources transaction into two categories. First，the anticipated and non-anticipated accidents resulted from bounded rationality. The non-anticipated accidents were accompanied with the complexity; Second, The uncertainty of information insufficiency due to concealment, deception and information distortion. This uncertainty increased the governance cost (Lin and Chen, 2007).

Robertson and Gatignon（1998）classified the uncertainty into external uncertainty(demand

uncertainty and technique uncertainty) and internal uncertainty (behavior uncertainty). Williamson （1985）also suggest the influences of transaction frequency. As the frequency of transaction increased, the cost of communication, coordination, contracting and negotiation increased and all of them increased transaction cost.

Anderson and Gatignon(1986) advised that in control of the inter-organization cooperation relation with the transaction cost prospect, the entrants would applied the power of control to get the majority of the profits and rewards. The cost of control will include: the duty of decision, the commitment of resources, the shifting cost, and the loss of flexibility in operational allocation. In the other way, the contracts will actually define the behaviors in the transaction. And these behaviors mean the trust, norm and the ethic duties in the relation structures（Nahapiet & Ghoshal, 1998）. Williamson（1979）,Kogut（1988）,Pisano（1989）suggested that from the rights and obligations relations, the commitment in anticipation internalized the external environment and efficiently made the transaction under control. However, no contracts could identify all the behaviors needed to be normalized and the rights and interests conflicts produce speculation behaviors. That is the reason we need the relation capital based on trust（Besanko, Dranove, & Shanley, 1996; Ahuja, 2000; Kale, Singh & Perlmutter, 2000）. From the prospect of transaction cost, both the trust and reliable information in the inter-organization cooperation relation could reduce the transaction cost efficiently （Uzzi, 1977; Nahapiet & Ghoshal, 1998; Kale, Singh & Perlmutter, 2000）. It came our proposition 8:

Proposition 8: The higher the reliability, the high the level of cooperation relation in transaction cost prospect.

（3）The strategic allies relation in relation capital prospect

Hwong（2008）,in his "The Determinants of Integration Boundary of Enterprises in Taiwan through the Viewpoints of Supply Chain", empirically found that according to the difference between industries, the emphasis of strategic direction in the integration of supply chain varied. Yet, no matter which industry was considered, the partner relationship play a major and sophisticated influences in the relation capital besides the reducing of transaction cost and the build-up of capability. Thorelli（1986）, Kale, Singh & Perlmutter（2000）defined the relation capital as the inter-organization trust relation from closed cooperation. Since the trust relation brings the profits in future, it need to pay the cost to maintain（Alder & Kwon, 1999）. The strategic allies established by the relation capital could promote the speed of market responses, the flexibility of management, and the efficiency of operation（Lorenzoni and Fuller,1995）.

In the prospect of resource base, the relation capital relies on the key resources. It is inexpressible and non-tradable (Hou, 2000). Gulati（1998）pointed out that once the strategic allies constituted, the organization should recognized the key dependency between the environments and the organization. And there is a dependency between relation capital and the strategic allies. Barney （2008）suggested three opportunities in strategic allies: Improving operation performance, creating a more competitive situation and making the corporation enter or quit the market. Part of the

11

organizations participated the strategic allies only aimed to evaluate and select the technique standard in the industry. The changes in technology may be of competence enhancing and destroying （Tushman and Anderson, 1986）, and of recombination. It enhances the competition between the new entrants and the original, just like a revolution（Abernathy and Clark, 1985）. In the situation of resources inter-dependency in relation capital, the organizations organize inter-dependency relation from strategic allies.

The discourse between the transaction cost and the relation capital started from the economic issues. The transaction cost theory explained the economic efficiency in the corporate institutionalization, and proposed that due to the economies of specialization and the administrative and incentive limits of hierarchies, the strategic allies should be a efficient governance structure （Pisano, 1990）, unless in a special situation（Williamson, 1985; Grossman and Hart, 1986）. Coase （1977） recognized that the transaction cost between the organization would be zero, and suggested that there were two type of governances, the market governance and the hierarchy governance. If the cost for the organization to produce the product was higher than that purchased in the market, the market governance should be applied; otherwise, when the cost for the organization to produce the product was lower than that purchased in the market, the hierarchy governance should be applied to extend the organization and internalized the transaction. The transaction cost theory help us to realize the R&D research boundaries selection. While technological change happened, the abandon of original technology and the preservation of commercialization ability should be trade off and the original organization had the ability to vertically integrate the transaction profits which injured the external organizations（Pisano, 1990）. There exist a constrained interdependent relation between the relation capital and transaction cost. The transaction cost theory formally identified the norms and roles of the organization, while the dependency in relation capital based on the strategic allies the organizations constituted. It came the proposition 9.

Proposition 9: The higher the level of dependency is, the higher the strategic allies relation is in relation capital prospect.

While discussed the integration boundaries based on the resource base theory, transaction theory and the relation capital theory, the authors put the level of cooperation, the level of trust and the level of dependency into the framework of the boundaries integration as figure 7:
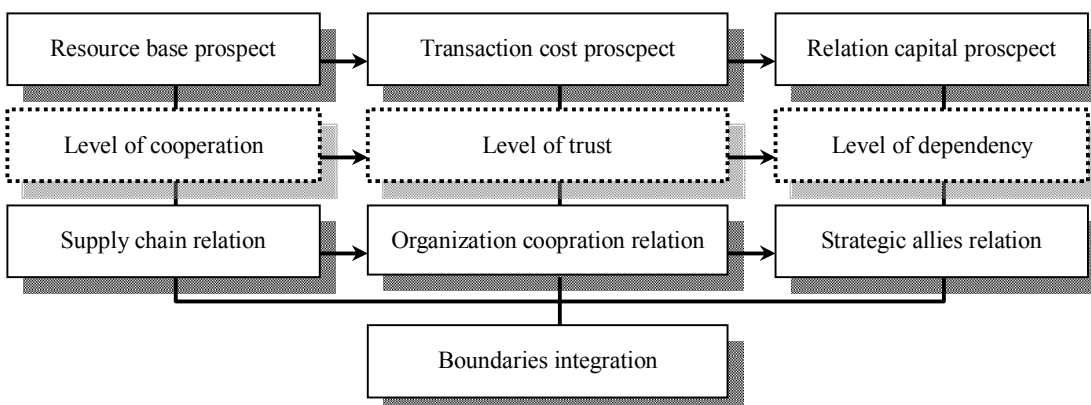

Figure 7: Modifies boundaries integration framework.

THE SUSTAINABILITY OF RELATIONSHIP CAPITAL, INTEGRATION BOUNDARIES, AND COMPETITIVE ADVANTAGE

Barney (2006) argued that a strategic alliance can achieve the sustainable competitiveness because "value," "rarity," "imitability," and "organization" based on integration boundaries can create economic benefits by taking opportunities and avoiding threats. Dyer and Singh (1998) emphasized the governing mechanism of self-enforcement in a strategic alliance can interpret the reason for competitive advantages of cross-boundaries. Comparing with the transaction cost theory, this argument for strategic alliances has more emphasis on relational capital management.

To the "imitability," Dyer and Singh (1998) further pointed out that the establishment and maintenance of relational capital is a process of interdependent paths under the strategic alliance. The trust and interdependence in business relationship network often are unique and intangible for competitors to easily imitate. Namely, the relational capital of strategic alliance can be seen as a particular and valuable resource. Relational capital filled with trust can provide immediate, credible, and valuable information, and then an organization can response faster than competitors to distribute resources including labor in hypercompetitive environment for effectively reducing management cost. According to Barney's (1991) theory based on RBV, the key issue is not "whether a strategy is easy to implement or not," but "whether the strategy is more easily to perform than the competitive rivals" if the cost of the strategy implementation is less than that can create value.

Through analyzing the relative costs of strategy implementation, the firms may have two errors occur. First, their own unique resources are overestimated. Second, their own resources uniqueness are underestimated. To properly understand the relative costs and the relatively competitive advantages of the strategy implementation, the firms' value, rarity, and imitability of the resources must be assessed appropriately. However, to have an accurate assessment for value of resource characteristics is still a big challenge; thus, the relational capital is a critical element for resources in integration boundaries.

Grant (1990) integrated the relational management with applications of competitive advantage in the market issues, he emphasized that competitive advantage comes from the barriers to resources imitability. On the one hand, these barriers can block competitors from their ability to challenge; on the other hand, the existence of the relationships between firms can make the "reputation resources" difficult to transfer. Moreover, causal ambiguity of relational management can effectively lead to uncertain imitability. Further, difficult to understand complicated relationships between organizations make competitive rivals difficult to speculate. Imperfect transferability of resources depends on trust in transaction cost and the degree of mobility of resources.

In terms of barriers to resources replication, the complexity of the relational basis for firms' cooperation will make competitors difficult to replicate after a long effort to build their resources network. Barney (1991) also considered that the delicate inter-organizational relations cannot be based on the systematic management and general control so that the competitors cannot easily replicate and replace the relationship. Therefore, the relational capital of the integration boundaries will be a positive impact on a sustainable competitive advantage. The assumption of the proposition

13

10 proposes the relational capital can be integrated into the strategic management process, and as shown in Figure 8:

Proposition 10: The higher the relational capital of boundaries integration, the higher will be sustainable competitive advantage for the strategic alliance.
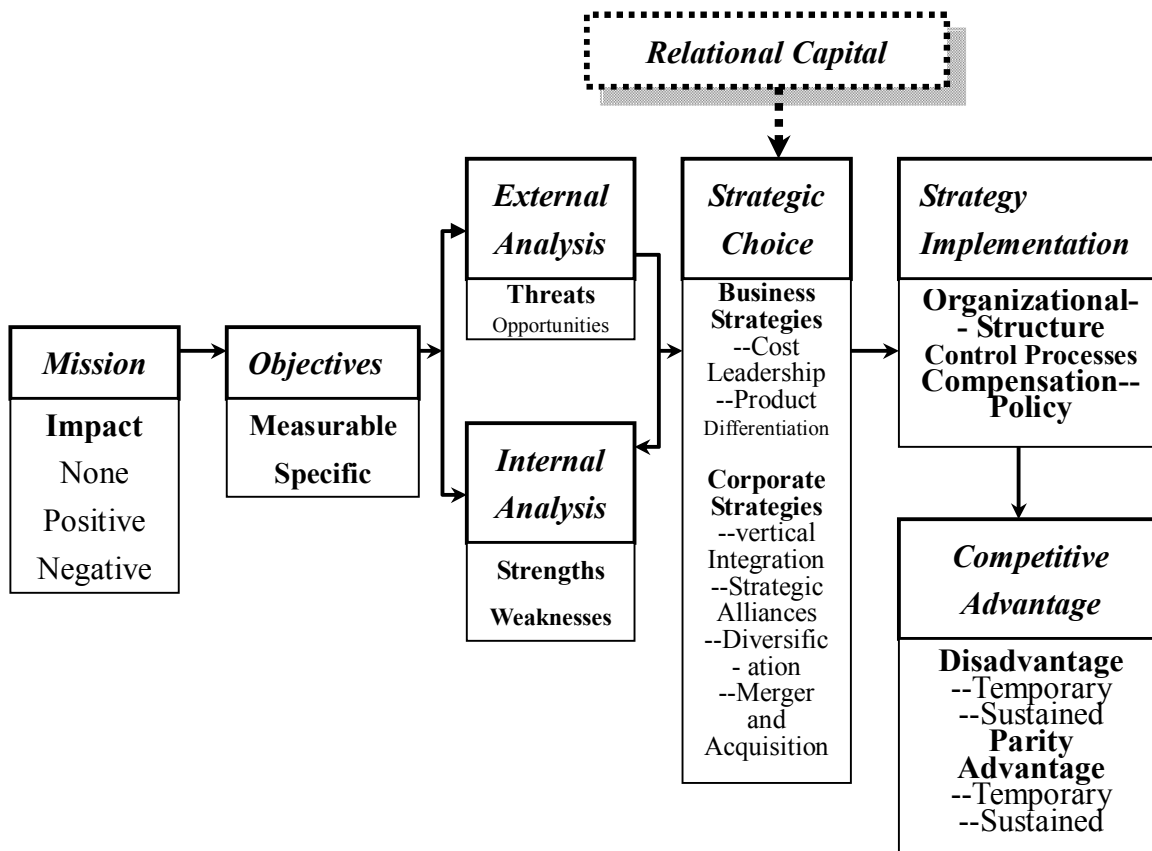
Figure 8: Relationship between Relational Capital and Strategic Management
Source:Modified from Strategic Management and Competitive Advantage (Barney & Hesterly, 2006, p. 15).

CONCLUSION

To participate in the strategic allies was the most efficient way for the corporation to get the external resources. However, without specific exchange relationships and in a constrained market space, it was hard to gain extra values beyond the ordinal exchange relation and so does it to gain the differences that constitute the competitive advantage. In the strategic allies, by way of managing the relationship capital, under the case formed by the mutual trust and interdependency, it positively influences the economic performance of the corporation. Besides that, stabilization of the cooperative relation, the preferential rights for chances selection, and the share of duties and risks, all those potential benefits in non-economic advantages need the modulation of relation capital. That's to say, the relationship may be trial yet play a major role in the resource element of the corporation.

In this study, the authors argued that owing to the different resources patterns the corporate occupied, the corporation operated differently. A unique way to execute the strategies faces a

14

particular difficulty. When thinking of the choice of strategies, the corporate weights the difficulty, the cost, the financial benefits, and the coordination of the strategies with the vision and vision of the corporate and the approaching of the goals. By applying the concepts of relationship capital and boundary integration, it enforces the corporate to gain resources, to integrate the resources from each organization in the allies to form the competitive advantages. Practically, the authors suggest that when the variance in exchange situation was high, the inter-organization communication was tough; the board members could resolve the problems while managing the relation capital to eliminate the obstacles and to purchase the benefits. Besides, in the process of the strategy management, it was recommended to review the particular resources that corporate had and the complementary resources that needed, to carefully evaluate the internal and external environments the corporate in, and to category the core resources and complementary resources, in order to allocate the proper resources for competitive advantage. In the study, the author adapted the relation capital into the strategic management processes and suggested further researches from other prospects.

REFERENCES

[1]. Barney, J. "Firm resources and sustained competitive advantage", *Journal of Management*, 1991, 12, 49-68.

[2]. Barney, J., "Strategic Management and Competitive Advantage—Concepts and Cases." chap.1,9.2006.

[3]. Bart, C.K., and M.C. Baetz., "The relationship between mission statements and firm performance: An Exploratory study." *Journal of Management Studies*, 1998, 35(6), pp.823-854.

[4]. Badaracco, J.L., and N.Hasegawa. "General Motors" Asian alliances. Harvard Business School Case, 1998, No.9-388-094.

[5]. Bozeman, Barry, and Albert Link., "Investments in Technology: Corporate Strategic and Publiv" *Policy Alternatives*. New York: Praeger, 1983.

[6]. Chao, Ching-Ping. "The determinanats of joint action in OEM relationships", Taipei., 1993.

[7]. Collis, D.J. "A Resources-based Analysis of Global Competition: The Case of Bearing Industry". *Strategic Management Journal*, 1991, 12, pp.49-68.

[8]. Conner, K.R., and C.K. Prahalad. "A resource-based theory of the firm: Knowledge versus opportunism." *Organization Science*, 7, pp.477-501. 〔1〕Dyer JH, and Singh H.,(1998). " The relational view: Cooperative strategy and sources of interorganizational competitive advantage," *Academy of Management*, 1996, 23(4): p660 -679.

[9]. Gulati R. and H.Singh., "The architecture of cooperation: Managing coordination costs and appropriation concerns in strategic alliances', *Administrative science Quarterly*, 1998, 43, 78-8.

[10]. Gulati R. and H. Singh., "The architecture of cooperation：Managing coordination costs and appropriation concerns in strategic alliances", *Administrative Science Quarterly*, 1998, 43, 781-814

15

[11]. Hitt, A.M., R.D. Ireland and R.E. Hoskisson.,"*Strategic management：Competitiveness and globalization*", West Publishing Company,1997.

[12]. Hill, C. and R. E. Hoskisson., "Strategy and structure in the multi-product firm,"*Academy of Management Review*,1987,Vol. 12, No. 2, pp. 331-341.

[13]. Hou, Jia-Jeng and Tang, Ming-Je.,"*Capacity Competition and Strategic Interaction: The Case of the Chemical Industries in Taiwan* ", The 17th Annual International Strategic Management Society Conference, October 5-8, 1997 in Barcelona.

[14]. Lin Jin-Kawn, Hsiu-Fang Yang.," The Competition of World's Leading Ports with Strategic Groups Analysis: Resourse-Based Perspective"*,Maritime Quarterly,* 2007*,*Volume 16, Number.2.

[15]. Lin, Nan., *ASocial Capital: Contending Paradigms and Empirical Evidence.*, Hong Kong *Journal of Sociology* ,2001,No. 2, pp. 1-38

[16]. Ming-Chang Huang; & Su-Chen Yang.,"The Relations between Resources Configuration in Strategic Alliance and Competitive Advantage",*Chung Hua Journal of Management*,2002,Volume 4, Number.2.,pp.59-75.

[17]. Mowery,C.D.,J.E.Oxley and B.S.Silverman.,"Technological overlap and interfirm cooperation:Implications for the resource-based view of the firm."*Research Policy,*1988,27,507-523.

[18]. Nueno, P. and Oosterveld, "Managing technological alliances", *Long Range Planning*, 1988,21（3）, 11-17

[19]. Pavitt,Keith.,"*Chips"and "trajectories."How does the semiconductor influence the sources and directions of technical change?*.In Roy MacLeod(ed). Technology and the human Prospect:Essays in Honour of Christopher Freeman:31-54 London:frances Pinter,1986.

[20]. Pisano, Gary P.,"The R&D Boundaries Of The Firm: An Empirical Analysis"*Administrative Science Quarterly;* Mar 1990; 35, 1; ABI/INFORM Global pg. 153.

[21]. Ramaswamy.K."The Performance Impact of Strategic Similarity in horizontal Mergers:evidence from the U.S. banking Industry."*Academy of Management Journal*, 1997,40(3),pp.697-715.

[22]. Stafford, R.E."Using co-operative strategies to make alliances work", *Long Range Planning*,1994, 27（3）, 64-74.

[23]. Teece., "Profiting from technological innovation：Implication for integration, collaboration, licensing and public policy", *Research Policy,*,1986,15, 285-305.

[24]. Ting-chung Huang.."The Determinants of Integration Boundary of Enterprises in Taiwan through the Viewpoints of Supply Chain", *Journal for SME Development* ,2008,Number.10,pp. 187-206.

[25]. Tsai, W.,"Knowledge transfer in intraorganizational networks: Effects of network position and absorptive capacity on business unit innovation and performance,"*Academy of Management Journal*,2001,Vol. 44, No. 5, pp. 996-1004.

[26]. Williamson, O. E.."Transaction Cost Economics: The Government of Contractual Relations",

16

Journal of Law and Economics,1979, 22(3), pp233-261.

[27]. Williamson, O. E..,"The Economics of Organization: the Transaction Cost Approach", *American Journal of Sociology*,1981,Vol.87, no.3, pp.548-577.

[28]. Williamson, O."Comparative Economic Organization：The Analysis of Discrete Structural Alternatives," *Administrative Science Quarterly*,1991,Vol.36, pp.269-296.

[29]. Wu, Se-Hwa, and Hsu, Frederick B."Toward a knowledge-based view of OEM relationship building- Sharing of industrial experience in Taiwan."，*International Journal of Technology Management.*,2001,Volume 22, Nos.5/6.

17

HOW TO LESSEN THE DISPOSITION EFFECT? ON THE INVESTMENT STUDY HYPOTHESIS

Kuo, Min-Hua

Associate Professor, Department of Finance, Shih Hsin University

111, Sec. 1, Mu-Cha Rd., Taipei, Taiwan, 886-2-22368225~63435

mhkuo@cc.shu.edu.tw

ABSTRACT

The prevalence of the disposition effect and its disadvantage to investment returns has widely gained empirical supports. However, none of the research, to our knowledge, gives investors any advice to overcome it. This study proposes the investment study hypothesis and suggests that investors' intention to make investment studies contributes to lessening the disposition effect. By means of a nationwide survey in Taiwan on the individual stock investors, this study collects 1672 effective samples and demonstrates significant evidences to support the study hypothesis. That is, when investors have greater intention to make investment studies (including the time and the media to devote), they keep losses shorter and are thus subject to less disposition effect.

The results represent that if investors are willing to do the investment study, they are more likely to be rational oriented and less dominated by the sentiment of regret averse, which contributes to lessening the disposition effect.

Key words: disposition effect, the investment study hypothesis

INTRODUCTION

Ever since Shefrin and Statman(1985) extended the Kahneman and Tverskey's Prospect Theory to the field of investment and proposed the disposition effect (DE thereafter), DE has been widely documented in the capital market (such as Lakonishock and Smidt 1986; Ferris, Haugen and Makhija 1988; Odean 1998; Weber and Camerer, 1998, Kuo, Kuo, Chiu, and Fan, 2005, etc.). The DE indicates that investors have the tendency to hold losers too long and sell winners too soon. What dispose investors to such effect? Except for the asymmetry of risk preference over gains and losses as argued by Kahneman and Tverskey's (1979) prospect theory, several other reasons were suggested, such as cognitive dissonance (Goetzmann and Peles 1997), quasi-magical thinking (Shiller 1999), insufficient self-control(Shefrin and Statman 1985; Glick 1957), and mental accounting, pride seeking

and regret averse (Shefrin and Statman 1985) and so on. These potential reasons belong to, roughly speaking, two psychological biases: perceptional bias and emotional bias.

The disposition effect is found disadvantaged to investment performance (Benartzi and Thaler 1995; Odean 1998; Grinblatt and Keloharju 2001), but little advice is provided to help investors to avoid it. Simply advising investors to hold winners longer or to sell losers faster is null and void if the DE is driven by psychological factors as mentioned. We need certain mechanisms to overcome the psychological forces and intrigue correct actions. People change behaviors when they gain new information and thus modify their expectations (Beaver 1968; Grossman 1976). Accordingly, more relevant information might equip investors with better judgment on the price trend and less uncertainty, leading investors to be more rational oriented and subject to less disposition effect.

HYPOTHESIS

Based on the above reasoning, it is plausible to argue that study is valuable for investors in terms of its contribution to reducing perception errors and the sense of uncertainty. Those who are more willing to devote themselves to making investment study would have more precise perception on stock value. Investors would be less reluctant to realize losses if they "know", say, the losers are likely to devalue further; they would be more willing to hold winners longer if they believe the winners are likely appreciating more. Because if they mistakenly keep losers longer or sell winners sooner, they would turn to even bigger regret. Therefore, the more disposed to make investment study of investors, the less disposition effect they would show. Based on this reasoning, we propose the investment study hypotheses as follows.

H1: If investors are more inclined to make investment study, they have more tendency of selling losers.

H2: If investors are more inclined to make investment study, they have more tendency of keeping winners.

H3: If investors are more inclined to make investment study, they show less disposition effect.

DATA SOURCES AND VARIABLE DEFINITIONS

To test the hypotheses, we use the databank of Individual Investors Sentiment Index in Taiwan, a nationwide survey targeting on individual stock investors. Three surveys in the databank are regarding the disposition effect, investigated on August 2004, April and June

2005 respectively, and 2632 samples are collected. The surveys are made through phone call by the Professional Survey Center of Shih Hsin University, Taiwan. The sampling error is below 3% in 95% confidence level.

The disposition effect represents "selling winners too quickly and holding losers too long" (Shefrin and Statman 1985). Various techniques are used to measure the DE, such as the selling proportions of winners vs. that of losers (Odean 1998; Weber and Camerer 1998, etc.), the comparison of holding periods (Odean 1998; Shapiro and Venezia 2001) or the price changes (Kuo, 2010) of losers and winners, and so on. For example, Odean(1998) found the holding period of losers (median 124 days) are longer than those of winners (median 104 days) and the proportion of winner realization is 24%, compared with 15% in losers. Kuo (2010) reported that nearly one third of individual investors rejected to sell any losers, which leaves the average absolute return of losers is significantly greater than that of winners. This study adopts the same measure of Kuo (2010), which is constructed by two questions as follows:

● On the average, how much do you earn when you sell the stocks in gains?
  ☐0~5%　　☐6~10%　　☐11~20%　　☐ 21~30%　　☐ 31~50% ☐ More than 51%
  ☐ I do not sell.
● On the average, how much do you lose when you sell the stocks in losses?
  ☐0~5%　　☐6~10%　　☐11~20%　　☐ 21~30%　　☐ 31~50% ☐ More than 51%
  ☐ I do not sell.

We construct the DE measure by the difference of the price changes of losses and gains. We take the median of each bracket above as the proxy measure. For example, we take 8% to represent the region of 6%~10% and 15.5% to represent the region of 11%~20%, etc. We take 60% and 100% to represent "more than 51%" and "I do not sell" respectively. Thus,

DE_i = price change of losers to be realized$_i$ −price change of winners to be realized$_i$

If DE>0, it indicates the investor i is more willing to hold losers than winners, in terms of the price changes, before realizing the gains/losses, consistent with the disposition effect. If DE<0, it indicates the investor i is more willing to hold winners than losers, in terms of the price changes, before realizing the gains/losses, which is labeled as the reversed disposition effect in Kuo (2010). And if DE＝0, it indicates the investor i has no significant preference to dispose of winners or losers, in terms of the price changes.

As for the study measurement, there are two questions in the data bank as follows:

- If possible, how many hours a week at most will you spend in studying investment information?

- If possible, how many media a week at most will you access to get the investment information? (For instance, it is 5 media if you will access one magazine, two TV channels and two newspapers.)

The first question above is investigated in all three surveys but the second is only in one survey. Both questions are of subjective willingness, instead of objective actions. It is more plausible to inquire the subjective intention than if we ask the interviewees the real endeavor they make in studying. The latter may involve greater estimation error, because some pieces of time in reading/listening the relevant information (say, reading/listening on the way home/work, etc.) may be counted in by some interviewees while not by others. Therefore, there are at least two advantages to inquire the subjective intention. First, the measurement error is more limited. Second, it reflects the interviewees' intention or motivation of devoting themselves in studying on investment, which is even more relevant in exploring investing behaviors. Higher motivation usually represents better quality of actions. When only limited time is allowed to study, one with ambition will try to enhance his/her study quality by other routes, such as to discuss the investment in question over the interpersonal network.

In the following sections, we will examine the differences of studying efforts among disposition patterns and the estimation results of multi-regression analysis on observing the effect of study on reducing investors' inclination of holding-losers/selling-winners and the disposition effect. The conclusion follows in the last section.

EMPIRICAL ANALYSIS

Sample distributions

Regarding the sample distributions, the proportions of male and female are about the same (51.1% and 48.9% respectively); most of the interviewees are between 30~59 years old (81.7%), live in northern Taiwan (56.7%), highly educated (40.9%), have yearly income between NT$200,000 and 1 million dollars. (Table 1)

Table 1   Sample distributions

The samples come from three nationwide surveys targeting on the individual stock investors. The total samples of each demographic variable may be different due to missing data.

Variables		samples	%	Variables		samples	%
gender	male	1344	51.1	income	Less than $0.2M	524	22.0
	female	1286	48.9	(NT$)	$0.2~1M	1477	62.0

Age						
	total	2630	100	More than$ 1M	380	16.0
Age	Less than 29	290	11.1	total	2381	100.0
	30~39	717	27.3	education low	884	33.9
	40~49	768	29.3	middle	658	25.2
	50~59	632	24.1	high	1067	40.9
	More than 60	215	8.2	total	2609	100.0
	total	2622	100.0			
Location	Northern TW	1492	56.7			
	Middle TW	540	20.5			
	Southern TW	600	22.8			
	total	2632	100.0			

Investment study

Regarding the intention to make investment study, the individual investors would spend averagely 7.1 hours a week. There are 12.6% of them do not intend to spending any time in studying for investment, while 16% of them will spend more than 10 hours a week to study. They would access averagely 3.76 media to get related information (including financial newspaper, magazines, and TV programs). Most investors choose 1 or two media (46.9%).

Table 2   Sample distributions of investment study intention: time and media

The data of study time are from three surveys and those of study media are from one survey.

Time(T, hours)	0	$0<T\leq 3$	$3<T\leq 10$	T>10	Total
samples	332	934	946	420	2632
%	12.6%	35.5%	35.9%	16.0%	100.0%
No. of media (M)	0	$0<M\leq 2$	$2<M\leq 5$	M>5	total
samples	44	399	291	117	851
%	5.2%	46.9%	34.2%	13.7%	100.0%

Disposition patterns

Observing the average price changes of gains (losses), we find that three fourths of investors will realize their gains before the prices rise up to 20%; to our surprise, about 40% investors do not intend to realize any losses in hoping for the prices gaining back, and about one third realize their losses by the prices drop up 10%. (see Figure 1)
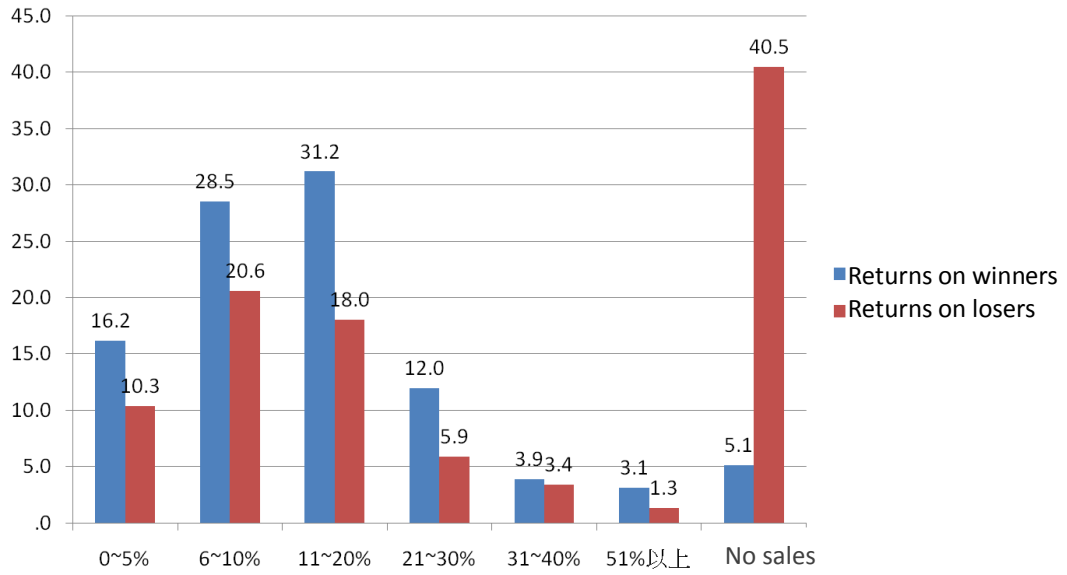
Figure 1   Sample distributions of the average price changes of the stocks held

The distribution of individual's DE shows two heights. One is around zero with one thirds of investors demonstrating the pattern of symmetry in terms of the price changes. The other height of DE is around 0.71~1, where having one thirds of investors, mostly due to the high proportion of investors reluctant to realize any losses. (See Figure 2)
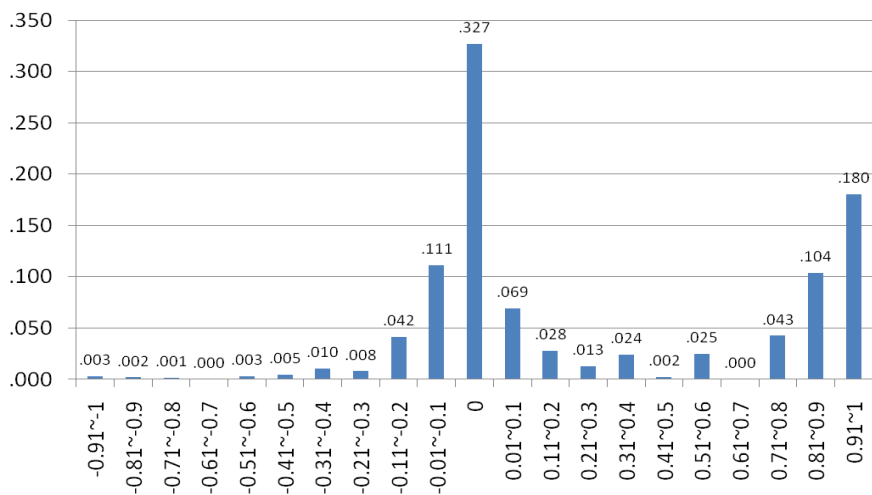
Figure 2   The distribution of DE

As a whole, most individual investors (nearly 50%) demonstrate the disposition effect; that is, they hold losers longer than winners in terms of the price changes; 32.7% of investors

have no difference between the price changes when disposing of winners and losers; 18.5% of individual investors are of reversed disposition effect, keeping winners longer than losers in terms of the price changes. (See Figure 3)
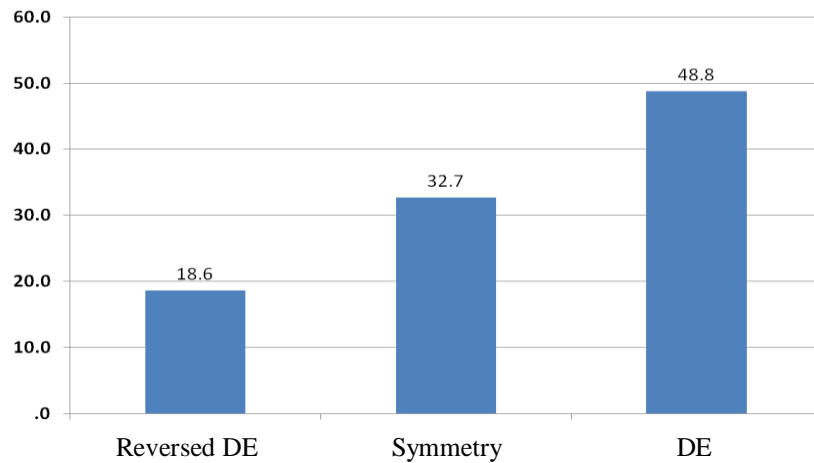
Figure 3 Sample distributions of disposition patterns

The disposition patterns and investors profiles

The disposition patterns are significantly related with gender, income, and education, according to the ANOVA results. The male investors with high income and high education are more inclined to show the reversed DE, while female with low income and low education are inclined to show the disposition effect.

Table 3   The independent test of disposition patterns and demographic variables

gender	male	female	total				χ^2(p-value)
reverse DE	253	144	397				52.222
	23.6%	13.5%	18.6%				(0.000)
symmetry	369	328	697				
	34.5%	30.8%	32.6%				
DE	449	593	1042				
	41.9%	55.7%	48.8%				
total	1071	1065	2136				
age	Lessen than 29	30~39	40~49	50~59	More than 60	total	
reverse DE	46	121	115	92	21	395	11.004
	20.2%	20.4%	17.6%	18.7%	12.5%	18.5%	(0.201)
symmetry	80	185	229	148	55	697	
	35.1%	31.3%	35.1%	30.1%	32.8%	32.7%	
DE	102	286	308	252	92	1040	

	44.7%	48.3%	47.2%	51.2%	54.8%	48.8%		
total	228	592	652	492	168	2132		
income	Less than $0.2M	$0.21~1M	More than $1M	total				
reverse DE	62	221	82	365				12.572
	15.6%	18.0%	24.9%	18.7%				(0.014)
symmetry	128	392	105	625				
	32.2%	32.0%	31.9%	32.0%				
DE	208	614	142	964				
	52.3%	50.0%	43.2%	49.3%				
total	398	1227	329	1954				
education	low	middle	high	total				
reverse DE	108	99	189	396				14.286
	16.2%	17.7%	21.1%	18.6%				(0.006)
symmetry	203	179	310	692				
	30.4%	32.0%	34.6%	32.6%				
DE	356	282	397	1035				
	53.4%	50.4%	44.3%	48.8%				
total	667	560	896	2123				

As for the DE value of each geographic sector, the average DE of female is 0.37, significantly higher than that of male (0.22), consistent with the previous empirical findings (such as Feng and Seaholes, 2005; Kuo, Kuo, Chiu and Fang, 2005, etc.). According to the results of ANOVA, there are significant differences among different groups of age, income and education ($P < 0.05$), and most are linearly related. Roughly speaking, the investors of older with lower income and lower education demonstrate higher DE. In addition, the disposition effect is more obvious for those living in the southern Taiwan than those in north.

Table 4   ANOVA of DE and Demographic Variables

variable	Item	samples	Average DE	F	p-value
gender	Male	1071	.22	87.290	.000
	Female	1065	.37		
Age	Less than 29	228	.19	8.194	.000
	30~39	592	.27		
	40~49	652	.28		
	50~59	492	.35		
	More than 60	168	.41		
Location	Northern TW	1224	.28	2.469	.085
	Middle TW	439	.29		
	Southern TW	476	.34		
Income	Less than $0.2M	398	.36	7.808	.000

	$0.2~1M	1227	.30		
	More than$ 1M	329	.23		
	Low	667	.36	17.136	.000
education	middle	560	.32		
	high	896	.24		

Table 5   The Difference Test on Study Time among the Demographic Sectors

Average studying time(I)	Average studying time (J)	Mean difference (I-J)	p-value	Average studying time(I)	Average studying time (J)	Mean difference (I-J)	p-value
Age				Location			
Less than 29	30~39	-.08404	.014	Northern Taiwan	Middle	-.00813	.741
	40~49	-.09269	.006		Southern	-.05254	.028
	50~59	-.15761	.000	Middle Taiwan	Northern	.00813	.741
	More than 60	-.21580	.000		Southern	-.04441	.128
30~39	Less than 29	.08404	.014	Southern Taiwan	Northern	.05254	.028
	40~49	-.00866	.728		Middle	.04441	.128
	50~59	-.07358	.006	Yearly income			
	More than 60	-.13176	.001	Less than 0.2M	0.2~1M	.06244	.014
40~49	Less than 29	.09269	.006		More than 1M	.12932	.000
	30~39	.00866	.728	0.2~1M	Less than 0.2M	-.06244	.014
	50~59	-.06492	.013		More than 1M	.06688	.014
	More than 60	-.12310	.001	More than 1M	Less than 0.2M	-.12932	.000
50~59	Less than 29	.15761	.000		0.2~1M	-.06688	.014
	30~39	.07358	.006	Education			
	40~49	.06492	.013	Lower	Middle	.04885	.052
	More than 60	-.05818	.138		Higher	.12862	.000
More than 60	Less than 29	.21580	.000	Middle	Lower	-.04885	.052
	30~39	.13176	.001		Higher	.07977	.001
	40~49	.12310	.001	Higher	Lower	-.12862	.000
	50~59	.05818	.138		Middle	-.07977	.001

The impacts of investment study on the price changes of winners and losers

From the previous analysis we find the great variation on the disposition tendency. We use regression models to analyze the impact of the investment study intention on the disposition tendency. Table 6 is the estimation results of the influences of study time intention from three surveys (1,672 samples) and table 7 is that of both study time and information media from one survey (593 samples).

According to the estimation results, all models show statistically good fitness except for

the model of price changes for winning stocks. Moreover, the study intention has good explanatory power on the disposition effect. Despite the insignificance of the study coefficients in models 2 and 5 (study does not make investors keep winners longer, we find the DE value and the price changes of holding losers are negatively related with the investment study time and the number of information media. The less study efforts investors are willing to make, the more disposition effect they are subject to. The intention to study helps to mitigate the disposition effect and mainly due to its effect on promoting the loss realizations. We believe it is because the knowledge gained from studying effectively reduces the uncertainty on stock value and thus reduces the resistance of realizing losses. Hypothesis 2 and 3 are supported by the empirical evidences while hypothesis 1 is not.

The regression models also show the influences of demographical variables. Female and older investors with lower income and living in the southern Taiwan significantly demonstrate higher disposition effect. The data are from three surveys; we need to control the potential influence of investigation timing, so we put each survey into control by dummy variables. The coefficients are insignificant. In other words, the effects of investment study intention on lessening the disposition effect are robust.

Table 6   The impacts of investment study intention (time) on the disposition effects: regression models

	Y=DE		Y=return on gains		Y=absolute return on losses	
	Coeff.	P-value	Coeff.	P-value	Coeff.	P-value
intercept	0.292	0.000	0.176	0.000	0.473	0.000
Time of studying	-0.004	**0.000**	0.000	0.549	-0.003	**0.001**
Gender (1=mail, 0=femail)	-0.127	0.000	0.011	0.288	-0.112	0.000
Age	0.004	0.000	0.000	0.926	0.004	0.000
location(1=northern Taiwan, 0=southern)	-0.051	0.059	-0.007	0.590	-0.060	0.022
location(1=middle Taiwan, 0=southern)	-0.033	0.318	0.016	0.284	-0.018	0.566
income(NT$M)	-0.040	0.031	0.007	0.394	-0.034	0.055
education(1=middle, 0=lower)	-0.003	0.921	-0.010	0.471	-0.012	0.674
education(1=higher, 0=lower)	-0.033	0.219	-0.026	0.037	-0.061	0.018
survey(1=#2，0=#１)	-0.001	0.978	0.011	0.354	0.011	0.661
survey(1=#3，0=#1)	0.005	0.860	0.001	0.904	0.007	0.773
Adjusted R-squared	0.053		0.001		0.055	
F-statistic	10.292		1.143		10.644	
Prob(F-statistic)	0.000		0.326		0.000	
No. of samples: 1672						

Table 7 The impacts of investment study intention (time and media) on the disposition effects: regression models

	Y=DE		Y=return on gains		Y=absolute return on losses	
	Coeff.	P-value	Coeff.	P-value	Coeff.	P-value
intercept	0.330	0.000	0.211	0.000	0.541	0.000
Time of studying	-0.004	0.027	0.000	0.676	-0.004	0.016
Channels of studying	-0.005	0.012	-0.001	0.236	-0.006	0.002
Gender (1=mail, 0=femail)	-0.155	0.000	0.038	0.018	-0.117	0.001
Age	0.003	0.015	0.000	0.561	0.003	0.029
location(1=northern Taiwan, 0=southern)	-0.023	0.597	-0.044	0.026	-0.067	0.124
location(1=middle Taiwan, 0=southern)	-0.014	0.781	0.008	0.733	-0.006	0.901
income(NT$M)	-0.007	0.822	0.001	0.932	-0.006	0.851
education(1=middle, 0=lower)	-0.044	0.337	-0.014	0.491	-0.058	0.202
education(1=higher, 0=lower)	-0.065	0.126	-0.022	0.254	-0.087	0.040
Adjusted R-squared	0.065		0.020		0.065	
F-statistic	5.567		2.317		5.575	
Prob(F-statistic)	0.000		0.015		0.000	
No. of samples: 593						

CONCLUSIONS

Existing studies on the disposition effect is mainly of descriptive research. The empirical evidences consider it an error due to its disadvantage to the investment performance. However, few advices are given to investors to reduce the disposition effect, leaving it a research gap to be filled. This study proposes the study hypothesis and argues that studies help to reduce the stock value uncertainty and thus decrease investors' resistance to realizing losses or keeping winners longer. Through the nationwide surveys in Taiwan, we get 1672 effective samples to do the empirical analysis and find the study hypothesis is significantly supported. That is, if investors are more intended to make investment study, their disposition effects are lessened. And it is mainly due to the study effect on promoting loss realizations; the study intention fails to make investors keeping winners longer. We suggest it us because study leads investors to be more rational oriented and thus less driven by the sentiment regret averse. In a word, back to basics to make investment study is helpful to lessen the disposition effect.

Reference

Beaver, William H., 1968, The Information Content of Annual Earnings Announcements,

Journal of Accounting Research 6, 67-92.

Benartzi, S. and R. Thaler, 1995, Myopic Loss Aversion and the Equity Premium Puzzle, *Quarterly Journal of Economics* 110, 75-92.

Feng, Lei and Mark Seasholes, 2005, Do Investor Sophistication and Trading Experience Eliminate Behavioral Biases in Financial Markets? *Review of Finance* 9(3), 305-51.

Ferris, S. P., R. A. Haugen and A. K. Makhija ,1988, Predicting Contemporary Volume with Historic Volume at Differential Price Levels: Evidence Supporting the Tisposition effect, *Journal of Finance*, 43(3):677-697.

Glick, Ira, 1957, A Social Psychological Study of Futures Trading, Ph.D. Dissertation, University of Chicago.

Goetzmann, William N. and Nadav Peles, 1997, Cognitive Dissonance and Mutual Fund Investors, *Journal of Financial Research*, 20(2), 145–158.

Grinblatt, M., and M. Keloharju, 2001, What Makes Investors Trade? *Journal of Finance* 56, 589-616.

Grossman, Sanford J., 1976, On the Efficiency of Competitive Stock Markets Where Traders Have Diverse Information, *Journal of Finance* 31(2), 573–584.

John R. Nofsinger, 2002, *The Psychology of Investing*, Prentice Hall.

Kahneman, Daniel, and Amos Tversky, 1979, Prospect theory: An analysis of decision under risk, *Econometrica* 47, 263-291.

Kuo, Min-Hua, 2010, Does the Prospect Theory Necessarily Mean Disposition Effect? An Extension of Disposition Effect, Annual Meeting of the Northeast Decision Sciences Institute, March 26-28, Alexandria, Virginia, USA.

Kuo, Min-Hua, N. F. Kuo, Y. C. Chiu and P. H. F, 2005, Gender and Investment Behavior: on Taiwanese Individual Investors, *Journal of Financial Studies* 13(2), 1-28.

Lakonishok, Josef, and Seymour Smidt, 1986, Volume for Winners and Losers: Taxation and Other Motives for Stock Trading, *Journal of Finance* 41, 951-974.

Odean, Terrance, 1998, Are Investors Reluctant to Realize Their losses? *Journal of Finance* 53, 1775-1798.

Schlarbaum, Gary, G. Lewellen, G. Wilbur and Ronald C. Lease, 1978b, The Common-Stock-Portfolio Performance Record of Individual Investors: 1964–70, *Journal of Finance* 33, 429–441.

Schlarbaum, Gary, G., Lewellen, G. Wilbur and Ronald C. Lease, 1978a, Realized Returns on Common Stock Investments: The Experience of Individual Investors, *Journal of Business* 51, 299-325.

Shapira, Z. and I. Venezia, 2001, Patterns of behavior of professionally managed and independent investors, *Journal of Banking and Finance*, 25, 1573-1587.

Shefrin, H. and M. Statman, 1985, The disposition to sell winners too early and ride losers too long: Theory and evidence, *Journal of Finance* 40, 777-790.

Shiller, Robert J., 1999, Human behavior and the efficiency of the financial system, in: J. B. Taylor and M. Woodford (ed.), *Handbook of Macroeconomics*, edition 1, volume 1, chapter 20, 1305-1340. Elsevier.

Weber, M., and C. F. Camerer. The Disposition Effect in Securities Trading: An Experimental Analysis, *Journal of Economic Behavior and Organization* 33, 167-184.

THE VALUE OF SPIRITUALITY WITHIN THE WORKPLACE: A DISCUSSION AND PROPOSAL FOR RESEARCH

Maureen L. Mackenzie
Associate Professor of Management and Leadership
Townsend School of Business, Dowling College, Oakdale, NY 11777
mackenzm@dowling.edu

Gilles Servant
Graduate Student
Laval University, Quebec City, Canada
ServantGilles@yahoo.ca

ABSTRACT

This paper considers the definitions of workplace spirituality, its value, and how the defining foundational concepts are manifested within organizations today. Also, if there is value, what is the process of educating future leaders so that they are prepared to create and sustain workplace spirituality as part of the cultural fabric of their organizations? Last, there is a call for dialogue between management practitioners and theorists as to whether the phrase, *Workplace Spirituality*, can thrive within the management practitioner's vernacular. Perhaps it is more effective to refer to the workplace spirituality concepts under their covert labels and avoid using the controversial "S" word. **Keywords**: leadership, management, spirituality, religion, manager-employee interaction, organizational behavior, spiritual philosophies.

INTRODUCTION

The discipline of management has continuously responded to its changing environment. As issues of diversity, gender equality, etc., emerged as social issues, with the potential to influence employee productivity, leadership strategies adapted. For example, years ago a discussion of "stress at work" was perceived as soft, negative, and weak. In a short time the work-stress conversation found its way into mainstream management discussions.

The goals of management have not change; the foundational philosophy of the management discipline is getting the work done, achieving organizational goals, and meeting stakeholder needs. Often companies will suggest that their stakeholders include clients, shareholders, and employees, but employees often have the least influence in the boardroom. Yet, without engaged employees, neither shareholder nor client will be satisfied (Jain 2011).

Management tactics may need to once again respond to recognize the modern worker and the emerging discussion of spirituality and the workplace. (Pawar 2009; Driscoll and McKee 2007; Sheep 2006; Leigh 1997). Essential to this discussion is the understanding that Spirituality is not religion (Leigh 1997). Rather, spirituality is part of the fabric of individuals, and individuals populate organizations. Managers are certainly not expected to become religious leaders, but rather to recognize the need many employees have for meaningful work and to feel connected to

each other and to their environment. (Milliman, Ferguson, Trickett and Condemi 1999; Pawar 2009; Sheep 2006).

This paper considers the definitions of workplace spirituality (WS), lessons gained from the spiritual philosophies, the value of WS, and how the defining WS foundational concepts are manifested within organizations today. Also, if there is value, what is the process of educating future leaders so that they are prepared to create and sustain workplace spirituality as part of the cultural fabric of their organizations?

BACKGROUND

Workplace spirituality emerged with perceivable interest in the early 1990s with books and articles published in respected journals, such as: *Journal of Managerial Psychology*, *Journal of Management Inquiry*, *Journal of Management Education*, and *Journal of Organizational Change Management* (Giacalone and Jurkiewicz 2010). The introduction of the Special Interest Group (SIG), *Management, Spirituality, and Religion*, within the Academy of Management, provided a respected venue for emerging research.

Pawar's (2009) research identified four foundational concepts upon which workplace spirituality can stand. They are Transformational leadership, Organizational Citizenship Behaviors, Organizational Support, and Procedural Justice. Pawar tracked these four concepts back to their founding disciplines, then their emergence in the field of Organizational Behavior, and finally linked them together within the emerging discipline of Workplace Spirituality (WS). Pawar's findings place WS as an important sub-discipline of the well-established discipline of Organizational Behavior.

Giacalone and Jurkiewicz show the research links between WS and other disciplines. The links include: Agency theory (self direction, individual existence), Ethics and social responsibility (moral behavior), Materialism (importance of possessions), Work values (importance given to particular work-related outcomes), Workaholism (excessive time spent at work which leads to detrimental life outcomes), Servant leadership (leadership that serves others), and Work-life balance (establishing greater equality in time spent at work and home), to name a few. (Giacalone and Jurkiewicz 2010, 17-18).

Leigh (1997) highlights stories of corporate programs and CEO testimony that suggests that "there is a new connection happening in many organizations between employees and management that is resulting in a happier workforce and real bottom-line improvements." (Leigh 1997, 26). Briskin suggests that "soul and self are inextricably linked to who we are both inside and outside of the workplace" (Cohen 1997, 57).

On the other hand, Pava (2003) concluded, after the review of popular books on WS, that satisfactory models of legitimate workplace spirituality have *not* been found because of the restrictive definitions of spirituality. Unfortunately, spirituality is too closely linked to religion, which has no place in the business world. Some feel that drawing spirituality into the workplace is divisive, excludes those who are not part of a particular tradition, can cause abuses, and may

lead to the manipulation of employees (Mitroff and Denton 1999; Karakas 2010, Giacalone and Jurkiewicz 2010).

Separating workplace spirituality (WS) from religion has been a strong focus within the literature. Giacalone and Jurkiewicz stress that WS must be separated from the "faith blanket in which it is frequently cloaked" (2010, 4). They further suggest that the work to separate WS from religion is "equivalent to surgically dividing conjoined twins" (Giacalone and Jurkiewicz 2010, 4).

Ultimately the workplace spirituality literature can be summarized into the employee's search for meaningful work, a workplace community, and transcendence of self-interest. (Karakas 2010; Pawar 2009).

DEFINITIONS FOR SPIRITUALITY AND WORKPLACE SPIRITUALITY (WS)

There are a range of definitions for workplace spirituality (WS). Karakas (2010) stated that there are more than 70 definitions of WS. What makes it more difficult is that there are no standard definitions for the root word, spirituality. Definitions of spirituality include concepts such as inner consciousness; enlightenment; "a specific form of work feeling that energizes action;" and a worldview with a path (Karakas 2010, 91). Karakas distinguishes spirituality from religion by characterizing it as a "personal, inclusive, non-denominational, universal human feeling; rather than an adherence to the beliefs, rituals, or practices of a specific organized religious institution or tradition" (Karakas 2010, 91).

Ultimately, spirituality can be referred to as the deepest values and meanings by which a person lives. Spirituality often is characterized by a need to feel connected to other people, the environment, and to some higher reality. Workplace leadership would *not* address the connectedness an individual worker may or may not have to a higher reality, but certainly it would address the employee's need to connect to a meaningful organizational goal, to the environment, and to others in the workplace (Pawar 2009; Duchon and Plowman 2005). An employee spends a majority of his or her waking hours at work; as a result, this desire to feel connected to others, and to the environment, will pervade the workplace. Leaders must articulate how their organization's work is beyond economic purposes and serves a larger social purpose. This articulation of transcendence will help the employee know how he or she is instrumental in the organization and how his or her work supports that social purpose (Giacalone and Jurkiewicz 2010).

Daniel (2010) pulled together the following definitions Workplace Spirituality (WS):

- The recognition that employees have an inner life that nourishes and is nourished by meaningful work that takes place in the context of community (Ashmos and Duchon 2000, 137).
- A framework of organizational values evidenced in the culture that promotes employees' experience of transcendence through the work process, facilitating their sense of being connected to others in a way that provides feelings of completeness and joy (Giacalone and Jurkiewicz 2003, 91).

- Spirituality in the workplace is an experience of interconnectedness and trust among those involved in the work process, engendered by individual goodwill; leading to the collective creating of a motivational organizational culture, epitomized by reciprocity and solidarity and resulting in an enhanced overall performance, which is ultimately translated in lasting organizational excellence (Marques 2005, 285).

Grzeda and Assogbavi (2011, 239) offer the following working definition for spirituality in management: It "consists of those management behaviors driven entirely by spiritual values, teachings, or beliefs, regardless of their source, creating connections between behavior and personal spiritual meanings which are cognitively acknowledged and affectively valued by the manager."

Marques, Allevato and Holt (2008, 85) developed a working definition that states, "spirituality in the workplace is an experience of interconnectedness among those involved in the work process, initiated by authenticity, reciprocity, and personal goodwill engendered by a deep sense of meaning that inherent in the organization's work; and resulting in greater motivation and organizational excellence."

Giacalone and Jurkiewicz (2010, 5) suggest that the lack of consensus around a conceptual definition for workplace spirituality hampers scientific study. Leigh (1997, 27) further suggests that "no one definition can encompass this entire phenomenon" but it can be said that there is an emphasis by corporate leaders to develop values-based businesses and to "recognize the need to help create meaning and purpose for their employees, and to link an organization more closely with its workforce."

SYNONYMS FOR CONNECTIVITY AND SPIRITUALITY IN THE WORKPLACE

As an organizational behavior professor, could you walk into your classroom and ask your graduate students if they value: fairness in organizations (procedural justice), the ability of their leaders to inspire employee loyalty, the influence to get employees to collectively work toward the higher order goals of the organization (transformational leadership), to create an organization that cares about the employee's well-being and contributions (organizational support), and sustains employee actions that benefit others without seeking a direct reward (organizational citizenship behaviors)? We suggest that those graduate students would agree whole-heartedly. Yet, would you be as successful to gain these graduate students' agreement if you asked them if they support workplace spirituality? Research shows that people are mixed on how they view the concept of spirituality and its relationship to formal religion. Mitroff and Denton (1999) found that 60% of their subjects viewed spirituality positively, while the remainder held negative views of spirituality as religion. Leigh (1997) found after interviewing CEOs that they prefer *not* to use the "S" word, referring to spiritual or spirituality.

Pawar (2009) found that the four OB concepts of procedural justice, transformational leadership, organizational support, and organizational citizenship behaviors, are foundational to workplace spirituality. Relevant is the insight that leaders "do *not* explicitly use the term spirituality or workplace spirituality. Rather, they include a different set of explanations for employee transcendence of immediate, narrow, or economic self interests" (Pawar 2009, 254). Pawar

further suggests that synonyms for workplace spirituality (WS) include: employee transformation, social exchange, reciprocity and more (Pawar 2009, 254).

In summary, theorists and practitioners have created new terms that describe the support of the modern workplace. Concepts such as team building, social responsibility, diversity, participative management, servant leadership, job enrichment, theory Y, workplace integrity and of course, business ethics, can be weaved together to describe our professional and human need to transcend self-interest, feel part of something larger, and to connect to other human beings, even within the workplace (Pawar 2009, Sheep 2006). The term spirituality mixed into the workplace vernacular may find sufficient resistance to keep it solely within the pages of academic journals.

IS THERE VALUE FOR SPIRITUALITY WITHIN THE WORKPLACE?

It is not our intent to tie religion to workplace spirituality as we do not expect religion to enter the work environment. The U.S. founders had the insight that religion was personal and our individual rights were worthy of protection. No collective religious beliefs should influence the capitalistic nature of a country that allows each person to benefit from the sweat of his or her own brow. That being said, we *do* wish to discuss the lessons which the spiritual philosophies and religions can offer to management theorists and practitioners. Also, there are other management philosophies that address these same issues of connectivity yet do not use the word spirituality. Business ethics, social responsibility, organizational culture and climate, transformational leadership, organizational citizenship behavior, organizational justice (Pawar 2009), valuing diversity, sustainability, connectedness, and conscientiousness, are all terms that have found a home in the workplace. No fear is attached to using these words and phrases at work; they all represent valued enhancements to organizational life.

Sheep (2006) cited Mitroff and Denton's Spiritual Audit of Corporate America (1999) when suggesting that there is evidence that leaders who consider their organizations to be spiritual, also see their businesses as more ethical. Garcia-Zamor (2003) suggested that "bringing spirituality into the workplace could create a different organizational culture in which the employee would be more satisfied and would have an improved performance" (Daniel 2010, 445). Although some scholars suggest that spirituality should be an end in itself and not be used to improve profits, there is a trend toward recognizing WS as tool to engage employees and improve performance, ultimately impacting organizational results (Karakas 2010, 92).

Karakas (2010) reviewed close to 140 workplace spirituality (WS) articles, specifically seeking insights on whether WS supports performance results at work. The findings are well summarized suggesting that WS enhances the employee's quality of life and well-being, helps the employee find meaning and purpose at work, and offers the employee a feeling of being connected to a community. The culture of the organization has a primary influence on the employee's perception of community, connectedness, and spirituality. Reder's (1982) conclusions were cited as "unequivocally" suggesting that "spiritually-based organizational cultures are the most productive and that by maximizing productivity they confer organizational dominance in the marketplace" (Giacalone and Jurkiewicz 2010, 6). Further it has been shown that cultural factors related to workplace spirituality override the economic-political influences on the employees' productivity (Giacalone and Jurkiewicz 2010).

Giacalone and Jurkiewicz (2010, 20) suggest that scientific research within the discipline of WS will bring improvements to workplace life, as long as they are "unfetted by legal and religious phobias." They identified the potential benefits that may emerge from successful collaboration and controlled study of WS and its impact on organizations. These potential areas of benefit include: Recruitment (how do we recruit candidates who desire a spiritual work environment), ethics (increased ethical decision making), creativity and innovation (are spiritual workers more creative?), public relations (repercussions for those accepting or rejecting spirituality), leadership (is job satisfaction influenced by spirituality), work-family issues (what is the relationship to work-life balance and spiritually at work?), and motivational/reward systems (are spiritual workers motivated by different factors than non-spiritual employees?).

LESSONS FROM SPIRITUAL PHILOSOPHIES AND RELIGIONS

We have stressed the controversy around the linkage of religion to spirituality within the workplace. There is no doubt that practitioners should not use religious rituals within their business practices, but the lessons that can be gained from these rich and deep spiritual philosophies can be translated into practices that business leaders can use to draw together the spiritual beings at work.

Sacred Pipe = Management Consensus

The Lakota's Sacred Pipe is the instrument used to ensure that all who meet come in peace. The Lakota's are a native tribe in the US, members of the Sioux tribe; their strength severely diminished since the 19th century. This proud tribe brings rich tradition; we can learn from its rituals. When a group of men, including the medicine man, would meet to make tribal decisions, especially the development of their treaties with the US government, the Sacred Pipe was central to the decision making and consensus-gaining process. The smoke from the Sacred Pipe travels upward to the six ancestral grandfathers; it represents the respect toward the decisions that had been made in the past (Neihardt 2008, 4). Neihardt cites Walker's book, *Lakota Belief and Ritual,* discussing how smoking the pipe brings peace among smokers and helps them remember the decisions made by their ancestors. At the beginning of a meeting, the members would pass the Sacred Pipe, allowing each man to reflect on the decision that was to be made that evening. (Neihardt 2008).

Each tribe had its own Sacred Pipe. The host of the meeting would present the pipe to those attending the meeting. The pipe's decorations represent the establishment of the tribe. When you see the pipe, you understand which tribe's leader is hosting the meeting (Neihardt 2008, 3). Smoking the pipe also represents to each man that he has the responsibility to be sincere, honest, and devoted to bring forward the best decision for the common good. Though the men would debate and perhaps argue as the decisions were being developed and made, once consensus was gained, all members smoked the Sacred Pipe to seal the decision. Most important, the pipe represents the solidarity among those who contributed to the decision. As everyone leaves the meeting, all are in common agreement that the best decision had been made (Neihardt 2008, 5).

In parallel, the Sacred Pipe still exists today in real-world business decision making. No, there is no smoking at the meetings, but there are minutes gathered which allow those in attendance to reflect on prior decisions. Also, attendance is taken so that the identity of each member at the meetings is recorded and made public for all to see. Most important, managers who fight it out during the meeting may find a social tradition at the end of the meeting to celebrate the consensus gained among all parties. Regardless of where the individuals stand during the debate, everyone walks out of the meeting supporting the group decision. Cocktails, dinner, cigars on a fancy balcony, may all be part of corporate celebrations that are reflective of the spirituality among the decision makers to work toward the common good.

Spiritual Reflection = Managerial Empowerment

Moberg and Calkins (2001) used the writings of St. Ignatius Loyola (1491-1556), the founder of the Society of Jesus, to draw out the essential ingredients that can be translated into a process of managerial reflection that may lead to better decision-making and true empowerment. St. Ignatius' process of reflection is over 500 years old. These Spiritual Exercises were developed for Christian believers; the structure of these exercises used "imagination, role-modeling, and the integration of reason with the emotions" as a means to enhance the spiritual reflection of these Christian disciples (Moberg and Calkins 2001, 257).

Moberg and Calkins bring this process forward to us today as a four-phase process that leaders can use in their daily routines. The process engages the manager's emotions and imagination. It is a strong model for reflection that begins with the individual's experiences and moves toward reflection, conceptualization, and then experimentation. Moberg and Calkins (2001) cite the testimony of Bell and Howell CEO, Charles Percy. He would reflect for 45 minutes each night before bed. This regular evening ritual is similar to the Hindu form of meditation. Percy's evening reflections were modeled after the exercises inspired by St. Ignatius' writings. His meditative reflections allowed him to consider specific questions or simply focus on conflicts, expectations, attitudes, and whatever else may have been on his mind.

The essential part of the process is that it allows the manager or the employee to draw on his or her real-world experiences; these experiences may otherwise be lost in the hurried life-style of today's over-worked leaders.

The benefit of this reflective process is two-fold. These reflective practices can be used by leaders, but they can also be taught to employees. As employees are given these advanced tools, they will be more apt to offer solutions to workplace problems as they draw upon their experiences within the realm of quiet reflection. Essential is that the reflective process remain absolutely voluntary (Moberg and Calkins 2001). Once again, a process that has a spiritual undertone must be presented to the employees cleansed of any religious reference.

In conclusion, Moberg and Calkins (2001) propose that St-Ignatius Loyola's approach touches many of the themes which are central to contemporary business ethics (virtue, role modeling, emotional engagement, interpersonal relationships). Done well, the Spiritual Exercises offer clear- headedness to both leaders and their employees.

Eastern Spiritual Philosophies = Modern Management Vision Statements

Corner (2009) extends the theory on the relationship between workplace spirituality and business ethics by integrating lessons from ancient Eastern spiritual traditions, such as yoga. The yogic self control ("Yamas" in Sanskrit) is comprised of five practices for harmonizing and deepening social connections in the Workplace within any community.

Most of the Eastern cultures are aware of yoga (in Sanskrit it means to join). Yogic practices include Asanas (static posture), breathing practices (pranayama), and meditation (Sivananda 2007). These Yogic practices are relatively new to the West, when compared to the ancient philosophies of Yama. The practice of yoga achieves a permanent state of peace, joy, and selfless dedication to humanity (Satchidananda 2004).

The five practices of Yama come from an uncertain period over 3000 years old (Vishnu-Devananda. 1981). They are designed to harmonize a person's social interactions. Harmonized relationships are needed because any discordance disturbs the mind (Corner 2009). Great examples of those who practiced the Yamas is Mahatma Gandhi, who practiced non-violence, "Ahima." Charles Darwin demonstrated the use of truthfulness and honesty "Asteya," when reflecting authorship of his book, "Origin of The Species" in 1856.

Driscoll and McKee (2007) also bring forward the connections between spirituality in the workplace, ethical organizational culture, and authentic transformational leadership. They provide examples of spiritual storytelling by leaders that give employees the capacity to understand the goals and objectives of the enterprise. A double win-win situation occurs: the leader becomes more attainable and may feel more efficient in his or her work. Simultaneously the work is accomplished by a conscious employee who's more aware of the businesses vision and values.

These spiritual philosophies can be foundational for the development of organizational mission and vision statements. Values-based organizations that build their processes on such spiritually healthy philosophies will enjoy longevity that will be felt for decades as the organization thrives in the changing environments which are prevalent in Capitalistic cultures.

Jainism ≠ Nazism

The philosophy of Jainism has similarity to Buddhism, Hinduism and the Tibetan philosophy. The religion prescribes a path of total non-violence towards all living things. For example, a root vegetable is not harvested, because it would be considered a practice of killing, which is against the philosophy of Jainism. We recognize that this practice may only exist with individuals who are fundamentalists, but demonstrates the full range of this non-violent philosophy. The existence of Jainism goes back over 2500 years. It's the smallest religion but among those considered most important (Foot 1999).

The Religious symbol of Jainism is the Svastika. Its definition is *the state of goodness*. The Svastika represents the "Kosmos;" the four worlds: men, gods, fauna, and the world of devils. The moral code of that religion possessed five vows (non-violence, sincerity, honesty, chastity,

and non-attachment of material). The form was adopted by Adolf Hitler and the Third Reich as its political party insignia in 1920. After World War II, this emblem was almost completely eliminated in the West due to its connection to the evil outcomes of the Nazi's. This is in complete opposition with the pure philosophy of Jainism. (Changing Times 2005).

The lesson is that a leader may be able to use the values of a spiritual philosophy to manipulate followers. This is warning that aligns with the concerns of some WS authors on the use of spirituality in the workplace.

Extreme Example of Religion at Work = Warning To Practitioners

Proulx's (2006) work focused on the presence of religion in the workplace. Though rare in Quebec, she did find more instances in the USA, where religious artifacts and practices were explicitly present in the workplace. She articulated in detail the story of one Canadian company, Tomasso Corp. This company produces frozen Italian meals and is owned by Mr. Jean-Robert Ouimet, who is a conservative Catholic. Proulx documented his model for leadership, which he had developed while earning his doctorate. Mr. Jean-Robert Ouimet was so confident in his business philosophy, that he said, "I challenge anyone to be successful in business without a spiritual presence." He further stated, "Without God, we don't go far" (Proulx 2006). He had no constraint in openly promoting his own values (Driscoll and McKee 2007). Proulx's (2006) work revealed less spirituality and more explicit religious presence within Ouimet's organization. Mr. Ouimet had developed his philosophy in the form of a Management Guide, which he titled, "Gold Book." The "Gold Book" (Le Livre Doré) was a direct summary of his doctoral thesis, which was titled, "Management Tools for Maintaining Happiness and Productivity." Proulx found religious artifacts in the form of religious posters, a dedicated room where employees could pray, the presence of prayer before board meetings, and the ability for employees to volunteer at charity organizations during working hours. Proulx commented that Mr. Ouimet even required a moment of silence before his meetings with her.

Mr. Jean-Robert Ouimet was very clear in his hiring practices. He was not shy in describing his hiring philosophy, which is, "To work for me, you must believe; otherwise you won't be happy working in my business." Before hiring a new employee, he would meet with both the employee candidate and his or her partner in a social setting. This provided Mr. Ouimet an opportunity to fully evaluate the spiritual life-style of the candidate. After hearing of this practice, Proulx asked Mr. Ouimet what may happen if the employee candidate arrived at the lunch meeting with his same-sex partner. Mr. Ouimet did not directly answer; there was complete silence. Ultimately his response to Proulx was "no homosexual has applied to my business." He did acknowledge that it may happen one day. He finished by saying, "God loves everyone" (Proulx 2006).

Proulx did not find any conflict within the Tomasso Corp., but it is most likely due to the owner's hiring practices that resulted in a like-minded work group of practicing Catholics. Everyone is on the same page. As of the writing of this paper, we are not aware of any filed discrimination suits. Oddly, the conflict arose from the owner's son, Mr. Robert Ouimet, Jr. Proulx suggested that observers of Tomasso Corp. were perplexed by the owner's explicit display of religiosity. While Ouimet Jr. was working for the father's parent company, he had

confided to a journalist that he felt that religion has no place in a commercial business. That conflict between the father and son nearly landed them in court.

On the opposite end of the spectrum Proulx (2006) offers the example of Cascades Corporation, which considers spirituality less important. For example, if an employee candidate includes on the application any reference to spirituality as a component of their personal life-style, there is a good chance that this person may not get an interview. The company-position is that an employee's spiritual activities should be private. The company prefers focusing on employee well-being and treating each with dignity and respect.

Proulx (2006) concluded by saying that there is a lack of consensus when discussing the meaning of Workplace Spirituality. WS means different things to different people and Proulx's extreme examples demonstrate this lack of agreement. The lesson to practitioners is to strike the balance where spirituality is inclusive, secular, non-threatening, and most important, non-discriminating.

ARE FUTURE LEADERS AND MANAGERS LEARNING ABOUT THE VALUE OF WORKPLACE SPIRITUALITY?

Managers gain their knowledge and skills from a variety of sources, some directly from an industry providing a lifetime of hands-on experience, others learn business skills from working in their families' businesses, others start businesses as entrepreneurs and learn as they grow, and many come from business schools. Individuals who intentionally plan a career in business will often enter a graduate business program leading to the Master of Business Administration (MBA) degree. The International Assembly for Collegiate Business Education (IACBE) and The Association to Advance Collegiate Schools of Business (AACSB) are the specialized accrediting bodies for institutions that offer business education programs. Although business education includes the disciplines of Management, Marketing, Accounting, Finance and Managing Information Systems, this article focuses on the discipline of management, which is a dominant discipline within graduate level business education (Mackenzie and Smith 2009). An appropriate management curriculum, as defined by the IACBE includes 1) management principles, 2) organizational behavior, 3) human resource management, and 4) operations management. The IACBE standards are established to "ensure that students understand and are prepared to deal effectively with critical issues in a changing global business environment" (IACBE Accreditation Manual 2006, 73).

The Workplace Spirituality literature is nestled within the discipline of Organizational Behavior (Pawar 2009), which is classified as a primary knowledge area for management education (IACBE Accreditation Manual 2006). It may therefore be assumed that workplace spirituality is part of the business curriculum. We suggest that is an assumption that cannot be made. We further suggest that it is unclear as to whether WS is explicitly discussed in business school. Perhaps the WS concepts are being taught, but without using the "S" word. Thompson (2000) asks whether you can train people to be spiritual. He suggests that "trainers who help people develop supervisory skills, teach them how to deal with difficult colleagues, or coach them on their career paths are providing training in spirituality – just under different names" (Thompson 2000, 18). His research had found that High Schools are starting to provide spiritual training

under the Character Education Movement, which gained impetus after the shooting at Columbine High School in Colorado (Thompson 2000).

Grzeda and Assogbavi (2011) suggest that business education still focuses heavily on teaching future leaders how to pursue profitability and exploiting every opportunity for continuous growth. They draw from the healing disciplines, and more specifically Tikkun olam, to recommend a conceptual approach for management education. Briefly, Tikkum olam, translated "repair of the world," encourages social action directed at the community or society with the intent of healing social ills. Grzeda and Assogbavi (2011, 243) suggest that management education needs to be authentically transformed by "essentially substituting an orientation toward *repair of the world* to replace profitability as the ultimate business goal."

Before we move forward with a transformation of management education, we recommend that we look at what *is* being taught in today's graduate programs. We recognize that general ethics and philosophy is traditionally taught within the general education curriculum of undergraduate programs. Business ethics is most likely integrated into the business and management curriculum. Yet, we suspect that spirituality is limited to the academic disciplines of theology and philosophy. Since management is the science of oversight, most often of labor, which translates to mean people, we suggest that managers should actively learn to engage and manage the spiritual person. Our proposal for research will allow us to better understand whether our future leaders are receiving the knowledge and skills they will need to fully engage their employees.

A PROPOSAL FOR RESEARCH

The purpose of this research proposal is to gain a better understanding of what WS concepts are being taught to graduate students within business education programs. The study is designed to be exploratory in nature (Morse and Richards 2002, 27-28) and will use descriptive statistics to describe the data collected, and qualitative methods to draw meaning from the data and to extract tentative conclusions and recommendations (Kidder 1981, 103). Any human subject research will include approval by the Dowling College IRB.

The study will be conducted in three stages. Stage *one* will include a review of the business curriculum for a random sample of graduate business programs. Stage *two* will include a questionnaire to graduate professors that will probe and explore the results from the stage one syllabus review. Stage *three* will include a questionnaire to current graduate students to explore whether they have observed any reference to spirituality within their current or previous workplaces. These students will be provided definitions of WS and asked whether they have seen any such behaviors demonstrated within their current or previous work environments.

Stage One Sample

The population and sampling frame for Stage One is the colleges included in the most recent publication of the *US World Report on Best Colleges*. The non-probability based, purposive sample drawn from this sampling frame will be randomly selected. Ten colleges with graduate business programs will be drawn from the group of *most* selective schools. Ten colleges with

graduate business programs will be drawn from the group of *more* selective schools. Ten colleges will be drawn from the group of *selective* schools. The sample size will be a total of thirty graduate business programs in the United States.

Stage One Program Review

The program website for the graduate MBA program will be used to gather stage one data. The following data elements will be collected:

- Is there any required course in the curriculum which includes the term spirituality in its title?
- Is there any elective course in the curriculum which includes the term spirituality in its title?
- Review the course description for any leadership courses and organizational behavior courses to determine if the term spirituality is included in the description.
- Capture the name and contact information for the professors teaching Leadership and Organizational Behavior courses within the program.

Stage One Syllabus Review

Search for the syllabus for the Organizational Behavior course offered within the graduate program. Review those syllabi for the following:

- Any evidence that Workplace Spirituality is being covered within the course.
- Any evidence that Workplace Spirituality – *like* concepts are being covered within this course. This will include concepts such as: ethics, consciousness, transformational leadership, organizational citizenship behaviors, servant leadership, etc.

Stage One Analysis

Descriptive statistics will be captured and reported.

Stage Two Sample and Questionnaire

One hundred percent of the professors teaching either leadership or organizational behavior will be contacted to explore and probe the results gathered from stage one. These professors will be asked to comment on the workplace spirituality concepts that are discussed in their classes. Open-ended questions will be used to gather rich data.

Stage Two Analysis

Content analysis will be used to reduce the data from the responding graduate professors and to draw tentative conclusions.

Stage Three Sample and Questionnaire

A sample of convenience of 50 current graduate MBA students will be provided descriptions of workplace spirituality definitions. The students will be asked to describe any observed behaviors or references they have experienced at their current job or any previous job.

Stage Two Analysis

Content analysis will be used to reduce the data from the responding graduate students and to draw tentative conclusions.

OPEN DISCUSSION FOR MANAGEMENT THEORISTS AND PRACTITIONERS

The literature suggests that there is a lot of controversy within the emerging field of Workplace Spirituality. Many scholars' basic aim is to "make the area of spirituality at work research more legitimate and mainstream to organizational studies" (Karakas 2010, 92). The purpose of our paper is to invite open discussion on the topic of spirituality in the workplace. To spur discussion, we have presented knowledge drawn from the literature as well as lessons that have emerged from the spiritual philosophies. We have also proposed a research process that will explore academic business programs to determine whether workplace spirituality has entered into the academic preparation of future managers. Perhaps we can now apply the once considered *New Age* philosophies to current workplace strategies. On the other hand, perhaps it more effective to continue to train and use the concepts of workplace spirituality under their covert labels and avoid the controversy in using the "S" word.

REFERENCES

Ashmos, D. and D. Duchon. 2000. Spirituality at work: Conceptualization and measure. *Journal of Management Inquiry,* 9(2), 134-145.

Changing Times. 2005 (May 6). *The Statesman.* New Delhi. Retrieved January 11, 2011, from Research Library. (Document ID: 833394131). p. 1.

Cohen, Sacha. 1997 (March). Review of the book, *Stirring of the Soul in the Workplace,* by Alan Briskin. *Training & Development,* 51(3), 57-58.

Corner, Patricia Doyle. 2009. Workplace spirituality and business ethics: Insights from an Eastern spiritual tradition. *Journal of Business Ethics,* 85, 377-389.

Daniel, Jose Luis. 2010. The effect of workplace spirituality on team effectiveness. *Journal of Management Development,* 29(5), 442-456.

Driscoll, Cathy, and Margaret McKee. 2007. Restorying a culture of ethical and spiritual values: A role for leader storytelling. *Journal of Business Ethics*, 73, 205-217.

Duchon, D. and D.A. Plowman. 2005. Nurturing spirit at work: Impact on work unit performance. *The Leadership Quarterly,* 16, 807-833.

Foot, John Edward Andrew. 1999. Non-violence in the Indian religious tradition. M.A. dissertation, University of Windsor, Canada. Retrieved January 11, 2011, from Dissertations & Theses: The Humanities and Social Sciences Collection.(Publication No. AAT MQ52548).

Garcia-Zamor, J.C. 2003. Workplace spirituality and organizational performance. *Public Administration Review,* 63(3), 355-363.

Giacalone, R. and C.L. Jurkiewicz. 2003. "Toward a science of workplace spirituality" In Giacalone, R. A., and C.L. Jurkiewicz (eds) *Handbook of Workplace Spirituality and Organizational Performance.* Armonk, NY: M.E. Sharpe.

Giacalone, R. and C.L. Jurkiewicz. 2010. "The science of workplace spirituality" In Giacalone, R. A., and C.L. Jurkiewicz (eds) *Handbook of Workplace Spirituality and Organizational Performance, 2nd edition.* Armonk, NY: M.E. Sharpe.

Grzeda, Maurice and Tov Assogbavi. 2011. Spirituality in management education and development: Toward an authentic transformation. *The Journal of American Academy of Business,* 16(2), 238-244.

IACBE Accreditation Manual. 2006 (September). IACBE. Section 3.2 Common Professional Components. URL: http://www.iacbe.org/ (see link for accreditation manuals and forms). (Last viewed November 19, 2006.)

Jain, Dipak. 2011. Presentation to attendees of the Sivananda Ashram workshop on January 2, 2011, Nassau, Bahamas. Note: Dr. Dipak Jain is the former Dean of the Kellogg School at Northwestern University, Evanston, Illinois.

Karakas, Fahri. 2010. Spirituality and performance in organizations: A literature review. *Journal of Business Ethics,* 94, 89-106.

Kidder, L. H. 1981. *Research Methods in Social Relations, 4th edition.* New York: Holt, Rinehart & Winston.

Leigh, Pamela. 1997 (March). The new spirit at work. *Training & Development,* 51(3), 26-33.

Mackenzie, M.L. and J.P. Smith. 2009 (Summer). "Management Education for Library Directors: Are Graduate Library Programs Providing Future Library Directors With the Skills and Knowledge They Will Need?" *Journal of Education for Library and Information Science.* 50(3), 129-142. ISSN: 0748-5786.

Marques, J. 2005. Socializing a capitalistic world: Redefining the bottom-line. *Journal of American Academy of Business,* 7(1), 283-287.

Marques, Joan F., Eugene Allevato, and Svetiana Holt. 2008 (Winter). Linking quality management and emotional intelligence to workplace spirituality. *Organization Development Journal*, 26(4), 81-88.

Milliman, J., J. Ferguson, D. Trickett and B. Condemi. 1999. Spirit and Community at Southwest Airlines. *Journal of Organizational Change Management,* 12(3), 221-233.

Mitroff, I. I., and E. A. Denton. 1999. *A Spiritual Audit of Corporate America: A Hard Look at Spirituality, Religion, and Values in the Workplace.* San Francisco: Jossey-Bass Publishers.

Moberg, Dennis J. and Martin Calkins. 2001. Reflection on business ethics: Insights from St-Ignatius' Spiritual exercises. *Journal of Business Ethics*, part 2, 33(3), 257-270.

Morse, J. M. & L. Richards. 2002. *Readme First for a User's Guide to Qualitative Methods.* Thousand Oaks, CA: Sage Publications, Inc.

Neihardt, John G. 2008. *Black Elk Speaks.* Albany, NY: Excelsior Editions, State University of New York Press.

Pava, Moses L. 2003. Searching for spirituality in all the wrong places. *Journal of Business Ethics,* 48(4), 393-400.

Pawar, Badrinarayan Shankar. 2009. Some of the recent organizational behavior concepts as precursors to workplace spirituality. *Journal of Business Ethics,* 88, 245-261.

Proulx, Marie-Hélène. 2006. Spiritualité au travail, Dieu à l'ouvrage (God at Work). *Jobboom,* 7(February 2). URL: http://carriere.jobboom.com/marche-travail/dossiers-chauds/2006/06/01/1609804.html [Last viewed January 11, 2011]

Reder, M. W. 1982. Chicago economics: Permanence and change. *Journal of Economic Literature,* 20, 1-38.

Satchidananda, S. 2004. *The Yoga Sutras of Patanjali*. Buckingham, VA: Intergral Yoga Publications.

Sheep, M.L. 2006. Nurturing the whole person: The effects of workplace spirituality in a society of organizations. *Journal of Business Ethics,* 66, 357-375.

Sivananda, Sri Swami. 2007. *Hatha Yoga*. Tehri-Garhwal: The Divine Life Society. ISBN: 81-7052-192-0.

Thompson, William David. 2000 (December). Can you train people to be spiritual? *Training & Development,* 54(12), 18-19.

Walker, James R. 1980. *Lakota Belief and Ritual*. Raymond J. DeMallie and Elaine A. Jahner (eds). Lincoln: University of Nebraska Press.

Vishnu-Devananda. 1981. *Mantras and Meditations, 2nd edition.* New Delhi, India: Excel Book.

THE INSTITUTIONALIZATION OF ETHICS
IN CORPORATE GOVERNANCE
Between efficiency and legitimacy

MERCIER Samuel, Professor of Management,
University of Burgundy, **Samuel.mercier@u-bourgogne.fr**

Summary

This paper offers a summary status of the introduction of formal ethical approaches in corporations. After having sketched the outlines of the notion of organisational ethics, we analyse the challenges and objectives of these practices, and propose a typology of the formalisation and of the mechanism accompanying it. We then go on to question the relevance of these practices.

Introduction

The themes of organisational ethics, of corporate social responsibility and of sustainable development (these three concerns tend to overlap) are the focus of growing interest since the late 1980s. This groundswell (nourished by several successive vogues) testifies to the evolution in collective representations concerning the role of the corporation in society. The desiderata of consumers, investors, employees and, in general, citizens, reveal a growing social demand for greater integration of ethics in corporate life.

Ethics is branch of philosophy which, for many centuries, has been attempting to propose a guide for the conduct of man in society. The proliferation of speculations on the issue highlights the difficulty of drawing a distinction between ethical and unethical. Hence no consensual definition of ethics exists: depending on the author, the concept oscillates between reflection about the notion of Good and the enunciation of normative rules.

We propose to qualify, as ethical, reflection that occurs upstream of action and which aims to distinguish between good and bad ways of acting.

Individual ethical behaviour eludes any easy definition. The task is further complicated if one tries to clarify what is the ethics of an organisation, a college of individuals or a set of contracts (implicit and explicit), mingling different, indeed conflictual interests. In such a context, ethics is a field of tensions, lying between the interest of the corporation, the general interest and the interests of others. The challenge is to find a balance where the interests of the various *stakeholders* cannot be achieved simultaneously.

Stakeholders "are the individuals or constituencies that contribute, either voluntarily or involuntarily, to its wealth-creating capacity and activities, and that are therefore its potential beneficiaries and/or risk bearers" (Post, Preston and Sachs, 2002, p.19).

Organisational ethics defines the way in which the corporation integrates its values and key principles in its policies, practices and decision making processes. This also includes the quest for compliance with legal and internal rules.

Ethical approaches of the corporation: origins and objectives

Since the late 1990s, many major corporations have adopted formal systems to manage the organisational ethics dimension and to consolidate compliance with the law.

Many factors were combined at the emergence of these practices (see Van Parijs, 2004, p. 597):

- a technological evolution that increasingly impeded the precise control of the effort and of the honesty of an ever more crucial fraction of the corporate employees;

- rapid globalisation of economic activity, which makes its regulation by national laws increasingly difficult, and, *faute de mieux* and through fear of supranational state of regulations, spurs self-regulation by corporate executives;

- growing sensitivity of corporations to their brand image. Corporations are faced with difficult choices, whose economic and social consequences are scrutinised and judged by public opinion. The ethical reputation of the corporation, in other words, its integrity as perceived by the *Stakeholders*, has accordingly become an important part of its capital.

Moreover, a growing number of executives consider ethics as a motor of success, and have the conviction that only corporations that have succeeded in clarifying their values are likely to succeed.

Ethics approaches appear to meet a dual objective:

- in response to environmental pressures, they appear as a tool for protecting the interests of corporate management, and for managing risks of conflict with stakeholders. In the United States, for example, the *Federal Guidelines For Sentencing Organizations* adopted in 1991 have tended to blur distinctions between legal and ethical aspects;

- they represent a tool for internal regulation and to standardise the organisational culture, or to limit the discretionary behaviour of the employees.

This emphasises the importance of the institutional and organisational context in which ethical problems and actions are played out.

Beyond these objectives, two main logics can be identified in ethics policies:

- cultural logic tries to standardise behaviours by eliciting the adhesion of the stakeholders to the values and aims of the organisation and by encouraging ethical aspirations;

- coercive logic is directed towards compliance and is based on respect for the rules, strict monitoring of behaviour, and the presence of a punitive system in case of violation.

These two systems are not mutually exclusive. Corporations can simultaneously seek the internalisation of values and compliance with the rules.

Analysis of the ethics institutionalisation process

Formalisation is the keystone of any ethics approach and comprises three main elements: organisational values, action principles and rules of conduct. These components, present in variable proportions depending on the organisational culture, are often formalised in distinct documents.

More specifically, an important distinction appears between the formalisation of values and management principles, and the formalisation of rules of conduct.

In the former case, the documents essentially contain considerations on the nature of the corporation and its purpose, with the aim of eliciting the loyalty of its staff. The formalisation of values and action principles offers the assurance to the stakeholders that the corporation recognises this responsibility towards them and the legitimacy of their expectations.

In the latter case, the codes of conduct chiefly contain an ethical dimension. This is certainly the simplest and cheapest tool for announcing the ethical intensions of the corporation and clarifying the nature of the responsibilities devolving on its staff.

The codes are formalised as prescriptions and systematically address the following topics: reminder of legal requirements, importance of integrity in relations with the stakeholders and particularly clients, types of conflict of interest to be avoided, sensitivity to the issue of the confidentiality of information.

In large corporations, the recent development of several ethics mechanisms attests to the importance of the innovations introduced into management systems.

The production of ethics charters is merely one aspect of the more far-reaching process of institutionalisation of ethics in corporate governance. Thus the formalisation has progressively set up several mechanisms ensuring fulfilment of the commitments avowed and aimed to establish trustful relations with the stakeholders.

At the Board level the *ethics committee* has the job of monitoring the application of the formalised ethics. It becomes the guarantor of the corporate ethics policy and in charge of the overseeing the compliance of executive and employee actions with legal and ethical standards.

To do this, it has to set up adequate procedures to update the formal ethics policy, ensure its dissemination and its application.

Within the corporation, *ethics officers* (or compliance officers) are progressively emerging, named by the management. Their responsibility is to check the observance of the ethics policy, its in-house dissemination, and also to report to their superiors on the behaviour patterns observed and to advise the company staff.

Formalisation is also accompanied by *ethics training seminars.* To be effective, ethics training must help to improve the ethical reasoning of the individuals:

- by promoting the identification of the ethics dimension present in any decision;

- by legitimising this dimension as an integral part of the decision making process;

- by helping individuals to apply an ethics analysis to everyday operations to identify emerging problems;

- by securing an awareness of the ambiguous nature of the situation. The right answer is never obvious and can be open to discussion.

Training also serves to familiarise the staff with the importance of standards of conduct in work and fosters a better understanding of the formal ethics policy and its application. It develops an interactive environment in which problems of an ethical nature can be discussed freely.

The organisation of a *periodic ethics audit* is one way to check whether the values and rules announced are actually implemented daily. This offers the opportunity to examine closely the practices of the corporation and to identify the factors leading it (or which are potential

incentives) to unethical behaviour. Practice helps to materialise the corporation's determination to follow through in an approach to continuous progress.

Major corporations (particularly mass merchandisers) are also encouraged to conduct audits of their suppliers to ensure the observance of certain ethical or social criteria.

The elaboration of *social responsibility* (or sustainable development) *reports* is a recent and fast growing practice. This enables the corporation to communicate to the stakeholders the progress accomplished in the ethical field. It is not a matter of presenting an embellished image to the public to persuade it of the corporation's pure intentions, but rather to share the dilemmas. This practice has become mandatory in France, for corporations listed on a regulated market (NRE Law of 15 May 2001).

Are these ethics approaches legitimate and effective?

Many benefits are anticipated from these approaches: greater familiarisation of the staff with problems of ethics, integrity and empowerment, determination to address the problems openly, to disclose ethics violations, and to seek advice from the management to solve the dilemmas.

Their advantage largely resides in their presumed aptitude to connect attitude with action. From this standpoint, an approach focussed exclusively on compliance is liable to generate very little ethics commitment, since it is based on a minimalist motivation: to elude punishment.

Similarly, the formalisation of a restricted ethics can lead to dangerous manipulation.

French employees question the validity of the standards thus announced by the corporation and appear to be sceptical concerning its wish to materialise the formalised commitments. Most of them are also hostile to an American style drift in ethics, instrumentalised and utilitarian. Some even express the greatest mistrust of mechanisms of Anglo-Saxon origin (see D'Iribarne, 2002) which are spreading in French corporations.

However, to leave the ethical judgement completely to the discretion of each individual can prove risky for the corporation. The company must formalise a frame of reference guiding behaviour while leaving sufficient leeway to its staff for them to mobilise their own ethical conscience.

We argue for the adoption of a measured position concerning the relevance of these approaches. They sometimes only constitute a façade, unconnected with real activities, but they can also culminate in concrete changes in the functioning of the organisations (and the ethics policies can be strongly connected to practices).

To analyse these efforts and investments exclusively from the legitimacy angle seems to us to represent a too deterministic view, overlooking the importance of the discretionary latitude in organisational decisions. The values and commitments of the executives certainly play an essential role (Treviño and Weaver, 2003) and the ethics policies can reflect their own commitment to ethical behaviour.

Depending on the institutional context, corporations always seek a compromise between moralism and idealism in ethics management and, generally speaking, between the logic of efficiency and the quest for legitimacy.

Bibliography

- D'Iribarne P. (2002), « La légitimité de l'entreprise comme acteur éthique aux Etats-Unis et en France », *Revue Française de Gestion*, No. 140, p. 23-39.
- Mercier S. (2004), *L'éthique dans les entreprises*, Collection Repères, No. 63, 2nd édition, Editions La Découverte, Paris.
- Post J., Preston L., Sachs S. (2002), *Redefining the Corporation: Stakeholder Management and Organizational Wealth*, Stanford University Press.
- Treviño L. K., Weaver G. R. (2003), *Managing Ethics in Business Organizations*, Stanford University Press.
- Van Parijs P. (2004), Economie, in Canto-Sperber M. (dir.), *Dictionnaire d'éthique et de philosophie morale*, Quadrige, PUF, Paris, p. 591-598.

INTERNATIONAL LITIGATION:
DISCOVERY RIGHTS VERSUS PRIVACY RIGHTS

Vicki M. Luoma
Minnesota State University
145 Morris Hall
Mankato, Minnesota USA 56001
Vicki.Luoma@mnsu.edu
507 389-1916

Milton Luoma
Metropolitan State University
700 East Seventh Street
St. Paul, Minnesota USA 55106
Milt.Luoma@metrostate.edu
651 793-1481

Penny Herickhoff
Minnesota State University
150 Morris Hall
Mankato, Minnesota USA 56001
Penny.Herickhoff@mnsu.edu

ABSTRACT

One of the most difficult issues that have arisen in international litigation is the conflict between discovery rules of common law countries and personal privacy rules of civil law countries. Common law countries place a high value on disclosure of information in litigation through the discovery process in order to achieve a fair and just result. Civil law countries place a higher value on personal privacy and do not have a discovery process similar to that of common law countries. The conflict arises when parties from countries with different legal systems are involved in litigation. This paper reviews this issue and the attempts to resolve it.

Keywords: Discovery, E-Discovery, Common Law, Civil Law, Cross-Border

INTRODUCTION

The conflict between common law countries and civil law countries over privacy rights in litigation has been raging for more than three decades with no solution. When litigation occurs in an international setting, particularly when it involves a party from a common law country and a party from a civil law country, the problem arises immediately. In civil law countries individual privacy rights take priority over all other considerations, while in common law countries disclosure of all information that may be relevant to the case or lead to the discovery of additional information that may be relevant to the case takes priority. The rationale for full

disclosure in common law countries is the high priority placed on arriving at the most fair and just result based on all of the relevant facts and evidence. Most of the highly reported conflicts occur in litigation involving parties in the United States and parties in countries with strong blocking statutes, such as the European Union. [20]

Even European countries such as Spain and Portugal without strong blocking statutes are so concerned with citizens' rights of privacy that they have enacted other laws to strengthen their citizens' rights of privacy. At the same time, common law countries have expanded their litigation rules that provide greater access to data. Spain and Portugal, along with 23 other European countries, have passed the European Union Data Privacy Directives with recent amendments that provide penalties for violations of data breaches. As civil law countries tighten their privacy protections, common law countries such as the United Kingdom and the United States have passed new rules that provide stiffer sanctions for failure to provide the data. Lawmakers in Spain and Portugal as well as in other European Union countries are aware that the limits imposed by the EU Data Privacy Directives not only conflict with the nature of the international realities of business, but specifically conflict with discovery obligations under common law rules. The conflict between these ideologies can stymie business opportunities and be costly from the point of view of both money and resources. [20]

THE BASIS OF THE DILEMMA

The common law countries include the United Kingdom, the United States, Australia, New Zealand, South Africa, Canada and other countries that once were part of the British Commonwealth. Most of Europe, Central and South America, and countries in Asia and Africa are civil law countries.

The difficulty between the two legal system starts with the development of the systems of government and legal systems. The common law countries developed from English law when the Magna Carta was imposed on King John in 1215 by the lords. The Magna Carta limited the power of the king and gave individual rights to landowners and barons. The Magna Carta is the original basis of individual rights valued by all common law countries. Laws can be written by the legislatures in common law countries, but most of the law has been developed through a series of legal cases that are bound together through a legal principle of stare decisis. The legal concept of stare decisis requires judges to follow precedents established by prior decisions. [14] The rationale for this concept is that similar cases should be decided similarly. Civil law developed primarily from Roman law and is codified into a collection of laws with the legislature being the primary source of law. The court system is inquisitorial, ruled by specially trained judges and not bound by precedent. [14, p. 27] In a civil law country there is a difference between private law and public law. Private law is the area of law "in which the sole function of government was the recognition and enforcement of private rights." [14, p. 31].
Lawyers in civil law countries mainly conduct adversarial dialogue and argue various parts of the law with the judges who are actively involved in the investigation and fact finding process. In common law countries the lawyers gather and produce the information necessary for the adjudication process. [14, p.31]. In common law countries civil case decisions are usually made by a jury after hearing testimony and considering the evidence admitted in the proceedings. The judgment is then enforced by the trial judge. The judge's role is to maintain order, to ensure the

proper procedure is followed, to interpret the law as applied to the case at hand. The judge does not take part in the investigation process. [14, p. 33]

In the common law countries, the priority in litigation is finding all relevant information bearing on a case or information that might lead to further relevant information in the litigation. As a result, priority electronic discovery rules and case law have been evolving rapidly. The case law in the various common law countries has been developing separately and in-sync with cases in other countries. Many common law countries cite cases decided in other common law countries as persuasive authority. [1] Despite the differences in privacy laws in these common law countries, individual privacy is not the main concern. All common law countries are attempting to determine litigants' obligations and rights in civil litigation with the aim of making sure that the litigation process is transparent and fair. [21]

In the past five years common law countries have rewritten their civil rules of procedure to accommodate the changing technology in use with electronic data. Common law countries have moved past the focus of retention and deletion of digital information to the more complex electronic discovery issues involving the actual search process in culling out the digital information. In general, common law countries now require litigants to have a plan and methodology in place for retrieving all electronic documents and requiring meaningful meet-and-confer sessions, agreements over search methods, claw-back agreements, and the like. The aim of these requirements is to find all the relevant information which may or may not be used in the trial. [20]

Common law trial courts are also considering the issue of appropriate search methods to make sure all relevant information has been found in the discovery process. For example, should the search be done via concept searching, artificial intelligence, or are key word searches good enough? The U.S. Federal Rule of Civil Procedure 26(b)(1) allows parties to obtain materials regarding any non-privileged matter that is relevant to any party's claim or defense. [6]The rule also provides that information requested for production need not be admissible at trial if it appears that the request is reasonably calculated to lead to the discovery of admissible evidence. (Federal Rule of Civil Procedure 26(b)(1), 2007) . [6] Litigants can risk court sanctions if they have not stored information in a retrievable format and have not preserved data from being deleted.

Transparency is a key term in the civil law countries, too, but it is used in a different context. Both systems adamantly assert that transparency in the litigation process is the goal, but each system views the term in a different context with a different meaning. [12] In contrast to the common law countries' aim of making data available and discovery transparent to the requestor, the civil law countries focus on keeping the process transparent and visible to the party from whom the information is requested. This notion of transparency includes obtaining expressed permission of the party from whom the information is sought. [16]

THE EUROPEAN DIRECTIVES

In an effort to provide a cohesive policy across the Europe regarding privacy issues, the European Union has enacted legislation known as the European Directives. The European

Directives consist of 34 articles. [2] The Union Data Directive 95/46/ adopted by the European Commission in 1995, established principles restricting the handling and production of electronically stored personal data. [2] One directive prohibits the further transfer of data to countries that fail to have "adequate" levels of protection for personal information. The EU Data Directive requires that before any personal data can be transmitted, the individual in question must freely give consent to the transmission. Further, the use of the data must be lawful, fair, adequate, relevant and accurate and may be used only as long as necessary and have adequate security to protect the data at all times. This directive allows the use of personal data in litigation only if the litigant can prove that the information exists, provide the location of the information (without searching) and can prove the information is essential to the litigation. Although all EU countries have adopted this directive, the countries interpret and enforce them differently. (Council Directive 95/46/EC) [2]Further, this Directive permits only the "processing" of personal data in limited circumstances. The requirement that it must comply with foreign legal statutes or regulations is not considered as a valid reason. This Directive also requires that all processing of personal data must inform the individual involved to know and approve of the data transfer methods. Unlike common law countries this Directive does not allow personal data to be retained past the necessary processing application. [2]

One of the most interesting provisions is that the EU Data Directive specifically prohibits the transfer of personal data to the United States because data stored in the United States is not considered secure. Further, the transfer of personal data to the U.S. for preservation purposes is not permissible under EU law. [2]

The conflict of discovery requirements and privacy laws is inevitable. Currently, 93% of multinational corporations are doing some form of partial outsourcing or out-tasking projects. [4]In addition, there are approximately 3,000 multinational corporations with potential litigation conflicts. [4]. There is also litigation involving citizens of one country damaged or injured by a company or person of another country. In common law countries the discovery process is considered the key to a successful litigation in arriving at a fair and just result. Attorneys routinely serve discovery requests to obtain all information that might be relevant or lead to something that is relevant in the litigation. The key to common law litigation is full disclosure. In most civil law countries there is no discovery process and most of the civil law countries do not allow the disclosure of any document that is not specifically allowed at trial. [3]

DISCOVERY VERSUS PRIVACY IN COURT

Exacerbating the problem is the fact that many of the civil law countries have passed blocking statutes that provide criminal consequences for violation of privacy rights in litigation while common law countries are sanction parties for failure to comply with discovery rules. From 1958 to the present United States courts have consistently meted out sanctions if a litigant does not comply with discovery requests. [7] With the ever increasing amount of electronically stored data, The United States and most common law countries require that the electronically stored information must be stored in a retrievable format and deletion of data must be stopped as soon as a potential litigant knows or should know that litigation is likely to be commenced.

When litigation involves the United States and any foreign litigant, the United States courts rely on the Restatement (Third) of Foreign Relations Law of the United States to compel litigants outside the county to provide proper responses to discovery requests. [22] The court's justification is that persons and companies that conduct business in the U.S., or are within the jurisdiction of the U.S., receive both benefits and legal protection of U.S. law, and therefore, they have obligations that include complying with the discovery laws. [22]

As early as 1958, the United States courts sanctioned a foreign corporation for failure to comply with discovery requests. In the Societe Internationale Pour Participations Industrielles et Commerciales S.A. (a Swiss company) v. Rogersin, the district court dismissed the plaintiff's case with prejudice when the plaintiff failed to comply with the court order. The District Court ordered the plaintiff to produce certain records of plaintiff's Swiss bank. [17]The Court found the records to be relevant and to be within plaintiff's "control," within the meaning of Rule 34 of the Federal Rules of Civil Procedure. The plaintiff failed to produce the records claiming the production would violate Swiss laws and that they were served with an order prohibiting their production by the Swiss Federal Attorney. Undaunted, the United States District Court found that plaintiff had shown good faith in its efforts to comply with the production order, but found that the bank records were in their control and that the records were crucial to the litigation. As a result of their failure to produce the documents, the District Court dismissed the case. On appeal the Supreme Court found that the District Court had wide latitude to deal with the noncompliance including dismissal or negative inferences; however, it remanded the case to the District Court to revisit the dismissal with prejudice and provide a different sanction. [17]

In one of the leading cases concerning the conflict between civil and common law litigation the court lost its patience with noncompliance. In Societe Nationale v. District Court, the pilot and passenger in a Rallye plane manufactured by the defendants, two corporations owned by France (Societe), sued for personal injuries resulting from the crash of the airline. In the beginning, Societe answered the complaint without challenging the jurisdiction of U.S. District Court in the State of Iowa and answered the first set of discovery requests without objection. However, when Societe was served additional discovery requests, it claimed the requests would violate French penal laws. [16] Societe also claimed that the proper venue should be The Hague Convention and responding to the discovery would violate French law since the information was located in France. The government of France argued that, "The Hague Convention is the exclusive means of discovery in transnational litigation among the Convention's signatories unless the sovereign on whose territory discovery is to occur chooses otherwise." [17]

In 1970, the Hague Convention on "The taking of evidence abroad in Civil and Commercial Matters" was passed in an effort to resolve some of the conflict that occurs in multi-national lawsuits. Although the United States and 15 other nations agreed to the convention, the United States has consistently maintained that the agreement is voluntary and not mandatory on litigants. The purpose of the agreement is to give litigants the opportunity to receive help in the collection of evidence from another country. The problem with the convention is that the process is a slow process and neither side of the argument is willing to compromise their positions. [21]

Further, in the Societe Nationale case the French government argued that Article 2 of their French Blocking Statute requires that the litigants secure a waiver from the French government

before providing information.[7] The court found that there was no indication that the defendants attempted to seek a waiver from the government nor was it likely that the French government would have consented if asked. (Société Nationale Indust. Aérospatiale v. U.S. Dist. Ct. for the S. Dist. of Iowa, 1987) Ultimately the court ruled that Societe, defendants must respond to the discovery requests.[18]

In a 2006 case, the plaintiff requested data, including metadata, kept in Italy by and an Italian corporation, and the court granted the request. The Italian corporation and its U.S. subsidiary, the defendants, filed for a protective order alleging The Hague Convention governs the evidence taking procedure and that the court did not have jurisdiction. The court cited the Supreme Court's decision in the Aerospatiale case and found that Italy's objection against American pre-trial discovery is not as important as the goal of the United States rules that emphasize a just, speedy, and inexpensive trial.[8]

In a 2008 case, an oil tanker sank off the coast of Spain. Spain brought an action against the American shipping company. The plaintiff (Spain) objected to the discovery demands served in 2004 that included producing electronic documents. Spain claimed it had failed to preserve its electronic data as required by U.S. rules and refused to provide documents. The judge ruled that the plaintiff's actions were subject to sanctions. The court further ruled that the U.S. Federal Rules of Civil Procedure discovery requirements were the proper rules for this international action. The trial court's decision was upheld by the higher court and found that the plaintiff had received adequate notice concerning preservation and spoliation possibilities. The court ordered attorney fees. [13] The court observed that "this litigation is in the Southern District instead of a court in Spain."[13]

In a 2009 case, the court found that The Hague Convention was only voluntary and that the contested discovery issues "should and shall be conducted under the Federal Rules and not under The Hague Evidence Convention." [11]Maasvlakte, a Dutch corporation filed two proofs of claim in a bankruptcy case. The debtors objected to the claims and the bankruptcy administrator sought to compel Maasvlakte to produce information and submit to depositions in the United States. Maasvlakte objected on the grounds that most of the requested information was in the possession of a French corporation and further insisted that discovery must be conducted through The Hague Evidence Convention. [11] Maaslakte was a member of a general partnership to install a power plant in the Netherlands. The partnership signed an agreement with Air Liquide Engineering, S.A. (ALE) in which they agreed that ALE would be the agent and attorney-in-fact of the partners and agreed to delegate all matters to ALE. Thus, in the bankruptcy case, Maasvlakte, argued that all requested information was in the possession and control of ALE. Maasvlakte further contended that ALE was a non-party and therefore not subject to the control under Rule 34 of the Federal Rules of Civil Procedure and the information was controlled under the French Blocking Statute. Maasvlakte argued that to comply with the discovery requests would result in sanctions and possible jail. In response, the bankruptcy administrator argued the position that documents were under the control of Maasvlakte and the court agreed. [11]

In yet another case in 2010 in which a U.S. and German company were in litigation over a breach of contract, the German company objected to providing electronic discovery based on the German Blocking Statute and The Hague Convention. The U.S. District Court ruled that

blocking statutes "…do not deprive an American court of the power to order a party subject to its jurisdiction to produce evidence even though the act of production may violate that statute." [17' p544 n.29.] The U.S. court also relied on the American Law Institute Restatement on blocking statutes and discovery, which reads as follows: "Blocking statutes that frustrate this goal need not be given the same deference by courts of the United States as substantive rules of law at variance with the law of the United States." [17]

In yet another case the plaintiff, a French company, objected to producing discovery claiming producing the documents would violate the French Blocking Statute and some of the documents do not exist. The court found that the documents were within the plaintiff's control and the search was not thorough enough. In addition, the court found that the plaintiff did not argue the consequences of the French Blocking Statute until the motion to compel. The court ordered that the plaintiff make a more thorough search for documents and awarded attorney fees. [17]

This series of cases involving countries with blocking statutes has underscored the position of the United States courts in practice. It is clear that U.S. courts will not tolerate excuses for failing to comply with discovery requests. Essentially, if the U.S. court asserts jurisdiction over a case, the U.S. Federal Rules of Civil Procedure will be applied and enforced.

THE EUROPEAN UNION RESPONSE

In the past U.S. courts have disregarded the consequences of blocking statutes in their discovery decisions and until recently there had been few consequences to those who provided discovery despite the blocking statutes. However, in 2009 a French lawyer was fined €10,000 under the French Blocking Statute for providing information in California proceeding. The statute in question was the 1980 French Blocking Statute (the amended version of the 1968 statute) that criminalized cooperation with discovery procedures.[7] The criminal penalties provided in this law include up to 6 months in prison and a fine up to €18,000. In 1998 the California Insurance Commissioner commenced an investigation against a French consortium for an alleged fraudulent take-over of the U.S. insurance company Executive Life and its assets in the early 1990s. The California Insurance Commissioner felt that the French consortium led by *Crédit Lyonnais* hid the identity of its controlling interest violating California laws regarding foreign ownership of insurance companies. As part of its investigation the plaintiffs had a French lawyer conduct an investigation to see if the executive board knew of the fraud. The lawyer for Crédit Lyonnais provided the discovery requests as demanded by the California court and was later charged and found guilty of violating the French Blocking Statute. The conviction was affirmed by the French appellate court. He was fined 10,000 Euros [9]

In addition to increased enforcement of the blocking statutes, Spain and Portugal, like many of the EU countries, have passed additional laws to protect and to enforce the data privacy of its citizens. On June 23, 2010 Spain passed the Organic Law 5/2010 amending the existing Criminal Code (Organic Law 10/1995). The purpose of this new law was to hold companies responsible for failure to comply with existing laws. Violation of this law or any criminal statute (including the data laws) can result in sanctions for companies that include the company's being suspended from operation, being involuntarily dissolved, being disqualified from public contracting, among other possible court actions.

Further, each individual member state of the EU has the ability to provide more stringent protections. In addition to the increased enforcement of the Blocking Statutes, Spain and Portugal have passed additional laws to enforce the data privacy of their citizens. On June 23, 2010 Spain passed the Organic Law 5/2010 amending the existing Criminal Code (Organic Law 10/1995).[23] The purpose of the new law was to make parties responsible for failure to comply with existing laws. Violation of these additional laws can render a party subject to the same sanctions that can apply with the basic blocking statutes. Thus, the EU has imposed draconian measures to enforce all laws blocking disclosure of private information that will inevitably be required in litigation brought in U.S. courts.

In addition to enforcing the EU directive, Portugal has a constitutional provision protecting its citizen's privacy rights. To enforce this constitutional right, Portugal has established a Comissao Nacional de Proteccao de Dados (Data Protection Authority) with the authority to monitor compliance. Failure to comply can result in criminal sanctions, fines, and prison. Both Spain and Portugal have made the penalties much stiffer for individuals or companies to comply with the discovery request involving litigation with a common law country. Although the EU has met to consider whether there might be a way to resolve the differences with the common law countries, but no resolution of the issue appears likely. [5]

THE SAFE HARBOR PROGRAM

In an attempt to resolve the conflict between legal systems, in 2001 a Safe Harbor Program was negotiated between the U.S. Department of Commerce and the EU governments. A second agreement was negotiated between the U.S. Department of Commerce and the Federal Data Protection and Information Commission of Switzerland. These agreements were made to allow U.S. companies to receive personal data from the EU country by a voluntary self-certification process in which the U.S. companies certify that they are upholding privacy standards for personal information. As part of the certification, the company must guarantee that it has included the following:

(1) Give notice to the individual whose data is be transferred;
(2) Give the individual the choice to opt-out;
(3) Give the party notice and choice whether that the data would be transferred to third parties;
(4) Give the individual access to his or her own personal information;
(5) Make reasonable security precautions to protect data;
(6) Include only relevant and accurate data; and
(7) Provide a complaint and enforcement mechanism.

While the Safe Harbor provided that complying corporations could transfer electronically stored information (ESI), it still did not allow the processing of ESI for litigation purposes. Even with this negotiated agreement and even if the companies complied with the Safe Harbor Program, the company could still be sanctioned under an EU Blocking Statute or the Data Protection Directive. On February 11, 2010, despite a signed agreement between a U.S. Bank and 27 EU countries that data could be transferred through the SWIFT network, the EU decided that the

privacy of its citizens was more important than any benefit of releasing their citizens' information. [2]

OTHER RESOLUTION ATTEMPTS

The European Commission initiated a dialogue with non-EU/EEA countries in an attempt to resolve the issue of the protection of private information when exporting personal data to those countries. Yet after the meetings, the EU decided to tighten its restrictions and provide more protection for their citizens' privacy. The Vice President of the EU, Viviane Reding, wants to widen the scope of privacy protection and to review the Data Protection Directive of 1995 to determine what steps need to be taken to improve privacy. The Vice President has announced in 2010 that "upholding EU fundamental rights in all EU legislation" is a priority. [5]
In the United States the Sedona Think Tank in the United States has been trying to work on this problem as well. It scheduled an international conference in Barcelona in 2009. While the working papers offered some suggestions for resolving the problem, nothing concrete has come of it. [4]

Both sides of the issue believe their laws, purposes and methodologies are correct. Multinational corporations need to establish a litigation team as soon as the company ventures into another country to conduct business. A local lawyer must be a part of that team. Obtaining an agreement of all employees for sharing data in case of litigation is a possibility, but the EU does require that employees must freely give permission. Companies must also do an assessment of each country's laws and develop a plan in case of litigation. It may also be prudent to maintain information in more than one country. Before outsourcing information, a risk assessment should be conducted. Some countries may be more able to comply with common law countries' requirements.

CONCLUSIONS

In conclusion, currently the only serious efforts to bridge the differences in legal systems has been to refer disputes to The Hague, but requiring litigants to go to The Hague to resolve problems is also not a reasonable solution. At the present time neither side is willing to compromise, so referring issues to The Hague is simply not a feasible alternative.

If there were movement by the EU, the common law countries could limit discovery to known relevant information and not the information that might lead to relevant information. This action would introduce the concept of proportionality in discovery based on the size of the case. However, at this point neither side is willing to compromise. The most a multinational company can do now is to keep this issue in mind when determining where companies are located before deciding to conduct business.

A recent effort by the American Legal Institute and the European Institute for the Unification of Private Law ("UNDROIT") shows more promise than earlier efforts. These organizations have established a system of legal rules that are intended for use during international litigation and transactions. The system of rules is borne of practical considerations and gives signatories the ability to create a bilateral agreement that specifies the rules that will be followed in legal

proceedings should the need arise. In that case, more expansive discovery procedures can be specified by the parties that might otherwise not be permitted. Flexibility in defining the terms of the agreement between parties would give the parties the most freedom in defining their relationship in a legal context. While this sort of an agreement could work well between businesses whose relationship is of a contractual nature, rights in litigation involving some torts with third parties may not be so easily resolved. [3]

Ultimately, there are no simple and universal answers to the challenges and issues posed in cases of international litigation where the jurisdictions involved have enacted legislation that are diametrically opposed to each other. All organizations that operate in multiple international jurisdictions must be fully aware of the rights and duties of all potentially interested parties and prepare for the challenges. It would not be reasonable for EU companies to do business in the United States and expect special dispensations in lawsuits because of their country's laws are different. Similarly, companies from common law countries should expect the discovery rules to change when litigation is commenced in an EU country. The adage "When in Rome …" has never been more appropriate.

REFERENCES

[1]Australian Securities and Investments Commission v Macdonald , 1169 (NSWSC 2008).

[2]Council Directive 95/46/EC, art. 2(b), O.J L 281.

[3]Daley, M. a. (2008). A Practical Guide to Navigating the Competing Currents of International

[4]Data Privacy and e-discovery. *The Sedona Conference* , 1-30.

[5] Dainnow, Joseph, The Civil Law and the Common Law: Some points of Comparison, 15 Am. J. Comp. Law 419 (1966-1967)

[6]Federal Rule of Civil Procedure 26(b)(1), 26B(1) (2007).

[7]French Penal Code Law No. 80-538 (1980).

[8]Hagenbuch v. 3B6 Sistemi Elettronici Industriali SRL, WL 665005 (ND ILL March 8, 2006).

[9]In re Avocat “Christopher X,” Cour de Cassation, Appeal t,, n. 07-83228 ((French Supreme Cour December 12, 2007).

[10]In re: Application of Operadora DB Mexico, S.A. De C.V., 2435750 (M.D. Fla. May 28, 2009).

[11] In re Global Power Equipment Group, Inc., 418 B.R. 833 (Bankr.D, Del, 2009)

[12] LAW 110–322—SEPT. 19, 2008 122 STAT. 3537

[13]Linde v. Arab Bank, WL 1456573 ((E.D.N.Y May 22, 2009).

[14] Merryman, J. The Civil Law Tradition (2d ed. 1985)

[15]Reino de Espana v. Am. Bureau of Shipping, Inc., WL 3851957 (S.D.N.Y. August 2008).

[16] Sedona Conference Cross Border Discovery Conflicts: A practical guide to navigating the Competing Currents of International Data Privacy and e-discovery (WG6) Public Comment Version 2008

[17]Societe Internationale Pour Participations Industrielles et Commerciales S.A. v. Rogers , 357 U.S. 197, 213 (1958).

[18]Société Nationale Indust. Aérospatiale v. U.S. Dist. Ct. for the S. Dist. of Iowa, 482 U.S. 522 (U.S. Supreme Court 1987).

[19] Sofaer Global Hedge Fund V Brightpoint, Inc, 2010 WL 4701419 (S.D. Ind, November 12, 2010

[20]Sorell, Matthew, editor, "Forensics in Telecommunications Information and Multimedia" Second International Conference , e-forensic 2009 Adeladide, Australia, - "Complying across Continents: At the Intersection of Litigation Rights and Privacy Rights. Milton H. Luoma and Vicki M. Luoma 186-195

[21]The Taking of Evidence Abroad in Civil and Commerical Matters (The Hague March 18, 1970).

[22] The Restatement Third of Foreign Relations Law

[23] The Spanish Organic Law

SHORE HOME INVESTMENT PROPERTIES MANAGEMENT

Dawn Konrady, Richard Stockton College of NJ, e-mail: dawn.konrady@stockton.edu
Katie Watson, Richard Stockton College of NJ, e-mail: kwat327@gmail.com
Haidong Liu, Richard Stockton College of NJ, e-mail: haidong.liu@stockton.edu

ABSTRACT

Shore Home Investment Properties (SHIP) is a unique house watching firm. Customers rely on SHIP to monitor and maintain their shore homes during their absence. The owner of SHIP has a goal of expanding into ten shore towns along the coast of South Jersey from Atlantic City to Cape May in the next five years. The current information systems (QuickBooks) may cause problems during the expansion. The case study outlines six recommendations for SHIP information systems. Implementing the suggestions will streamline SHIP business operation and bring SHIP to the forefront of property management companies.

About Shore Home Investment Properties (SHIP)

Shore Home Investment Properties (SHIP) is a house watching firm servicing homeowners in the towns of Stone Harbor and Avalon, New Jersey. The company was established in 2001. Customers rely on SHIP to monitor and maintain their shore homes during their absence. The services required by each customer are personalized to fit their specific needs or desires. Competing property management firms in the area may offer some similar services, but SHIP offers more extensive and unique services.

SHIP is a very small company because it currently serves a very small region. Over the past few years SHIP has handled between 30 and 40 accounts. The downturn in the economy has had a drastic effect on the company. In 2008 SHIP had revenues of $71,806 and a net income of $27,778. In 2009, SHIP suffered a 12% loss of revenue yielding $63,250 for the year and only $21,075 in net income. For a complete listing of expenses, refer to the income statement in the appendix. The company has experienced a loss in revenue due to the poor economy. Some previous SHIP customers no longer needed the company's services because they sold the properties that SHIP managed due to no longer being able to afford shore real estate. Most of the customers that SHIP lost required a wide variety of services which meant that they brought in a higher amount of revenue. While a few customers may have been lost, SHIP also continues to gain more customers through referrals from current customers. However, the new customers have not been requiring as many services and therefore are not generating the same level of revenue per customer. Also, some of SHIP's current customers decreased the services they required in an effort to save money. This explains why the number of accounts SHIP handles has stayed the same, but their revenues have decreased over the years. It is important that the company continues to grow because it needs more customers in order to earn more profit. Without positive word of mouth, SHIP would gain very few new customers.

When SHIP acquires a new customer, a representative of the company sits down with them to discuss the specific services they require. SHIP offers a broad range of services including, but

not limited to: checking the customer's home once a month during the winter when they are not present, inspecting the customer's home daily when they are not present to empty the dehumidifiers, and helping customers set up their outdoor furniture for the summer and pack it up for the winter, just to name a few. SHIP will also arrange for the customer's car to be dropped off at the local community airport before they land and take the car back to their residence when they leave because the local airport is not large enough to provide long term parking.

SHIP's customers are not local, and as such are not familiar with local contractors and services. SHIP is also a general contractor and hires local contractors to do work on a customer's house. If a customer needs their bathroom redone SHIP will arrange for the carpenter, plumber, and painter to arrive to complete the task. SHIP will be sure to schedule the workers and make sure that the house is open when the contractors are supposed to arrive. The contractors are still hired by the customer and not by SHIP so the customer is ultimately responsible for approving the plans and paying the bills. Since the contractors are outside employees, SHIP can hire multiple contractors to work on different properties simultaneously. This arrangement also allows SHIP the opportunity to hire a specific worker if a customer has a preference. Since these contractors are not SHIP employees, SHIP saves on liability insurance by not having a staff of contractors employed directly. This is one of the largest differences between SHIP, a house watching firm, and the property management firms that SHIP competes with.

In the Avalon and Stone Harbor area, there are currently four competing property management and house watching firms. Some new companies are currently being formed. The types of services that are offered are very similar, but SHIP strives to make sure every customer experience is personalized to fit the specific needs of the customer. Despite the amount of competition in the area, SHIP is the leading house watching firm for Avalon and Stone Harbor; a position the company would like to maintain. In the future, they would like to expand their target market to other shore communities in the Cape May and Atlantic County regions. The owner of the company has a goal of expanding into ten shore towns along the coast of South Jersey from Brigantine to Cape May in the next five years. This may be a difficult task at first because the only technology that is currently in use is QuickBooks to maintain bookkeeping, keep track of customer payments, and manage operating expenses. SHIP also utilizes a website, but it is extremely outdated.

The current information systems may cause problems during the expansion. It would be beneficial to the customers and the company itself if its information systems were updated. There are four areas where an improved information system will assist with SHIP's goals of expanding:

- Creating a database management system which includes a better customer database, inventory system, and accounting system
- Giving the employees the ability to access information from the field via a property maintenance system
- Increasing marketing
- Improving the website and creating an online customer access system.

In 2010 SHIP expects to have revenues of about $52,000 and a net income of about $23,000. With the increase of clients from expanding the company, the operating costs will increase proportionately with the exception of incurring new expenses based on the initial setup costs for the new information system. However, the benefits of having a new system in place will quickly outweigh those initial costs. The streamline style operation backed by an information management system will improve revenue, shorten the process and allow employees to become more efficient, and create happier customers that will be able to see that the services they pay for are taking place.

The Database Management System

SHIP currently maintains paper files on each of its customers. For SHIP to enhance the tracking of customer information, streamline productivity, and ensure accuracy ,a database management system (DBMS) should be developed. The advantage of using a DBMS is that the incorporated information systems can change as the needs of SHIP change over time. This allows for the development of new systems without destroying the existing data. The DBMS will be designed as a relational DBMS that allows for communication between three separate systems or programs and relays that information to the main customer information system database. The three systems that will be connected are an inventory system, accounting system, and property maintenance system. The DMBS will serve as a platform integrating various programs or applications that are related to the three main systems. Figure 1. on the next page gives a visual representation of the DBMS and its associated systems.

Figure 1. Proposed SHIP DBMS
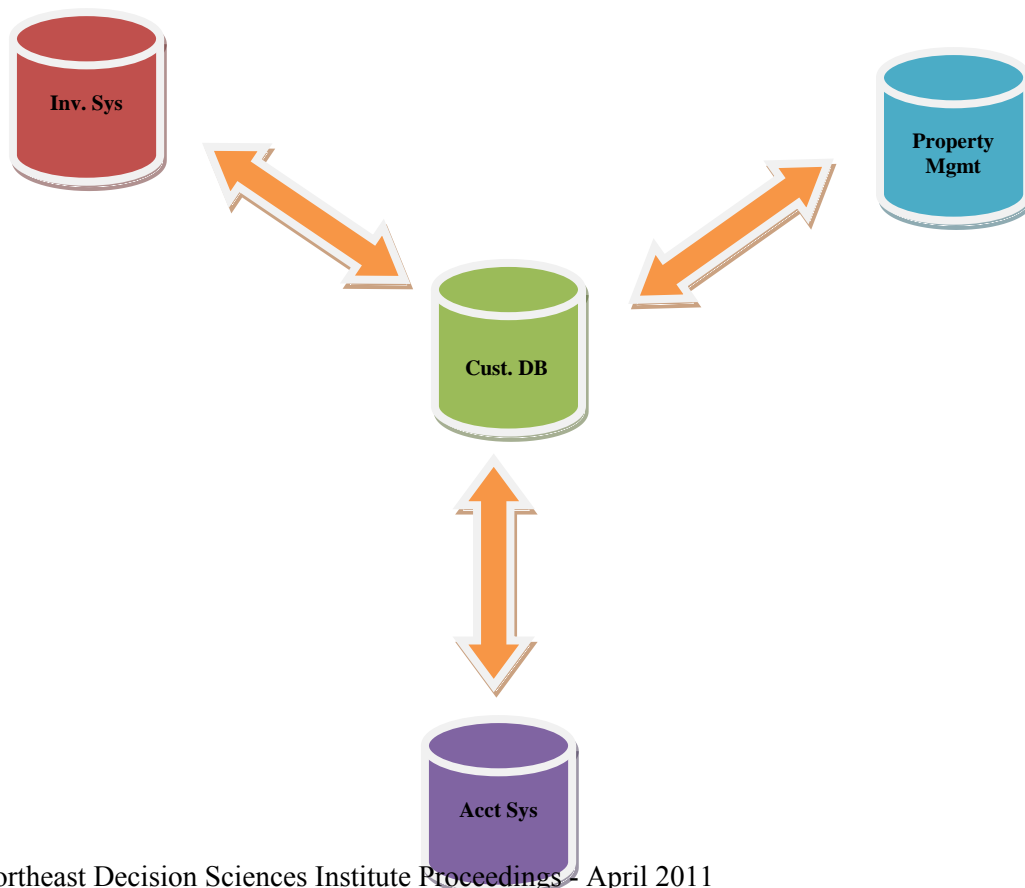
Customer Information Database

The Customer Information database is the central system that all other systems relate to. It will be used to record specific information about each customer. The creation of the Customer Information database allows for SHIP to centralize all its information so that all staff would be able to access it at any given time. The company can increase the personal information that it collects so that a personalized level of service can be created and customized for each customer. Capturing this customer information will reflect attention to detail and foster trust and confidence that the customer has chosen the right company to perform the necessary services. It shows that SHIP takes pride in getting to know its customers and is paying attention to subtleties in customer preferences. The information that is recorded in the Customer Information database will then be available across all systems so at anytime either the customer via a new website or an employee in the field will be able to access customer information.

Inventory System

A unique feature of SHIP is stocking a property with whatever food, beverages, or supplies the property owner would like prior to their, or their rental customer's, arrival. SHIP has accounts with various local vendors allowing the company to purchase items and then create a bill that can be applied to a customer's account for later payment. Currently SHIP has no inventory system in place which allows for quick referencing of what supplies or products are needed for the property or a record of the client's product preferences. With the implementation of the recommended DBMS, this will allow SHIP to keep a running record of what supplies or products are currently in the house, what is running low, and what the preferences of the customers are.

Accounting System

The accounting system for SHIP is complex because all of the customers have different arrangements to pay their bill. Customers have the option to pay monthly, quarterly, or annually depending on the services that they require. Managing when payments are due for each customer can be challenging. It may be easier to have a set of rules, but since everything is personalized at SHIP, they are not able to standardize the bill paying schedule.

With this in mind, the company needs a way to electronically manage when bills for certain customers are due. This will assist in the process to create and send out invoices. Once the company increases the number of customers, it is important that a program can determine when each customer is supposed to pay their bill. This is where an online customer access system will help with the accounting system. Customers will always be able to view their bills, receive notifications through the access system when it is time to pay the bills, and actually pay their bills through the customer access system. Online banking is becoming more popular, so it is important to have that as an option for bill paying.

The president of SHIP is also responsible for paying some of the utility/maintenance bills for a few of SHIP's customers. He pays for items such as cable, landscaping services, cleaning services, or maintenance out of a joint account he has with the customer. The president of SHIP also purchases household items, such as televisions or computers for some customers. The

majority of the money for these purchases comes directly out of the joint accounts but sometimes the president pays for these items himself. He must carefully keep track of these expenses so that he can be fully reimbursed for his purchases. These accounts should also be included in the online access system so that customers can monitor the bills paid and items purchased for them.

It is also important for the company to generate financial reports as needed. An upgraded system will make it much easier to track incomes and expenses. The company needs to monitor their accounts receivables to know who has paid their bills on time. An upgraded system will be able to pull all of the pertinent information together to form financial statements such as balanced sheets or income statements so that the employees do not need to spend their time doing it.

Property Maintenance System

Currently, each customer fills out an initial property management authorization form that contains basic information about the property and the customer. It includes an area for describing the service plan chosen, preferred contractors or vendors for services such as plumbing, electric, landscaping, HVAC, etc. Again, all of this information is recorded on paper and filed. There is no electronic organizing of this information. The Property Maintenance System will hold all this information and relay it to the main Customer Information database. Once this information is captured, any employee will be able to access this information and determine who they need to contact when it is time to arrange for the pool to be opened, who to call when there is an HVAC emergency, etc.

The best part about creating the property maintenance system is that the information in the system can easily be accessed through an enterprise application via a smart phone. This capability allows the employees to instantly record information on each of the properties he or she visits as well as take pictures and then upload them to the client's account. Once the information is captured it is instantly changed in the customer's maintenance history. Many open source and cost applications exist so it would just be a matter of finding one that meets SHIP's needs and is fairly inexpensive, and easy to deploy.

By having this portion of the DBMS available and viewable to the customers, it will instill additional confidence in the company and the services that are being paid for because the customer will be able to see a service history through a customer account web page. Previously, customers would find a piece of paper on their counter detailing the visits and any issues addressed. Months could go by before the customer would see this. If there was an issue during an inspection, the employee has to manually look up the customer contact information back at the office, which is not a productive use of time. With the proposed new system, customers will be able to see the information whenever they log onto the website and access the maintenance/visit log.

Increasing Marketing

In 2001 when the company was established, an immense amount of marketing was conducted to inform potential customers about the services that SHIP offered. Various direct mail campaigns were carried out to individuals who owned second homes in Avalon or Stone Harbor. One

campaign involved sending pizza-like boxes that were personalized by describing a potential customer's shore home and the services that SHIP had to offer. Assembling and personalizing the pizza boxes required a tremendous amount of time and effort. It was also very expensive to purchase the materials needed and to ship the finished products to the potential customer's primary residence. After the initial marketing campaign, SHIP did very little advertising. One advertising method that has been utilized since the beginning is the use of yard signs. On most of the houses that SHIP maintains, there is a sign in the front yard advertising that it is a SHIP property. The yard sign, unfortunately, does not explain what the company has to offer. In order for the yard signs to be more effective, they should explain that SHIP can take care of your house when you are not present. Adding a simple phrase like "*Do you know your house is safe? These owners do. Call SHIP to find out more.*" That phrase would draw people in and make them wonder what the company had to offer. The only other type of advertising that has been continuously used is word of mouth. This can be a very effective advertising method, but it is very difficult to control. Most of the customers seem to be very pleased with the service that they have received, so they only spread positive word of mouth.

If the company would like to expand their target market to other shore towns, they are going to need more marketing than word of mouth. Word of mouth will not be effective for introducing the company into a new market because it will be difficult to make the connections between current customers and potential customers when they are located in different towns. SHIP could use a direct mail campaign to introduce the company to new customers. However, the direct mail campaign should only involve flyers or letters explaining the company. Creative marketing, such as pizza boxes, should not be necessary.

Another way the company could advertise is in local publications, such as the weekly newspaper that each island publishes. The local residents or visitors are likely to read it and recognize the ad from the information that they received in the mail. SHIP should also focus heavily on electronic advertising. This should be a priority for the company. Currently, the SHIP website does not appear in any search engine results. This should be changed so that prospective customers can find the website after they search for items such as "SHIP, Stone Harbor." An e-mail campaign would be effective to explain what the company has to offer. E-mails are relatively easy to design and send out. The company can send out different e-mails, such as ones that correspond with the seasons, to make prospective customers think differently about the various types of services that SHIP has to offer. Before an e-mail campaign begins, the website for the company would have to be improved. This is because whenever someone opens an e-mail, they will be directed to the website for more information. The company needs to make sure that they have a modern website.

Improving the Website

Currently, SHIP does not have a very engaging website. The current website was designed in 2001 when the company was started and it has not been updated since. SHIP's homepage is very plain and simple. This may make customers believe that the company is unprofessional because they do not have a modern website. The information provided on the website is sufficient for the company. It explains what the company has to offer as well as information about the towns where SHIP is currently working. Also, it contains contact information for current and

prospective customers who would like to get in touch with employees at the company.   In order to compete in this technologically advanced time, SHIP needs to upgrade their website.

SHIP should make their website more modern. The graphics could be upgraded. Something more than just plain text on the page would be more attractive to a viewer. The only pictures on the website were taken of a few houses on beautiful days. It does not describe the services of the company well. Inclement weather scenes, such as a flood or a snow storm, could be shown so visitors understand that some bad weather may affect their houses when they are not present. Another approach would be to include visitors relaxing when they are on vacation because they are SHIP customers, whereas non SHIP customers have to be busy maintaining their property while they are on vacation. It is important to use the website as a marketing tool.

The website could also be used as a customer service tool. Implementing an online customer access system or dashboard would be very valuable for this company. Customers would enter this system through a log in feature on the website. They would require a secure user name and password to enter their specific site to ensure their privacy is protected. The property management applications for smart phones, proposed earlier, will allow the employees to update the property management system (maintenance log, property photos, status of construction, etc.) right from the customer's house while they're there doing a check, and it will appear in the customer access system. The customer could then enter the access system and view when their house was checked and see the updates on its condition. This access system would provide an easy way to communicate between the customers and the SHIP employees. This access system would also provide additional services, such as access to checking payment statuses, submitting online payments, etc, which would come from the development of the accounting system and associated application.

Conclusion

If SHIP implements the six improvements identified, it will have no problem achieving its goal of extending to a total of 10 shore towns over the next 5 years. The increase in revenue will allow for the expansion of staff which will be able to handle the increase in business and property maintenance. Creating a database management system which includes a better customer database, inventory system, and accounting system; giving the employees the ability to access information from the field via a property management system; increasing marketing; and improving the website; are all feasible and relatively inexpensive recommendations for expanding the customer base. Considering that SHIP currently has no type of electronic infrastructure, implementing the suggestions we have made will bring SHIP to the forefront of unique house watching/property management companies and become a serious competitor.

Appendix

<div align="center">

Shore Home Investment Properties

Income Statement
For the year ending December 31, 2009

</div>

	2008	**2009**
Revenue		
Service Revenue	71,806	63,250
Total Revenue	71,806	63,250
Expenses		
Advertising and Promotions	3,182	3,773
Auto Expenses	14,691	12,049
Bank Charges	9	45
Dues	425	830
Internet and Computer Services	400	1,350
Legal and Accounting Fees	995	1,300
Meals and Entertainment	6,816	4,655
NJ Corporate Income Tax	4,903	3,488
Phone	2,109	1,567
Postage	347	533
Rent Expense	3,600	3,000
Repairs and Maintenance	4,262	5,777
Supplies	2,491	3,361
Utilities	349	448
Total Expenses	44,578	42,175
Net Income	27,228	21,075

Can we afford to be in the Trash Collection Business? – Is Privatization the Answer?

Lynn Ruggieri Associate Professor of Accounting
Michelle Maczka Class of 2011
Roger Williams University

Carlos was tired, he put in a full day on the truck, and the meeting was lasting longer than usual. There were a great many items on the town council agenda. He would wait; the men were counting on him. He was nervous, he was not accustomed to speaking at a public forum but their jobs were on the line. He needed to show the town council members why this was a bad idea.

Background

Barrington Rhode Island, a quaint suburban community of primarily single family dwellings, is located nine miles southeast of the capital of Providence. The Warren river and Narragansett Bay flow mightily through the town and no point is more than two miles from salt water. It is a small town with a total land area of 16 miles only 9 of which is land while the remaining area consists of inland waterways. It is a picturesque small town known for the quality of its schools and its standard of living.

The Town operates under a charter form of government, which provides for a five member Town Council elected at large, for staggered four-year terms. They are responsible for enacting local legislation, and an appointed Town Manager executes the laws and oversees the administration of the town government.

Barrington is primarily a suburban residential community. Residential property accounts for 89.9% of the total tax roll. In terms of real estate tax revenue, 88.62% of the Town's tax revenue is derived from residential real estate. The town operated at a surplus for fiscal year 2009 adding $1,183,282 to net assets. See exhibit I and II for detail.

1

In 2000 Barrington was ranked 22nd in population among the thirty-nine cities and towns in Rhode Island. Based on the U.S. Bureau of the Census, population in the Town from 1950 to 2000 was as follows:

Years	Population
1950	9,105
1960	14,850
1970	17,375
1980	16,174
1990	15,849
2000	16,819
Source: U.S. Bureau of the Census, 1950-2000	

There are approximately 6,000 homes in Barrington. It is not inexpensive to live in Barrington. The median sales price of a home was $353,500 in 2010 and $305,000 in 2009. The town is also known for its high property taxes with a rate of $16.10 on real estate per $1,000 based on 100% valuation. The taxes on a median priced home were $5,691 in 2009.

The town offers the one of the top three school systems in the state and a multitude of services including police, fire and a department of public works. The Town of Barrington Public Works Department provides a full range of services including street maintenance, snow removal, garbage and yard waste collection and disposal, and maintenance of Town property.

To Privatize or not?

The department of public works, which employees a staff of eight, provides the following waste collection services:
- Refuse collection – once a week
- Recycling – once every two weeks
- Yard Waste – once a week

With property taxes already high and no end in the economic recession in sight the town had decided to consider privatizing trash collection as an alternative to raising taxes. Proposals were requested from trash hauling companies and the lowest bidder was MTG Mega Disposal at $828,000.

2

Other Bids received were
 Coastal Disposal: $935,820
 Waste Management: $964,000
 M&M Disposal: $984,495
 ABC Disposal: $1,439,521

The terms of the contract with MTG Disposal provide that for the sum of $828,000 per year commencing on October 1, 2010 through June 30, 2016 MTG will provide the following waste collection services:
- Refuse collection – once a week
- Recycling – once every two weeks
- Yard Waste – once a week from April to December

MTG also agrees to maintain worker's compensation and liability insurance. The contract price shall be adjusted after the completion of 12 months based on the annual evaluation of the cost of living price index as published by the department of Labor, bureau of Statistics Boston Labor Division of the Northeast Region.

The discussion grew heated at times but the bottom line was a basic financial question — would contracting with a private vendor save taxpayers money?

Carlos remained patient while Peter DeAngelis, town manager, outlined the reasons for outsourcing. Mr. DeAngelis said that if the town did not accept the Mega contract and continued the trash collection through the Department of Public Works (DPW), it would have to invest in two new trucks immediately —
- $147,000 refuse vehicle and
- $210,000 recycling truck

He explained that costs of the DPW fell into two categories – operational and capital. Operational costs included compensation and benefits for the workers and fuel and repairs for the fleet of 7 trucks used. He provided the following costs for the 10 month period of September 1, 2010 through June 30, 2011. June 30 is the end of the fiscal year.

3

OPERATIONAL COSTS

Salary		$ 279,452		
Overtime		$ 13,391		
Medical		$ 68,398		
FICA		$ 22,403		
Worker's Compensation		$ 27,923		
Pension		$ 11,988		
Total Compensation & Benefits			$ 423,555	
Insurance		$ 7,593	$ 7,593	
Fuel & Repairs				
Refuse vehicle fuel		$ 23,941		
Recycling vehicle fuel		$ 14,205		
Repairs		$ 27,151		
Total Fuel & Repairs			$ 65,297	
Total Operational Costs			$ 496,445	

It is anticipated that these costs will rise each year in the future as follows:
1. Salary increase of 2%
2. Medical increase of 10%
3. Worker's compensation and liability insurance increase of 5%
4. Repairs increase of 5%
5. Diesel fuel increase of .25 for refuse vehicle (9,279 gallons were used in 10 months)
6. Diesel fuel increase of .25 for recycling vehicle (5,506 gallons were used in 10 months)

In addition, Mr. DeAngelis stated, if the town remained the provider of trash collection the following capital expenditures would be incurred for equipment

PROJECTED CAPITAL EXPENDITURES			
YEAR		AMOUNT	
2		$	360,000
5		$	215,000
4		$	219,000
5		$	200,000
6		$	228,000
TOTAL		$	1,222,000

4

"If we want to continue in this business we're going to need to step up and purchase the equipment," Mr. DeAngelis said. "The town manager said the DPW currently has four refuse trucks at a cost of about $200,000 each and three recycling trucks which cost about the same. The $1.4 million fleet requires regular replacements, said Mr. DeAngelis, which had not been a problem until budgetary constraints at the state level began trickling downhill toward Barrington."

The Town Planning Board Committee for the Town of Barrington meets monthly and its members are appointed by the Town Council for 3-year terms. In accordance with Chapter 5, Capital Improvement Program, of the Code of the Town of Barrington, the Planning Board shall prepare an annual capital budget and six-year Capital Improvement Program for submission to the Town Manager. The Recommended Capital Improvement Program for the Fiscal Years 2010-2016 was submitted January 9, 2009 and contained the following funding plan for refuse and recycling vehicle replacement

PROJECTED CAPITAL EXPENDITURES		
YEAR		AMOUNT
2		$ 360,000
5		$ 170,000
4		$ 170,000
5		$ 170,000
6		$ 170,000
TOTAL		$ 1,040,000

The DPW currently employs a force of 30 workers. "There are eight positions associated with the privatization move," Mr. DeAngelis said.

Mr. Pedroza, at last taking his turn to speak, said he views those eight positions in a different light than Mr. DeAngelis. "Those are eight men with families, kids, mortgages," Mr. Pedroza said. "I had one of the refuse guys come up to me last week. He's got two little girls. His wife is out of work. He asked me what was going to happen. He wanted to know if he was going to lose his job. "Those same eight guys they're talking about eliminating, they also plow in the winter, salt and sand, and do fields in the spring."

 The personnel matter weighs heavily on Mr. Pedroza and union secretary Kevin Braga, as do a few other issues surrounding the matter. For starters, said Mr. Pedroza, the bid submitted by Mega has some significant holes. He said the yard waste collection services offered by Mega would start in April and end in December, a big step backward from the full-year yard waste collection the DPW currently offers.

5

Carlos Pedroza, chairman of the DPW union, said the analysis figures are not accurate. "The numbers provided by the Town Manager for capital are not actual, and are misleading. A review of the last 10 years of actual capital costs as listed in the yearly budgets for new refuse and recycling vehicles is less than $80,000, not the $350,000 listed in the Town Manager's summary of in-house services," Mr. Pedroza stated. "The consideration and evaluation of any bid must be based on a comparison of actual cost, not assumption of projections. He provided the following capital projection figures for the purchase of refuse and recycling equipment:

PROJECTED CAPITAL EXPENDITURES		
YEAR		AMOUNT
2		$ 360,000
5		$ 210,000
4		$ 110,000
5		$ 110,000
6		$ 120,000
TOTAL		$ 910,000

"Plus, why was Mega so much lower than the other bids? Something doesn't add up," he said. The next low bid to Mega's $828,000 was Coastal Disposal's $935,820. The three other companies ranged from $964,000 to more than $1.4 million. Mr. Pedroza thinks the lower bid might imply a reduced level of services, including fewer returns for missed pickups. "Right now, residents just have to call and we'll go back and pick it up by the end of the day. Mega's not going to do that," he said.

Mr. Braga, who works on one of the town's recycling trucks, said he has plenty of experience with Mega Disposal. Mr. Braga lives in East Providence, which contracts with Mega for refuse and recycling pickup."If you don't follow their rules, if everything's not in the bin, they'll just leave it there," Mr. Braga said. "I can tell you we [the DPW] pick up everything."

Mr. DeAngelis said he would not agree to hire a private firm unless the quality of services remains the same, adding that he would also request an operations manager who would oversee any calls placed by residents to the DPW about pickup services. He said his decision to explore privatization was not initiated by any complaints filed by residents over refuse, recycling or yard waste pickup. "At the end of the day, there has to be a high level of services... that's as important as the savings produced," he said. Mr. DeAngelis said that if the town should decide to hire a private company, he would do everything in his power to find jobs for the released DPW employees at the contracted disposal firm.

6

Ed Grindle and Joe Moraise are not interested in any job offers from Mega Disposal. The two men are employees of the Barrington Department of Public Works and are being told that should the town go private with its refuse collection, they — along with six other DPW employees would be offered jobs with the private contractor, Mega. The only problem is, Mr. Grindle and Mr. Moraise have already worked for Mega and have no interest in returning to that company. "Not at all," Mr. Grindle said, when asked if he would want to work for Mega again. "That's just not for me. I can remember what it was like." Mr. Moraise agreed."I plan to retire right here," Mr. Moraise said, referring the Barrington DPW."I'm very happy here."

Mr. Moraise and Mr. Grindle said they can recall quite clearly what work was like at Mega Disposal, which is based in Seekonk, Mass. Pay was based on the size of the route the individual worked — Mr. Grindle, who was a driver, said he was paid about $140 a day for his route. The pay did not increase year-to-year, meaning that he did not get a raise for the three years he worked for Mega. The laborer on the back of the truck earned less, added Mr. Moraise, who worked for Mega for about three years.

"I can remember I got one week vacation for the whole time I was there," Mr. Grindle said. "The co-pay for health coverage was very high... No sick days. You call out sick, you didn't get paid." Mr. Moraise said his family could not survive with the benefits offered by Mega.

Mr. Pedroza also said the town was getting a bad deal if they agreed to sell their seven refuse and recycling trucks to Mega for $350,000 as part of the deal. "One of those trucks is just a one-year-old vehicle we have right now. We paid $220,000. They're offering to pay $350,000 for all the vehicles? No way. Trade-in value we'd get more for the vehicles," he said.

Both men said the town would lose control over the collection services if they contracted with a private vendor. "Right now, the town has control over us, they hire us, they can fire us," Mr. Braga said. "Another contractor, the town will have no control over who they bring in. Are they safe workers, felons? Does the town want to subject its residents to that?"

The DPW workers also said there was no need to switch from the bi-weekly recycling pickup to the once-a-week pickup that Mega is promising. "We're looking at eight guys here with families," said Mr. Pedroza. "I suggest they love working here for this town. I don't see you going out and getting another company. You don't know what you're getting. You get rid of these trucks, you get rid of these guys and you're at the mercy of Mega. You abide by their rules."

7

Barrington Town Council Vice President Jeff Brenner said that officials owe it to the public to at least examine potential cost-saving measures. "It's not just about the money. It's about the money and the services. If you can get the same level of services for less money, it makes sense to privatize," he said, adding a few moments later, "I've lived in town for 19 years. I have no complaints on the services of the DPW. Mr. Brenner, along with Council President June Speakman, voted against drafting the contract with Mega. Both Mr. Brenner and Ms. Speakman said they did not feel that there was enough evidence for savings that could be made with the privatization switch. They also voiced displeasure at the pending termination of eight Barrington Department of Public Works employees.

During the meeting, Mr. Brenner tried to explain the situation as he saw it: He said there were two plans — Plan A is the town's current situation where the department of public works handles all refuse, recycling and yard waste collection; Plan B is hiring a private contractor. "I just don't see where the savings merit going from Plan A to Plan B. So, for that reason I would not move forward of privatization of collection services ... for this year," he said. Ms. Speakman said she would support contracting with a private vendor for the collection services if the savings warranted it, but after examining the figures proposed by Barrington Town Manager Peter DeAngelis and Finance Director Dean Huff, she said making the switch does not make good sense.

"For me, these men took a zero percent raise, and they were fabulous in the aftermath of the storm. Then to tell these men that the town will terminate their positions, I thought that was not the way to proceed. ... I did not find the savings to support the change," Ms. Speakman said.

Councilors Kate Weymouth and John Lazzaro both supported the town manager's initiative to privatize. Ms. Weymouth said she trusted Mr. DeAngelis' management skills and leadership. Mr. DeAngelis has said that as part of the deal with Mega, the private company would offer jobs to the eight DPW workers terminated from town employment. "We aren't considering the fact that Peter has a pretty good track record of negotiating contracts," Ms. Weymouth said

Residents react

The public attending the council meeting appeared mostly opposed to the privatization of refuse, recycling and yard waste collection services. Residents Mary Teixeira, Joel Hellmann and Mark Mancino spoke against contracting with a private contractor while Dan Justynski shared comments supporting the move.

Mr. Justynski said he manages millions of square feet of working space for his own job and continually questions why the town is still in the rubbish-collecting business. He added that

8

measures could be taken to ensure the performance of the private trash collection company. "There are ways to contractually obligate the vendor," he said.

Ms. Teixeira said the town has gotten itself into this predicament by deferring the purchases of new trash and recycling trucks. Now, she added, officials seem willing to sacrifice their control over the service."I don't live here for the schools. I live here because the streets are swept, the roads are plowed. ... We have a beautifully manicured town hall lawn. ... I don't see the point in give up our control to somebody else."Ms. Teixeira also questioned the amount of savings that could be realized by switching to a private collection service.

Mr. Hellmann told the town manager and councilors that eliminating the refuse and recycling department from the DPW is not the answer to budgetary issues. He said the trash collectors did a standout job following the floods in the spring, and deserve the town's praise not the threat of termination."I can't support privatization.

Mr. Mancino wasted little time voicing his position on the issue: "I strongly oppose the outsourcing of trash collection. These DPW workers aren't disposable ... what about the human element here? Our quality of life in Barrington exceeds a lot of other cities and towns. Have we looked at any other areas to save money — administrative positions, the consolidation of purchases with other municipalities for vehicles? "Why do we look at personnel before we look at purchases? Let's not make any mistake about it, I can assure you the Mega wages and benefits are inferior to what the town is paying. Otherwise, they'd be working at Mega right now. I don't minimize your decisions, but I think it's about time we treat our municipal employees with dignity rather than as digits on a spreadsheet."

9

Exhibit I

Income statement for the Town of Barrington

Town of Barrington, RI		
Income Statement		
For the Fiscal Year Ended June 30, 2009		
Revenue:		
Program Revenue:		
Charges for Service	$ 6,118,840.00	
Grants & Contributions	$ 6,108,682.00	
General Revenue:		
Property Taxes	$ 49,739,166.00	
State Aid & In Lieu of Tax	$ 3,574,912.00	
Other Revenue	$ (116,935.00)	
Total Revenue:		**$65,424,665.00**
Expenses:		
General Government	$ 2,292,203.00	
Public Safety	$ 6,724,665.00	
Public Works	$ 4,014,723.00	
Public Library	$ 1,379,549.00	
Senior Services	$ 130,052.00	
Cemetery	$ 124,133.00	
Recreation	$ 478,607.00	
Inspections	$ 100,900.00	
Conservation of Health	$ 46,500.00	
Community Services	$ 271,691.00	
Interest on Long-Term Debt	$ 836,727.00	
School	$ 44,462,190.00	
Sewer	$ 3,379,443.00	
Total Expenses:		**$64,241,383.00**
Net Income for 2009:		*$ 1,183,282.00*

10

Exhibit II Balance Sheet

Town of Barrington, RI		
Balance Sheet		
June 30, 2009		
Assets:		
Current Assets:		
Cash & Equivalents	$ 18,362,471.00	
Investments	$ 4,447,821.00	
Accrued Interest Receivable	$ 207,575.00	
Real Estate & Property Tax Receivable (net)	$ 2,366,514.00	
Betterment Assessment Receivable	$ 4,080.00	
Sewer Use Receivable	$ 3,021,768.00	
Due from Federal & State Government	$ 991,691.00	
Other Receivables	$ 358,138.00	
Inventories	$ 15,145.00	
Total Current Assets:		$ 29,775,203.00
Non-Current Assets:		
Land	$ 12,148,064.00	
Depreciable Buildings, Equipment, etc (net)	$ 69,973,072.00	
Construction in Progress	$ 12,715.00	
Other Assets (net)	$ 576,673.00	
Total Non-Current Assets:		$ 82,710,524.00
Total Assets:		**$ 112,485,727.00**
Liabilities:		
Current Liabilities:		
Accounts Payable & Accrued Liabilities	$ 3,919,320.00	
Due to State Government	$ 173,388.00	
Claims Payable	$ 203,164.00	
Accrued Interest Payable	$ 199,101.00	
Compensated Absences Payable	$ 58,885.00	
Long-Term Debt - Current Portion	$ 3,837,562.00	
Total Current Liabilities:		$ 8,391,420.00
Non-Current Liabilities:		
Long-Term Debt	$ 28,701,757.00	
Compensated Absences Payable	$ 603,415.00	
Other Post-Employment Benefits (net)	$ 449,393.00	
Total Non-Current Liabilities:		$ 29,754,565.00
Total Liabilities:		**$ 38,145,985.00**
Total Net Assets:		**$ 74,339,742.00**

11

Can we afford to be in the Trash Collection Business? – Is Privatization the Answer?

CASE SYNOPSIS

The town of Barrington was faced with the immediate replacement cost of two refuse and recycling vehicles. The town manager thought this was the time to consider privatization of trash collection. It would also mean the termination of eight long time employees. Bids were accepted and the low bid of $828,000 was weighed against the future cost to maintain trash collection within the town department of public works.

The town must answer the question — would contracting with a private vendor save taxpayers money? And which of several capital projection costs should be used? This case was developed from primary sources from Town Council meetings, financial reports from the town and cost projections provided by the Town Manager.

COURSES AND LEVELS FOR WHICH CASE IS INTENDED –

Courses: Managerial Accounting, Entrepreneurship, Human Resource Management, Public Administration
Level: Undergraduate

KEY ISSUES –
1. Decision analysis
2. Cost comparison – projection of costs
3. Advantages/disadvantages of privatization
4. Human resources – personnel decisions

RESEARCHMETHODS-

1. Data Sources –
   - Town Council Meetings
   - Town of Barrington Financial Statements – Fiscal Year Ended June 30, 2009
   - Town of Barrington Refuse Recycling Contract
   - Cost projections prepared by Town Manager
   - Town Planning Board – Recommended Capital Improvement Program report

12

DISCUSSION QUESTIONS

1. Prepare a cost analysis for the 2 scenarios for a 5 year projection–
   - leaving the trash collection within the town of Barrington by the department of public works
   - Contracting services with MTG Disposal

   The capital costs differ between the town manager, the planning committee and the union. The costs provided are for a 10 month period therefore the amounts have to be annualized before computing projections.
   See attached exhibits 1-3

2. Privatizing Trash Collection – what are the advantages and disadvantages

Advantages:
   - Contractor buys all equipment and pays for upgrades, maintenance and repair, as well as their employees' salaries and insurance.
   - Price is set: the exact cost is known for the next 5 years.
   - No labor disputes: Outsourcing trash collection eliminates the town's need to involve itself in union negotiations or salary disputes.
   - Less risk: If a contracted employee does something wrong and is sued, the contractor, not the city, assumes full financial and legal liability.
   - Guarantees: contracts include performance standards and completion guarantees.
   - Tangible results: Residents can measure whether contracts are working by results: reliable trash pickups.
   - Pay for what you want: A contract can be tailored to the needs of the residents.
   - Incentive to provide quality of Service - profits ultimately which depend on the quality of service being provided and the public response to it.

Disadvantages

   - Union resistance: Strong, politically influential unions will surely lobby against losing their jobs to an outsourced firm.
   - Rate hikes: Outsourced departments could raise rates at renewal time.
   - Forfeits: If a contractor forfeits the job or goes out of business, the city may have to scramble for a replacement, possibly inconveniencing residents.
   - Loss of control of the personnel performing the service
   - Since the equipment was sold it is difficult to change the decision made to privatize

13

REFERENCES

1. Rhode Island Economic Development Corporation. Accessed November 1, 2010, from www.riedc.com/dataandpublications/municipal-tax-rates
2. Town of Barrington Rhode Island, Comprehensive Annual Financial Report, For the Fiscal Year Ended June 30, 2009.
3. Barrington Town Council Meetings 5/26/10, 6/7/10, 7/5/10, 8/2/10

EPILOGUE:

Members of the Barrington Town Council voted 3-2 in favor of drafting a contract with the private waste collection company, Mega Disposal, at a meeting on Monday, Aug. 2 at the town hall utilizing the projection of costs submitted by the town manger.

14

THE OIL SPILL CASE OF BRITISH PETROLEUS

Joshua Banger, Richard Stockton College of NJ, e-mail: bangerj@go.stockton.edu
Shaoping Zhao, Richard Stockton College of NJ, e-mail: szhao@stockton.edu

ABSTRACT
This case discussed and analyzed issues and consequences of last year's explosion on the Deepwater Horizon oil platform of British Petroleum (BP) in the Gulf of Mexico. It also discussed safety violations of BP and corresponding operations improvement decisions.

Introduction

British Petroleum (BP) is one of the largest oil corporations in the world, and it produces more oil and gas for the United States than any other company [10]. British Petroleum wields a great deal of power and influence because of this. The company's business category and financial strength are showing in the table below from BP's Annual Report and Accounts 2009

Facts and figures:

Sales and other operating revenues [1]	$239 billion (year 2009)
Replacement cost profit [2]	$14.0 billion (year 2009)
Number of employees	80,300 (at 31 Dec 2009)
Proved reserves[3]	18.3 billion barrels of oil equivalent
Service stations[4]	22400
Exploration and production	Active in 30 countries
Refineries (wholly or partly owned)	16
Refining throughput	2.3 million barrels per day (year 2009)

For further information, please see the Annual Report and Accounts 2009 on BP's website

The company first began in 1901 when William Knox D'Arcy was granted the right to search for oil in the desert of what was then known as Persia. Little progress was made by D'Arcy until 1908, when he discovered the greatest oil field of its time. The **Anglo-Persian Oil Company was formed out of this find.** Winston Churchill realized the potential for an untold amount of wealth and oil for the British military and with his help, Britain became the major stockholder in this company [12].

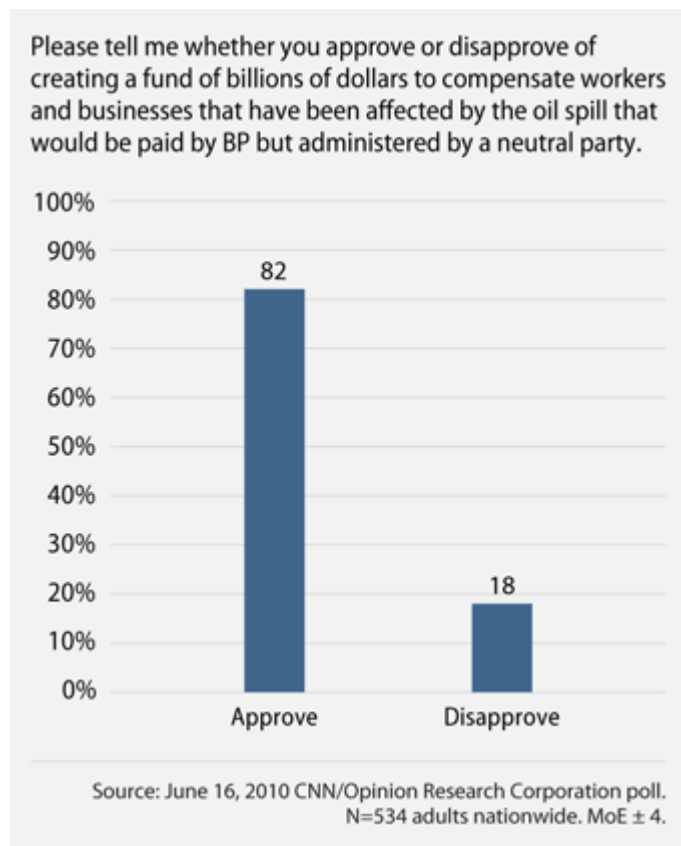Throughout the 1920s and 1930s, the Anglo-Persian Oil Company was very profitable as the world started to rely more and more on oil. However, their workers often toiled in brutal conditions where disease and discomfort flourished. Many of the native citizens resented the company's presence because of the way their workers were treated. In 1935, the company renamed itself to the Anglo-Iranian Oil Company [12].

In 1954, the company renamed itself again to the British Petroleum Company [10]. It sold countless millions of barrels of oil to the Western World from the Middle East and started to expand its borders in the 1970s. In the 1980s and 1990s, British Petroleum started to gain a large presence in the United States [12]. This influence in the United States has been wrought with disasters in recent years due to gross negligence from the high level managers. The most notable was an explosion on the Deepwater Horizon oil platform in the Gulf of Mexico, which occurred on April 20th, 2010. This disaster killed eleven workers and is considered to be the worst oil spill in history [4].

Current Issues of BP Spill

The Deepwater oil spill ended on July 15, 2010, and the company faced consequences during the event and will continue to face them for a long time afterwards. On June 16, President Obama pressured BP to make it pay for the damages it caused. The corporation had to set up a $20 billion escrow to pay for the damages to the local businesses and workers impacted by the oil spill. (See Chart 1)

Chart 1

Please tell me whether you approve or disapprove of creating a fund of billions of dollars to compensate workers and businesses that have been affected by the oil spill that would be paid by BP but administered by a neutral party.

	Approve	Disapprove
	82	18

Source: June 16, 2010 CNN/Opinion Research Corporation poll.
N=534 adults nationwide. MoE ± 4.

This graph shows that the large majority of people want BP to reimburse those who have been negatively impacted by the spill that occurred on April 20, 2010.

It also had to suspend dividend payments to shareholders to make sure that it would be able to meet these and any future obligations for this oil spill. This escrow account has been taken over by an outside party to make sure that the money is distributed fairly. This happened after numerous complaints were filed about the complexity and unfairness in using British Petroleum's own claims processing center. President Obama also placed a six month freeze on deep sea drilling in the Gulf of Mexico and required BP to give $100 million to the workers who were idle because of this [15]. President Obama made it clear that this was not the limit of the company's liability because the full extent of the disaster has not been fully realized. The company's stock had taken a massive hit because of the disaster itself, and the stock became even less desirable when investors got word that the company would be liable for such a large sum of money. The fact that dividend payments were halted also made the stock unattractive for investors. To this day, BP stock has got up some, but it is nowhere near where it was in the time preceding the oil spill. The oil spill had cost the company $32.2 billion for the third quarter of 2010, resulting in a net loss of $16.97 billion during this time [17].

In addition to the amounts that the United States government is making BP pay to the workers and businesses affected by their actions, the Clean Water Act states that BP can be fined for $1,100 per barrel of oil leaked and up to $4,300 per barrel if it is found that criminal negligence caused the accident [14]. The spill lasted for roughly 3 months and government estimates place the amount of gallons of oil leaked to be about 206 million, so the oil giant is at risk of losing billions of dollars more. Currently, BP lawyers are challenging these governmental estimates in an effort to lower the fines. They are claiming that the government had used invalidated assumptions and estimates that did not take into account some of the equipment BP had in place that would have lessened the impact. The company is expected to argue that the real damage is between twenty percent and fifty percent lower than what the government places the estimate at, but it is unlikely that it will win this argument in its entirety. The government used multiple independent experts utilizing different techniques to arrive at its figures, so BP will have to present an extraordinarily persuasive case. However, the amount that BP will actually be fined has yet to be determined. It could save between $1.1 billion and $10.5 billion if it wins the case.

Class action lawsuits have been a frequent occurrence for BP since the oil spill. These lawsuits include oil related damages and the controversial methods that BP employed to clean it up. One case was on June 18, 2010, when the Eastern District of Louisiana filed a case against the company for the use of the chemical Corexit in the clean-up process. The plaintiff argued that British Petroleum knew the toxicity of this chemical, but ignored the risks in an effort to clean the oil up as quickly as possible [13]. This dispersant is said to be more toxic than the oil itself, but it is less visible.

British Petroleum is currently on thin ice in the eye of the public and the government, and it must work to prevent another catastrophe from happening. There are currently numerous requests and petitions for either the assets of British Petroleum to be seized and for the disbanding the company because of criminal negligence. If another BP oil disaster occurs in the near future, the company could be dealt a fatal blow. Investors could lose all faith in the company or the United States government could ban the company from operating anywhere near U.S. territories.

To this day and for the foreseeable future, British Petroleum will be facing the ramifications of the Deepwater Horizon oil spill. The company chose to hide as much of the environmental damage as it could from the public by spraying the affected areas with dispersants. These chemicals caused the oil to go under the surface of the ocean and millions of gallons of oil are currently hidden from sight. Previous oil spills, including the Exxon Valdez spill in 1989, show that oil traces can persist around the disaster site decades after the incident [16]. Clean-up crews do their best efforts to contain the damage, but no amount of cleaning can completely restore the affected areas back to normal. For years to come many sea creatures that fishermen depend on for their livelihoods will have too much oil to be used for human consumption and hurricanes could push the hidden oil into the swamps and other coastal areas of the Gulf of Mexico. It is still too early to tell how many sea creatures perished as a result of this spill, including dolphins, sea turtles, and endangered fish, but the public will sure to have a new surge of distaste for the company responsible for the worst environmental disaster in United States history. Undoubtedly, BP will spend billions of dollars on fines, lawsuits, and public relations over the next several years.
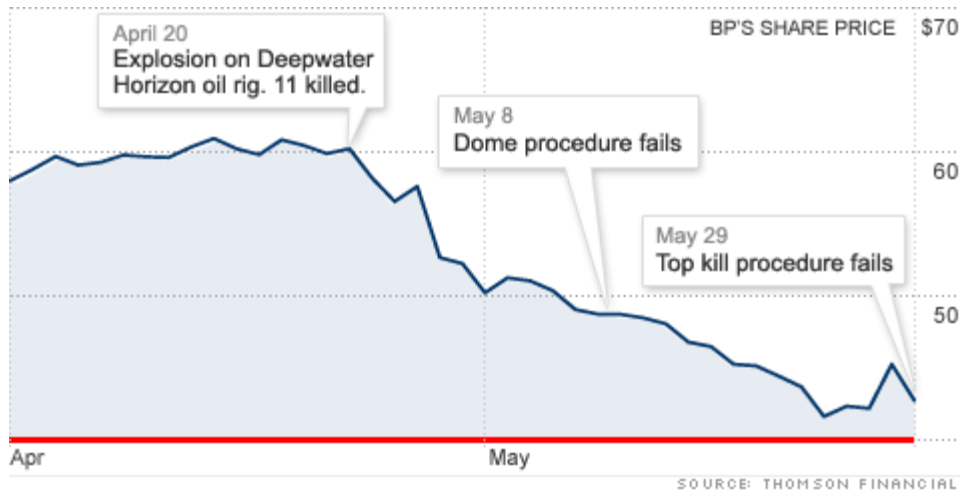
The Deepwater Horizon oil spill ended months ago, but the ramifications of this event will last for years for both the environment and British Petroleum. New findings are going to come up over the next year or two. These will decide if criminal negligence was the primary cause for the event, the extent of the damage, any long-term restrictions that will be put on the company, and many other things that will affect BP at this pivotal moment in its life.

Safety Violations

Tom Jones has just been promoted to CEO after the former CEO was forced to step down due to an overwhelming demand by stockholders and the public for his termination following BP's latest catastrophe. He knows that his company is on the verge of another tragedy. British Petroleum faces a problem in their lack of compliance with safety regulations because it has been placing short-term profits over basic safety and ethical practices for years [9]. Jones knows that their priorities need to change drastically not only for their financial sake, but also for the sake of the millions of people who are affected when their carelessness leads to disasters. Jones is a more compassionate man than his predecessor, and he is more concerned about improving safety for the environment, the local people, and his employees. Many local businesses are often affected when these oil spills occur and most people want the responsible parties to compensate them for their losses. He is also highly motivated to keep his company in the best shape possible, which is why he was recently elected to CEO. He is an experienced businessman and knows that an oil spill can be extremely detrimental to the company in asset damages, clean up costs, lawsuits, and stock prices (see Chart 2 on next page).

Tom Jones pores over his company's records in their entirety for the first time. The amount of safety regulations broken by British Petroleum is astonishing, and he cannot believe how badly things were being run. Corners had been cut at nearly every possible opportunity by his company's managers to save money in the short term. He does a quick comparison to other similar corporations and finds that ninety-seven percent of all of the Occupational Safety and Health Administration violations in the refining industry were found in BP's oil refineries (see Chart 3 on next page).

Chart 2

This graph shows the rapid decline of British Petroleum's stock prices following the Deepwater Horizon oil spill on April 20, 2010. The company lost roughly half of its stock value at its lowest point [6].

Chart 3

This graph illustrates the number of OSHA safety violations of the top five oil companies. Note that British Petroleum is far above any of the others (Overview of BP's Safety Record).

Jones is appalled as he pores over the hundreds of safety violations that his company is currently breaking, with thirty of them being serious risks [9]. He sits at his desk and thinks to himself how greedy the decision-makers at British Petroleum had really been to have such a blatant disregard for safety.

Jones plans to put a stop to this if he can. He knows that at one point cutting safety expenses may have been profitable, but now that the company has grown so complex and large, these safety concerns are going to come back to haunt him and the corporation. With so many outstanding violations, disasters are nearly guaranteed to occur at some point if nothing is done to fix them. When the inevitable disasters do happen, their money-saving tactics backfire. Not only did they lose about half of their stock value at one point because of the Deepwater Horizon spill, but they had to place $20 billion in an escrow fund to pay to those financially impacted by the environmental damage [2]. "Why has no one at this company learned how bad it is to break all of these safety regulations?" he asks to himself. "There is no question that eventually we will lose more money than we saved." The problems are so largely ingrained in the operation of British Petroleum that Jones has trouble even figuring out where to begin to help improve his company before another disaster struck.

Improving Operations Management Decisions

British Petroleum will need a thorough and long process to fix these problems, and Jones decides that the best way to implement changes is to start from the top. He plans to schedule a video conference with his regional managers to tell them of the company's drastic change in management practices. First, though, he goes over in his head how to implement some of these changes. He feels that new policies need to be put into place to prevent future incidents from happening. Sparing no expense on safety equipment, better training and treatment of workers, and changing where BP operates are all things that need to be addressed.

At the strategic level, BP will have to rethink their answer to the question of where to locate their facilities. The Deepwater Horizon spill lasted for months because that oil platform was located in very deep ocean water. When disaster struck, no humans could go down that deep to fix the problem, and it took too long for an effective solution to be implemented with their current levels of technology. The company currently has vital equipment in places where they have no way of reaching in a timely manner in the event of an emergency. The costs of researching and developing appropriate submersibles for all of their deep ocean oil platforms will be extremely high at this point in time, though. Jones weighs the benefits against the risks and decides that these oil platforms had been a bad idea from their inception and makes the decision to terminate their use because it in the best interest in the company. The costs saved in drilling for oil in deep water, where oil is more plentiful, is not worth the cost of developing and testing the necessary safety equipment and vehicles until current technologies improve.

On the tactical level, Jones considers the many problems relating to his workforce. Due to the lack of adequate training, BP faces a shortage of trained workers. He used to be a general laborer at one of BP's oil platforms in the United States for years until he worked his way up the corporate ladder. He knows firsthand of how clueless he had been whenever things out of the ordinary occurred and of how long his shifts had often seemed. He reviews some employee

timesheets and sees that frequently the workers have been working sixteen to eighteen hour shifts to compensate for the shortage of skilled workers [8]. He knows that a person working for such a length of time will not be fully alert to what is going on around him or her. This causes avoidable accidents to occur because of inevitable worker carelessness. Jones needs to emphasize the importance of adding more shifts and spending more resources to train the workers. He knows that it takes years to fully train an oil employee to fully understand how to react in virtually any possible scenario, but he finds it necessary. This will ease the burden on the few workers who are qualified to do this work by splitting the load amongst more people.

He begins thinking on the operational planning level. British Petroleum needs to make some changes to which workers it assigns to specific tasks and to which jobs have priority. Once the tactical planning has been implemented, these newly trained workers need to be placed only in areas where they are confident in. The workers who are in charge of safety and vital equipment need to have a very high priority, and only select individuals will work with these pieces of machinery. They must also be given only the best equipment available to assist with their jobs.

Jones knows that working conditions had not always been ideal for the lower-level workers in the United States, but he is shocked when he does some research and sees how bad it is for the workers in parts of the world. This is especially true in countries where labor laws are not as well enforced as in the United States. Those who work for BP in Columbia have been heard complaining about low wages, no job security, unsafe working conditions, and a number of other disturbing things [5]. Jones makes a note to emphasize the importance on improving working conditions for all employees to his managers. Jones will use some of the corporation's discretionary income to go towards making life better for these people. Taking steps toward better job design and worker empowerment will benefit the company, the workers, and society as a whole. This will lead to happier workers, more safety, more productivity through better efficiency, and a better public image.

Even with all of these radical changes, Jones realizes that oil drilling is a risky business. He has thought long and hard about the things that his company should do in the aftermath of an unavoidable disaster to reduce damages. Jones feels that being honest with the public is the best way to improve their image when things go wrong. He remembers how the former CEO spent too much effort and money on improving their public image and initially lied about how much oil had been pouring into the Gulf of Mexico. This tactic failed as the oil continued to spew out of the ruined well, and the problem became worse and worse because nobody knew how to stop it. He will make sure that all nearby operations halt in the event of an oil rig explosion until the disaster has been contained. All available personnel will be sent over to quickly repair the damage and clean up the spilled oil and debris.

Tom Jones met with the regional managers and discussed his new strategies and policies for the company. Though radically different than they were used to, these managers had always done as they were told and immediately told their subordinates of the changes and made sure that they were implemented. Jones made it very clear that their jobs depended on it. It took almost a decade for these changes to work their way down to the lowest levels of the corporation, but eventually the company became unrecognizable. Avoidable disasters ceased and British

Petroleum became known worldwide as an ideal place to build a career, investors saw slower but steadier profits, and it flourished overall.

Restoring Public Image

Public image, business reputations are vital for BP back to the feet and recover its operation. Strategically, damage controlling is the first priority BP has to take care for after the disaster. Frankly, the company didn't do any good in the beginning. Covering up, lighting the affect result, even lowering spill total were scattering from BP interview on TV and news paper. Under sharp critics, BP improved. The company sets up a BP's Gulf Coast Restoration Organization and opens a communication place on its website "Gulf of Mexico response". Public can access the place for updating video of Gulf Coast Restoration progress and Facebook Q&A sessions handled by BP CFO, COO and VP respectively. These efforts may pay off positively for reviving BP business image in the long run.

REFERENCES

[1] Bergin, Tom. "Analysis: New BP Boss Should Boost Safety, Asset Sales." *International Business Times*. 26 July 2010. Web. 12 Sept. 2010. <http://www.ibtimes.com/articles/38408/20100726/analysis-new-bp-boss-should-boost-safety-asset-sales.htm>.

[2] "BP Sets Aside $20 Billion Fund For Oil Spill Victims." *News One*. Associated Press, 16 June 2010. Web. 12 Sept. 2010. <http://newsone.com/nation/associatedpress2/bp-sets-aside-20-billion-fund-for-oil-spill-victims/>.

[3] Davis, Mark M., and Janelle N. Heineke. *Operations Management: Integrating Manufacturing and Services*. Boston: McGraw-Hill, 2005. Print.

[4] Frayer, Lauren. "Crews in Gulf of Mexico Begin 'Static Kill' Procedure." *AOL News*. 3 Aug. 2010. Web. 12 Sept. 2010. <http://www.aolnews.com/gulf-oil-spill/article/bp-tests-for-static-kill-as-spill-deemed-worlds-largest/19578603>.

[5] Hall, Claire. "British Petroleum in Colombia: Colombian Army Attacks Striking BP Workers." *Global Research*. 7 June 2010. Web. 19 Sept. 2010. <http://www.globalresearch.ca/index.php?context=va&aid=19600>.

[6] Hargreaves, By Steve. "BP's $70 Billion Whipping." *CNN*. 02 June 2010. Web. 12 Sept. 2010. <http://money.cnn.com/2010/06/01/news/companies/BP_analysts/index.htm>.

[7] Krause, Thomas R. "What Caused the Gulf Oil Spill?" *The Corporate Social Responsibility Newswire*. 19 Aug. 2010. Web. 12 Sept. 2010. <http://www.csrwire.com/press_releases/30321-What-Caused-the-Gulf-Oil-Spill->.

[8] Leopold, Jason. "EXCLUSIVE: New Documents, Employees Reveal BP's Alaska Oilfield Plagued by Major Safety Issues." *Truthout*. 15 June 2010. Web. 19 Sept. 2010. <http://www.truth-out.org/documents-employees-reveal-bps-alaska-oilfield-plagued-by-major-safety-issues60470>.

[9] Morris, Jim. "Renegade Refiner: OSHA Says BP Has "Systemic Safety Problem"." *The Center for Public Integrity*. 16 May 2010. Web. 12 Sept. 2010. <http://www.publicintegrity.org/articles/entry/2085/>.

[10] Press, James. "The British Petroleum Company Plc." *Funding Universe*. 1998. Web. 12 Sept. 2010. <http://www.fundinguniverse.com/company-histories/The-British-Petroleum-Company-plc-Company-History.html>.

[11] Teixeira, Ruy. "Public Opinion Snapshot: Public to BP: Pay Up, Clean Up." *Center for American Progress*. 21 June 2010. Web. 19 Sept. 2010. <http://www.americanprogress.org/issues/2010/06/snapshot062110.html>.

[12] Tharoor, Ishaan. "A Brief History of BP." *TIME*. 2 June 2010. Web. 12 Sept. 2010. <http://www.time.com/time/business/article/0,8599,1993361,00.html>.

[13] "BP & Nalco Holdings Class Action Lawsuit Filed Over Alleged Spraying Of COREXIT®9500 In Gulf Of Mexico Oil Spill." *ClassActionLawsuitsInTheNews*. 18 June 2010. Web. 07 Dec. 2010. <http://classactionlawsuitsinthenews.com/class-action-lawsuits/bp-nalco-holdings-class-action-lawsuit-filed-over-alleged-spraying-of-corexit®9500-in-gulf-of-mexico-oil-spill/>.

[14] Cappiello, Dina. "BP Challenge To Oil Spill Size Could Affect Fine." *Breaking News and Opinion on The Huffington Post*. 3 Dec. 2010. Web. 07 Dec. 2010. <http://www.huffingtonpost.com/2010/12/03/bp-challenge-to-oil-spill_n_791970.html>.

[15] Mason, Jeff. "BP Agrees to $20 Billion Spill Fund, Cuts Dividend." *Business & Financial News, Breaking US & International News | Reuters.com*. 16 June 2010. Web. 07 Dec. 2010. <http://www.reuters.com/article/idUSN1416392020100616>.

[16] Nelson, Karen. "Effects of BP's Oil Spill Might Persist for Years." *McClatchy*. 8 Aug. 2010. Web. 07 Dec. 2010. <http://www.mcclatchydc.com/2010/08/08/98767/effects-of-the-bps-oil-spill-might.html>.

[17] "Oil Spill Bp 2010 Wiki." *Bp Oil Spill Details*. 2 Aug. 210. Web. 07 Dec. 2010. <http://bpoilspilldetails.co.cc/oil-spill-bp-2010-wiki.php>.

LYCOMING ENGINES' LEAN TRANSFORMATION

Pamela Wynn, Bloomsburg University, 400 East Second St. Bloomsburg PA, 17815,
pwynn@bloomu.edu, 570-389-4591
Christian Grandzol, Bloomsburg University, 400 East Second St. Bloomsburg PA, 17815,
cgrandzo@bloomu.edu, 570-389-4521

Abstract

This teaching case documents the radical transformation of Lycoming Engines from an
organization losing millions of dollars to a Shingo Silver Medallion in just 4 years through the
rigorous implementation of Lean Six Sigma philosophy and methodology.
Keywords: Lean Six Sigma, Transformation Leadership

Introduction

This case was inspired by a field trip taken by students in Bloomsburg University's Operations
Management courses to Lycoming Engines in December 2007. At the time, Lycoming Engines,
an aircraft engine manufacturer, was embarking upon its' most aggressive operations-based Lean
improvement efforts. Upon arrival, the students were greeted by leaders and employees of
Lycoming, who displayed passion and humility as they relentlessly pursued a vision that would
transform it from a failing operation to a leading example of the transformative power of Lean
principles and tools.

Lycoming's Lean success was affirmed in 2008, when it was awarded the prestigious Shingo
Silver Medallion, perhaps the most rigorous external validation of Lean in the country. In
evaluating Lycoming's performance for the award, lean experts spent days turning over every
stone of their operation for the assessment. When we learned of the award, we approached
Lycoming and asked if we could tell their story.

This case fills a gap in the management case literature particularly in the area of Operations
Management by affording students and opportunity to engage in decision making, analyze
management's actions, and apply their disciplinary knowledge in the context of a Lean
transformation. The case examines the cultural enablers of a Lean transformation, the use of
Lean Six Sigma principles and tools, the time horizons of a Lean transformation, the leadership
qualities embodied by transformational leaders, and the use of policy deployment. It is targeted
for use in upper-level, undergraduate business courses such as Operations Management, and it is
also appropriate for Organizational Behavior, Supply Chain, Labor Relations, and other
undergraduate and graduate Management courses.

Overview of the Case

This case describes Lycoming Engines, the world's leading producer of piston aviation engines,
as it emerges from a company on the brink of failure to a premier manufacturer. In 2004,

Lycoming was a fairly typical example of the mass production orientation. There were large batch sizes, mountains of inventory, long lead times, little in the way of long-term supplier relationships, and lengthy product development cycles. Sixty percent late on orders was considered normal, and nearly 80% of engines that were delivered to customers had at least one quality issue. Furthermore, Lycoming had recently experienced a crisis of failed engines that resulted in a surge of liability costs from less than .5% to over 10% of gross earnings and an impending loss of their reputation.

Lycoming also had a broken relationship with its unionized workforce. Massive outsourcing and a union strike in the 1990s resulted in a hostile relationship between the workers and management. The two sides had failed to agree to a contract for nearly 9 years. Excuses, finger-pointing, lack of respect, and poor safety records dominated the workplace. It was so bad that former employees once said that they had "escaped Lycoming."

Combined with the effects the 2001 9/11 attacks had on the aviation industry, Lycoming was decimated. The company was losing tens of millions of dollars per year, incurring its biggest operating loss of $40 million in 2002. When compared to a roughly $125 million annual revenue base, the prospects for reversing the losses were dim.

In 2004 Ian Walsh was hired as the Senior Vice President and General Manager of Lycoming. He started the company's transformation by bringing in a new leadership team, developing insight into the company's problems, crafting a new vision for the future, and rebuilding the relationship with the workforce. He initiated and sustained a rigorous Lean Six Sigma program that radically changed the culture, mindset and outcomes at Lycoming. The results were record revenue levels, superior competitive positioning, and the celebrated Shingo Silver Medallion in 2008.

However, the lull in challenges facing Lycoming since its transformation would be short lived. The dramatic economic declines that accompanied the recent recession hit Lycoming hard once again. It had to layoff roughly 25% of its workforce at a time when the employees had finally regained confidence in management and bought in to Lean Six Sigma. Confronted with these severe new realities, Ian Walsh and his leadership team debated the next courses of action, and Walsh must now decide how to guide Lycoming through the new challenges presented by the economic downturn.

DECISION PROCESS INVOLVING OCCUPATIONAL HEALTH AND SAFETY ISSUES

Monique Lortie[1], Sylvie Gravel[2], Jessica Dubé[2], Henriette Bilodeau[2]

1. 1.Sciences biologiques, Université du Québec à Montréal (UQAM); 2. Département d'organisation et ressources humaines, École des sciences de la gestion, UQAM

ABSTRACT

In the administrative and decision sciences field, issues related to populations health have attracted a lot of interest, probably because the government is a major stakeholder and the financial stakes are enormous. On the opposite, little research and literature have focused on the decision process when occupational health and safety issues are concerned. Although the government is present through various boards [e.g. US' National Institute for Occupational and Safety Health and Occupation al Health and Safety Organization (NIOSH, OSHA), Québec' Commission de la Santé et Sécurité du Travail (CSST)] the decision process remains essentially internal to the organizations. Presently, the importance of the costs related to occupational problems is well acknowledged as well as the actual accelerated rising of the costs related to mental health problems. In Business School, occupational health and safety are seen as domains being under the responsibility of human resources. Everyone agrees that human resources management may be narrowly linked to the occupational health problem. However, we are often short of teaching material – especially in the form of documented case studies - to demonstrate these interactions. To overcome this gap, eleven case studies were developed. Basically, the cases involved organizations that were struggling with complex occupational health problems. The goal was to provide documents allowing the processing of various questions. Among these ones: 1. How the actual problem could be related to previous decisions? On the base of which elements did the organization decided to 'attack' and eventually solve the problem?

Study design : The targeted organizations were enterprises which were facing a complex occupational health problem involving the human resources management for their last three years. That problem could have not been solved when data were collected. Human resources or occupational health directors received an interview questionnaire in advance, in order to give their consent and to get prepared for the interview. Eight of the eleven interviews were realized with both human resources and occupational health directors. Interviews lasted around 90 minutes and were recorded. Most organizations provided datas and material (e.g. statistics, reports, etc.). The material was analyzed with Nvivo (version 8).

Results: Four issues will be presented and illustrated through extracts of different case studies: 1. Why the organization decided to take an active role in the solving of the problem (occupational problems are often recurrent). 2. How a (the) previous decision(s) played a significant role in the emergence of the problem to solve. 3. The interactions between the societal context and the decisions. 4. Difficulties to foresee the impact of decisions or actual situation on the human resources management.

1. **What decided the organization to act:** In all cases, the organization decided to move actively and to invest itself in order to solve the problem into a climate of major crises. In most cases, the problem manifested through an important increase in absenteeism or a severe lack of motivation at work. Usually, there were previous clear signals. In half the cases, the crisis developed over a short period of time. In the first case to be described, the sudden event that was decisive was four workers occupying strategic positions that went on a simultaneous sick leave. In the second case, a major epidemics event leads to a massive absenteeism problem.

2. **How previous decisions had a significant impact.** In most cases, the actual crisis was partly related to a previous decision where potential impact was not foreseen. The critical decision could come from the outside as well as from the inside. In the first case, the organization decided to introduce a ten hour shift schedule in order to attract outside workers. In the second, an outside organization decided to implant a new policy.

3. **The interactions with societal questions**. Most organizations felt deeply affected by societal context. Some tried to cope actively, other felt exceeded. In the first case, the organization changed its production to get access to new customers groups. In the second case, the organization, which target young consumers, faced with delinquent problems.

4. **Difficulty to foreseen impact.** The various cases show that organizations have difficulty to forecast how a decision based on production or economic criteria –or an absence of decision - could change the work situations or impair on the capacity of the workers to get adjusted or to accept a situation. In the first case, the organization underestimated the impact of losing competencies as the work was seen as simple. In the second case, the organization used to be seen as very a attractive one, has now difficulty to recruit and retain younger workers.

The discussion will focus on the elements that an organization needs to keep in mind to anticipate if a decision can have or not a major impact on the work, the workers and the occupational health and safety.

PUBLICATION OF SOCIAL RESPONSIBILITY NEWS AND ITS IMPACTS ON MINING SECTOR INVESTORS' BEHAVIOR

Karine Casault
MBA Student, UQAM
3700 rue Saint-Antoine Ouest, H4C 0B1
Montreal

Casault.karine@courrier.uqam.ca
k.casault@cbleue.com
514-924-3391

ABSTRACT

The debate over the main responsibility of enterprises is immemorial [5] [7], and opposes those who still claim that the sole responsibility of the company is to be profitable and those who claim that the responsibilities to various stakeholders are much more extensive. Shareholders are therefore a public trial for businesses. We therefore propose a study that consider the multiple and complex relationships between different actors including: Canadian mining companies, the home state enterprises (Canada), the media, investors, but also underlying variables, including social responsibility, the legislative framework, the history of the area, socially responsible investing, etc. Through a content analysis of corporate social responsibility news dealing with Canadian mining companies and their international activities, we are looking to investigate the link with their diffusion and the reactions of mining sector investors.

Keywords : CSR, Investor's Behavior, Media

PROBLEM

Corporate social responsibility (CSR) has been widely debated, sometimes in a controversial way, in the past three decades with no real consensus emerged. This debate is partly derived from the proliferation of different conceptualizations of CSR. The debate about the link between financial performance and social responsibility also remains unsolved. Indeed, empirical research on the relationship between CSR activities and corporate financial performance has brought only inconsistent results, influenced by a variety of methodological problems [4][1][6][9] [2] [11][8. However, almost all agree that it is highly strategic, according to a study of Oracle Corporation and The Economist Intelligence Unit, interviewing 136 managers from various industries and 65 investors[1], 85% of them put social responsibility as a paramount consideration in investment

---

[1] The Importance of Corporate Responsibility, The Economist Intelligence Unit, 2005.

decisions. This leads one to wonder exactly how much and what are the immediate impacts of the release of news about social responsibility on investors?

SCIENTIFIC OBJECTIVE

The aim of our study is to clarify the disapointing empirical results disappointing from previous scientific research and discover the presence of a link between the release of social responsibility news and investor behavior in order ultimately to promote socially responsible business in a given sector ; the mining sector, particularly at risk because of the nature of its activities. By adopting a different methodology and a specific industrial sector, the proposed study addresses the issue with an innovative perspective.

METHODOLOGY

The methodology privileges content analysis, both in terms of social responsibility news and shareholder resolutions passed at meetings. It is also envisaged to conduct a triangulation to gauge investors to determine the extent to which the publication of news, from the media or from the company itself will influence their behavior. Content analysis will be conducted using the press analysis method Morin-Chartier [3], which by hindsight it imposes, by its systematic approach to dissect all documents dealing with a particular organization, enables to identify scientifically fervor used by the media, in minimizing room for interpretation. It aims to identify and understand the meaning and impact of information transmitted by the media in interpreting both its quality and quantity. Through the method based on the key concept of "information unit", it is possible to go further in tracing in the press content, not only the most debated topics, but also the fervor of each medium, the evolution of news discourse in time, events or major issues that have affected the coverage, efficiency of spokespersons.

Mining companies whose market capitalization was the largest on the Toronto Stock Exchange TSX Venture were selected for analysis. Indeed, as their weight on the Canadian economy is important, the public eye, the eyes of shareholders, the eyes of government and media are turned to them. That is, the Toronto Stock Exchange (TSX) hosts the largest group of mining companies in the world. At the end of 2008, 55% of public mining companies were registered there, or 1,427 companies, compared to 684 to 216 in Australia and London.
The TSX is a global destination of international project finance. It imposes standards of scientific and technical stringent to registered companies (National Instrument 43-101). This standard is increasingly seen as the benchmark for global mining disclosure, which can lead to think that the disclosure of companies that are registered there will be more complete and uniform, allowing analysts and financial journalists to access the information they need and thus promoting dissemination.

CORPUS

We conducted a search through the Eureka database, which provides access to over 3204 electronic, web or written medias, daily or weekly, as well as press releases, newswires, and news agencies. For each company we reviewed all the items found with the search by name of mining company to retain only those dealing with social responsibility.

Table 1 : Corpus of Canandian mining companies studied

Mining company	International activities	CSR News	Total News Volume (2009)	CSR Report	Metals
Barrick Gold Corporation	Yes (South America, Australia, Africa)	135	1235	Yes	Gold, minerals
Potash Corporation of Saskatchewan Inc.	Yes	63	626	Yes	Potash
Goldcorp Inc.	Yes (Latin and South America)	37	760	Yes	Gold, minerals
Teck Resources Limited	Yes (Chile-Peru)	155	932	Yes	Copper, coal, zinc

HYPOTHESIS

Recent events in terms of social responsibility in the mining sector had a direct impact on the share price of the company involved inclined us to think that there is an immediate effect of the publication of news on social responsibility behavior of investors in the sector and it is measurable through the observation of the fluctuating stock price, but also by studying the behavior of shareholders.

For instance, the company Manhattan Minerals Corporation has seen its price falls 30% in the days following the holding of a referendum where the population of Tambogrande voted 93% against the continuation of the mining project [4]. Cambior has seen its price falls by 23% between Friday, August 18 and Monday, August 21, 1995 following the breakdown of the tailings dam of the Omai gold mine in Guyana which has resulted in the release of 4 million cubic meters of mine waste containing cyanide into the Omai River. Within five days after the announcement of the bursting of a dam of a tailings pond and waste rock storage of the Los Frailes mine in Spain, the action of the mining company Boliden Ltd. lost 28% of its value at the Toronto Stock Exchange. July 15th 2002, the Kivalina Relocation Planning informed Teck Cominco of its intention to proceed against the Company under the provisions of the Clean Water Act, against which the company has committed over 3,000 crimes. Within five days following its announcement, the share value of the company fell 10% [4].

Although these events are major and mean significant costs for the companies involved, they certainly affect their assets and have a direct impact on investors, these cases suggest that events of social responsibility interest and modify the behavior of investors.

By the NEDSI Conference in April, we plan to be able to present our results.

REFERENCES

[1] Aupperle, K.E.; Carroll, A.B. and Hatfield, J.D. (1985) An Empirical Examination of the Relationship between Corporate Social Responsibility and Profitability. Academy of Management Journal, 28(2): 446-463.

[2] Balabanis, G.; Phillips, H.C. and Lyall, J. (1998) Corporate social responsibility and economic performance in top British companies: are they linked? European Business Review, 98(1): 25-44.

[3] Chartier, L. (2003). Mesurer l'insaisissable : méthode d'analyse du discours de presse. Sainte-Foy, Presses de l'Université du Québec.

[4] Cochran, P.L. and Wood, R.A. (1984). Corporate Social Responsibility and Financial Performance (1984) Academy of Management Journal, 27(1).

[5] Dodd, E. M. (1932). "For whom are corporate managers trustees?" Harvard law review **45**(7): 1145.

[6] Gray, R.; Kouhy, R. and Lavers, S. (1995) Corporate Social and Environmental Reporting. Accounting, Auditing & Accountability Journal, 8(2): 47-77.

[7] Hayek, F. A. (1969). "Three elucidations of the Ricardo effect." The journal of political economy **77**(2): 274.

[8] Moore, G. (2001) Corporate Social and Financial Performance: an Investigation in the UK Supermarket Industry. *Journal of Business Ethics*, 34: 299-315.

[9] Orlitzky, M; Schmidt, F.L. and Rynes, S.L. (2003) Corporate Social and Financial Performance: A metaanalysis. *Organization Studies*, 24(3): 403-441.

[10] Repetto, R. (2004). *Divulgation de données environnementales importantes dans les états financiers des sociétés d'exploitation*, Rapport de la Commission de coopération environnementale, 2004.

[11] Richardson, A.J.; Walker, M. and Hutchinson, l.R. (1999) Managing capital market reactions to corporate social responsibility. *Intemational Joumal of Management &*Retailing, 1(1): 17-43.

CODES OF CONDUCT AND CORPORATE SOCIAL RESPONSIBILITY: AN EMPIRICAL STUDY

Barbara Withers, School of Business Administration, University of San Diego,
San Diego, CA 92110, (619) 260-2380, bwithers@sandiego.edu
Kristine Ehrich, School of Business Administration, University of San Diego,
San Diego, CA 92110, (619) 260-7629, ehrich@sandiego.edu
Maling Ebrahimpour, University of South Florida, St. Petersburg,
St. Petersburg, FL 33701, (727) 873-4786, bizdean@stpete.usf.edu
Anika Jensdotter, School of Business Administration, University of San Diego,
San Diego, CA 92110, anikajensdotter@yahoo.com

ABSTRACT

A study of the perceptions of supply chain management professionals toward CSR was conducted to determine whether exposure to codes of conduct and ethics training influenced their behaviors in the workplace. Survey results revealed that most companies have corporate codes of conduct in place. Having a corporate code of conduct did not have a significant impact on employees' commitment to CSR, but mandatory training on the codes did significantly impact the level of commitment towards CSR. The level of perceived commitment to CSR impacted how often the respondents witnessed various unethical behaviors within their respective companies. Higher levels of inappropriate behaviors were observed among vendors' employees when one's codes of conduct were shared with vendor employees. True commitment from top management makes a significant difference in how employees perceive and respond to codes of conduct and social responsibility.

INTRODUCTION

A shift toward corporate responsiveness began at the beginning of the twentieth century, however the phrase corporate social responsibility (CSR) did not gain common usage until the 1960's when firms were challenged to provide more than just goods and services to society[6, 29]. As social and environmental concerns grew, society turned its focus towards large corporations, claiming that with great power came great responsibility [20]. Consequently, firms were pressured not only to maximize shareholder value, but also to contribute to sustainable development and social welfare in the global community.

In 1994 Milton Friedman's concept of maximizing shareholder value was challenged by Freeman's publication on strategic management [14]. Contrary to Friedman, Freeman argued that managers held both a fiduciary and social responsibility to all stakeholders in the company, reinforcing the need for CSR adaptations in order for companies to remain competitive and

maintain a positive reputation in the marketplace [14]. Consequently, in the late 1990's, as more companies turned attention to social obligations, firms were challenged to modernize Friedman's traditional microeconomic theory of solely maximizing shareholder value [3].

Policy makers have defined CSR as "a fundamental concept whereby companies integrate social and environmental concerns into their business operation and in their interactions with their stakeholders on a voluntary basis" [12]. This definition suggests that, while CSR is an essential aspect of the business marketplace, CSR is, in essence, not obligatory. However, because investors have modernized evaluations of performance based on the company's ability to meet social responsibilities and the demand for ecological products have increased exponentially, companies must adopt CSR initiatives if they wish to stay competitive and maintain a positive reputation [32].

As CSR implementation has the capability to repair damaged reputations and increase competitive advantages in the market, the social initiatives sought after by companies have been widely criticized as a marketing tool rather than an honest aim towards improving social welfare [16]. Oil giant, Shell, developed a program to install health clinics in Nigeria in order to maintain a stable working environment in the area and to improve perceptions of the company. However, after only one-third of the projects proved functional, the audit showed that Shell was trying to buy off the locals with gifts rather than trying to better the society [16]. Some companies, such as Shell, seem to approach CSR solely from a cosmetic perspective; however CSR is, in theory, a central business issue that, if used properly, can have a substantial impact on most business operations [32].

First coined in 1982, Supply Chain Management (SCM) did not gain common usage till the mid 1990's when businesses, large and small, began outsourcing operations to developing nations [1]. As components of the supply chain became more compartmentalized, management lost oversight of day-to-day operations, which increased opportunities for unethical practices and exploited human and environmental resources [35].

According to some scholars, the supply chain has become the most important strategic tool in today's business word [30]. The supply chain holds an intricate balance of methodologies and internal responsibilities that have expanded to include externalities left on the surrounding environment. As more firms rely on outsourcing to lower production costs, ethics has become a major component in the supply chain and furthered the rise of CSR initiatives [5]. As early as 1998, strategic partnerships among suppliers and customers started to become a major industry trend. According to a study by Witt [34] an 85% of companies ranked SCM as essential to success [34].

Ethical issues have become more prominent in the marketplace because, by nature, CSR initiatives tend to promote some form of right-action, thereby attempting to improve unethical conditions [35]. As a result, ethics has become a hot topic, and business leaders are increasingly turning their attention to ways to incorporate ethical representation in business operations. Evidence of this prominent message among business leaders was revealed through a simple Google search on the words "ethics" and "business ethics" that yielded over 82 million hits.

Ethics is defined by Stahl and Grigsby [28] as doing the right thing right the first time. Therefore, as ethical scandals surface within and among the supply chain, companies are adopting new codes of conduct to improve viability in the marketplace [31]. After facing sever allegations of sweatshop labor in 1996, Nike took an operational approach towards corporate responsibility and put a strict code of conduct in place that allowed Nike to monitor all aspects of its supply chain. Nike has since taken a leadership position for worker protections in the textile industry and has plans to aid other companies to improve existing factory conditions [31].

Studies have shown that companies with strong CSR programs tend to attract a committed workforce because employees are proud to work for socially responsible organizations [4]. Top management tends to hire people with similar values that have a pivotal effect on the behavior of employees thereby reinforcing employee productivity and motivation through enhanced organizational identification [4]. Also, CSR companies with less management ability tend to manage CSR with less skill, so it has been argued that successful CSR initiatives tend to produce greater benefits when coupled with experienced managers [9]. Furthermore, companies can counteract unethical conditions through top management's commitment to a positive ethical climate found within CSR implementation [2].

In light of the trend in recent years for increasing numbers of firms to establish formal codes of ethics as standards of behavior for both employees and suppliers, we were interested in determining whether or not these codes appear to be influencing behaviors. Our research sought to determine whether the existence of codes of ethics or codes of conduct have impacted employee and/or vendor behaviors with respect to actions that would be perceived as constituting socially responsible behavior.

METHODOLOGY

Questionnaire Development

Our questionnaire was developed in a six-step process. The first step was a comprehensive review of the literature on ethics and the supply chain. Within this literature review, particular effort was made to identify articles that addressed ethics in the context of Corporate Social Responsibility (CSR). Based on the literature review, the second step was to draft the survey instrument. The third step of the questionnaire development process was an initial test for clarity by graduate students at a major US university. Using information from this test, the questionnaire was modified to improve understandability. The next step was a test for clarity, relevance, and technical accuracy and was administered to members of the Supply Chain Management Institute (SCMI) affiliated with a west coast university. The final step was a refinement of the survey instrument based on the results of the survey responses from these SCMI members.

The survey instrument contained a series of questions to capture demographic information about the respondents and their respective firms. The other questions in the survey were designed to capture information about firms' actions relevant to ethics and social responsibility. Most of these questions utilized a five-point Likert scale with five representing the strongest or most positive response and one representing the weakest or most negative response. Another set of

questions was designed to assess observed behaviors to which respondents answered yes (coded as 1) or no (coded as 2). For the purposes of this study, mean responses for both types of questions were analyzed.

Sample Selection and Characteristics

The population for the survey was comprised of supply chain professionals /members of the Institute for Supply Management ISM). A random sample of 5000 members, representing twelve major industries, was provided by ISM from its membership list. The industries selected were Agriculture; Mining; Utilities; Construction; Manufacturing; Wholesale Trade; Retail Trade; Transportation and Warehousing; Information Services; Finance and Insurance; Professional, Scientific, and Technical Services; Administrative and Support and Waste Management and Remediation Services; Educational Services; Health Care and Social Assistance; Accommodation and Food Services; Other Services (not Public Administration); and Public Administration.

Of the 5000 names and addresses provided by ISM, 63 names were eliminated because of incomplete mailing addresses. Hard copy surveys were mailed to the remaining 4,937 addressees obtained from the sample draw. Nine surveys were returned to sender, reducing the population size to 4,928. Of these, a total of 421 completed and usable survey forms were returned, for a response rate of 8.5%.

For the purposes of this study, we were interested in the beliefs, behaviors, and perceptions of respondents who are specifically associated with the procurement function and who would have direct knowledge of or interaction with vendors and customers. A subset of all respondents to the survey whose job title identified them as individuals working directly in a procurement capacity was selected for our analysis. Two hundred forty-eight respondents fell into this category and constituted the sample used for our analysis in this study. Of these respondents, 18.1% were in lower level positions of responsibility (i.e., buyers), 41.5% were in middle management positions, and 40.3% were in upper management positions.

Responding firms in the subset represent both multinational and US-based companies: 59.3% of the respondents indicated their company is a multinational company and 36.3% indicated their companies are not multinational. A small portion (3.6%) indicated they do not know whether their companies are multinational. The vast majority of the respondents' companies are headquartered in North America (87.5%); 11.3% of respondents' companies are headquartered in Western Europe. The remaining companies have headquarters scattered in Central and South America, Eastern Europe, or Asia. The vast majority of respondents' firms (90.7%) identified North America as their primary sales region; 87.5% indicated that their firms' primary sources of purchases were also from North America.

Survey respondents work in all of the industries represented in the original sample but the majority are employed in three industries: Manufacturing (42.3%), Utilities (14.9%), and Finance and Insurance (10.9%). Responding firms represent a wide range of firm sizes in total annual sales but the majority would be classified as large: 28% of the firms have annual sales up to $100 million; 20% of the firms have annual sales greater than $100 million and up to $1

billion; and 52% have annual sales greater than $1 billion. Firm size, as represented by number of employees, was similarly represented: 28.2% of the firms have 1000 or fewer employees, 36.3% have from 1001 to 10,000 employees, and 35.5% have more than 10,000 employees.

Findings and Discussion

To establish whether companies see a need to communicate corporate values to their employees and vendors, we asked whether respondents' companies have a policy regarding business ethics (e.g., a code of ethics, a policy statement on ethics, a code of conduct, or any other guidelines on ethical behavior). Ninety-five percent of the respondents indicated that their companies do have policies regarding ethical behavior. To assess the prevalence of the dissemination of these standards of employees, we asked respondents to identify the proportion of company employees to whom the code of conduct had been communicated. Eighty-one percent indicated that their company's code of conduct had been communicated to at least 76% of its employees.

Having established that the majority of employees are exposed to corporate codes of conduct, we were interested in determining whether this exposure impacted their commitment to social responsibility. Respondents were asked to indicate their perception of the level of employee commitment to social responsibility using a scale of 1 to 5, with 1 representing "not committed" and 5 representing "completely committed." ANOVA revealed that employees of companies that have shared their code of conduct with the vast majority of their employees (i.e., more than 75%) were perceived as having a stronger commitment to CSR (mean = 3.93) than companies to whom fewer than 75% of employees were exposed to the corporate code of conduct (mean = 3.67). The difference is not significant (p = .10) but the direction of the difference was as expected.

While the results mentioned above looked for the influence of making a company's code of ethics available to employees, we felt it would be more insightful to see if mandatory training of employees on the company's code strengthened the commitment to CSR. The results showed that training did strengthen employees' commitment to CSR (p < .001): for companies with mandatory training, employees were perceived to be more committed to CSR (mean = 4.02) than were employees at companies without mandatory training (mean = 3.42). Similar results (p < .001) were found with respect to top management commitment. Respondents from companies with mandatory ethics training perceived their top managers as being more committed to CSR (mean = 4.34) than respondents from companies without mandatory ethics training (mean = 3.67).

Next we were interested in whether employees' levels of commitment to CSR would impact their observation of illegal conduct or conduct violating a company's code of ethics within their company. The illegal conduct or violations of company policy that were addressed are listed in Table 1.

As indicated previously, the perception of employee commitment was assessed on a scale of 1 to 5, with 1 representing "not committed" and 5 representing "completely committed." We found that the level of perceived commitment by employees impacted how often the respondents witnessed certain unethical behaviors within their respective companies. Any level of

commitment to social responsibility is likely better than none; however, being moderately committed doesn't result in the same behavior as being completely committed. This isn't to say that total commitment to CSR is necessary to see adherence to ethical behavior within the company, but it does appear that complete commitment results in significantly more adherence than does moderate commitment for certain behaviors. For example, in comparing those who state their employees are "completely committed" to those who are "moderately committed," we find significantly less a) sexual harassment (M_{comp} =2.00 M_{mod} =1.82, $p < .04$), b) lying to employees, customers, vendors and the public (M_{comp} =1.95 M_{mod} =1.70, $p < .01$) and c) abusing or intimidating other company employees (M_{comp} =1.95 M_{mod} =1.71, $p < .02$). However, this commitment did not impact all behaviors, for example, when asked whether "discrimination on the basis of race, color, gender, or age" was witnessed, employees reported not witnessing these unethical behaviors regardless of their perception of commitment to social responsibility by employees.

TABLE 1. TYPES OF ILLEGAL CONDUCT OR CODE VIOLATIONS

Engaging in sexual harassment
Giving or accepting bribes, kickbacks, or inappropriate gifts
Falsifying records and reports
Lying to employees, customers, vendors, or the public
Withholding needed information from customers, vendors, or the public
Misreporting actual time or hours worked
Stealing, theft, or related fraud
Breaking environment and safety laws or regulations
Abusing or intimidating other company employees
Discriminating on the basis of race color, gender, age, or similar categories

Much has been written about the trend on the parts of companies to impose corporate codes of ethics on their suppliers, particularly in certain industries [4, 7, 18, 21, 25, 35, 36]. The literature reports that suppliers in the textile and clothing industries have been virtually mandated to adhere to the corporate codes of ethics of their buyers [21, 35, 36]. Unfortunately, much of the literature suggests that these suppliers are circumventing these codes of conduct in their own operations by presenting false information to customers' auditors, maintaining two sets of books, and coaching employees on their behaviors when auditors or customer representatives are in contact with suppliers' personnel [11, 18, 35, 36].

We asked respondents whether their companies communicated their corporate codes of conduct/policy with their vendors. Over 66% reported that the company's code of conduct is shared with vendors. Based on these responses, we then queried respondents about their observations of the same illegal conduct or violations of these codes (listed in Table 1) by their

vendors' employees. The results are interesting. We found that sharing one's code of conduct with one's vendors had little impact on the behavior of vendor employees. Only two categories ("misreporting actual time or hours worked" and "discriminating on the basis of race, color, gender, age, similar categories") showed significant positive relationships between the sharing of the code of conduct and the observed behavior of vendor employees. In fact, the results revealed that more inappropriate or illegal behaviors were observed when corporate codes of conduct were shared with vendors than when the codes were not shared.

It is assumed that top management commitment to CSR would set the tone for facilitating employee adherence to the company's code of conduct. To see if this is the case, we examined the perception of top management commitment to CSR (described above) with responses to a series of questions about respondents' comfort level in dealing with misconduct and unethical practices. Respondents who see top management as "completely committed" compared to those who see top management as "moderately committed" were significantly more likely to agree that "top management talks about the importance of ethics and doing the right thing in the work we do" (M_{comp} =1.05 M_{mod} =1.29, p < .007) as well as "the head of my company sets a good example of ethical business behavior" (M_{comp} =1.03 M_{mod} =1.21, p < .021). We realize that these are almost tautological, but they actually serve as a nice check on their responses (the questions were not near each other on the survey).

We also find that respondents in companies with top management who are seen as "completely committed" (as compared to those who are "moderately committed") to social responsibility are more likely to agree with a) "if I observe misconduct within my own company that could hurt my company's profits or effectiveness, I would report it to my supervisor or to the appropriate office" (M_{comp} =1.06 M_{mod} =1.25, p < .007) as well as b) "I can report unethical practices of people above my job level without fear of negative consequences" (M_{comp} =1.17 M_{mod} =1.54, p < .001). One reason for this feeling of freedom regarding the reporting of unethical practices is seen in the response to the question "If I report my ethical concerns, I will be seen as a troublemaker by top management." Those who feel top management is "completely committed" to social responsibility were more likely to disagree with that statement (M_{comp} =1.87 M_{mod} =1.63, p < .032). Interestingly, as stated above, employees in companies where top management is perceived to have "complete commitment" felt that they could report supervisors' unethical practices without fear of reprisal more than those at companies with "moderately committed" top management. However, level of commitment did not impact whether respondents felt at ease to report unethical practices of people who were below their job level (M_{comp} =1.06 M_{mod} =1.13, n.s.) indicating that regardless of an employee's perception of top management's commitment to CSR, virtually all employees feel comfortable reporting misconduct on the part of underlings.

Employees who see top management as "completely committed" tend to agree that "my company's concern for ethics and doing the right thing is an important reason that I continue to work here" much more than those who see top management as "moderately committed" (M_{comp} =1.09 M_{mod} =1.46, p < .001). Also, those who see top management as "completely committed" are more likely to say "my company is serious about doing good in the community around it" (M_{comp} =1.02 M_{mod} =1.17, p < .04).

CONCLUSION

Today social responsibility has become a necessity, however there still lacks a single consensus on how CSR should be implemented in organizations [13, 17]. Successful implementation of CSR cannot occur without integrating these efforts into SCM.

Our research has revealed that most companies have a corporate code of ethics or a policy on ethical behavior. Results indicate that sharing the company's code of ethics with employees resulted in improved ethical behavior on the parts of company employees, as reported above. In fact, mandatory training in a company's code of conduct leads to behaviors that are more consistent with a company's code/policy. Interestingly, sharing corporate codes of conduct does not appear to result in similar good behavior on the parts of the employees of vendors. In fact, worse behaviors were reported when the codes had been shared than when the codes had not been shared. One possible reason why the procurement respondents observed more bad behavior when their corporate codes of conduct had been shared with vendors could simply be that the procurement people were aware of bad behaviors and, therefore, more likely to observe/be aware of violations on the parts of their vendors' employees. Another possible explanation is that the respondents may not have had the opportunity to observe vendor employee behavior and, therefore, tended to respond "no" to the question of observing bad behavior, regardless of whether the code had or had not been shared. Alternatively, the results of the survey could also be revealing evidence suggested in the literature that vendor employees modify their behaviors in the presence of customer representatives and auditors [11, 25, 26].

In summary, we find that it is important for top management to demonstrate (by having a code of conduct, especially a written code, and sharing that code with employees) that they are committed to social responsibility. This demonstration impacts both ethical behavior of employees as well as whether or not employees feel at ease reporting misbehavior of those around them. It is also clear that many employees who work in organizations where social responsibility is demonstrated by top management, do so because they want to work for an ethical company. The results of this research indicate that true commitment from top management makes a significant difference in how employees perceive and respond to codes of conduct and social responsibility. These findings are consistent with the literature showing that successful implementation of major corporate initiatives (e.g., TQM, ISO 9000, CSR) has hinged upon top management commitment to these initiatives [7, 8, 10, 21, 27, 33].

There is still room to further investigate the role of sharing codes of conduct with vendors. It appears that the link between sharing the code and then observing ethical behavior isn't as clear-cut as it is with employees. However, as we stated above, it could be that there is just not as much opportunity for observing negative behavior (or any behavior) of vendor employees or that vendor employees are modifying their behaviors in the presence of customers. We are excited to investigate this and other possibilities in future research.

REFERENCES

[1] Ansberry, C., 2003. Outsourcing Draws Debate at Home, *Wall Street Journal- Eastern Edition*, 242 (9), A2.

[2] Badenhorst, J. A., 1994. Unethical Behaviour in Procurement: A Perspective on Causes and Solutions, *Journal of Business Ethics*, 13, 739-745.

[3] Balakrishnan, U., Duvall, T., Primeaux, P., 2003. Rewriting the Basis of Capitalism: Reflexive Modernity and Ecological Sustainability as the Foundation of a New Normative Framework, *Journal of Business Ethics*, 47, 299-314.

[4] Bammer, G., 2007. Improving the Management of Ignorance and Uncertaintly, *Handbook of Collaborative Management Research*, The Goolabri Group, 421-437.

[5] Boyd, E.D., Spekman, R. E., Kamauff, John W., Werhane, P., 2006. Corporate Social Responsibility in Global Supply Chains: A Procedural Justice Perspective, *Long Range Planning*, 40(3), 341-356.

[6] Carroll, A., 1997. A Three Dimensional Conceptual Model of Corporate Performance, *Academy of Management Review*, 4(3), 497-505.

[7] Carter, C.R. and Jennings, M.M, 2004. The Role of Purchasing in Corporate Social Responsibility: A Structural Equation Analysis, *Journal of Business Logistics*, 25 (1), 145-186.

[8] Crawford, K.M. and Cox, J.F., 1991. Addressing Manufacturing Problems through the Implementation of Just-In-Time, *Production and Inventory Management Journal*, 32 (1), 33-36.

[9] Demacarty, P., 2008. Financial Returns of Corporate Social Responsibility, and the Moral Freedom and Responsibility of Business Leaders, *Business and Society Review*, 114 (3), 393-433.

[10] Ebrahimpour, M. and Withers, B.E., 1993. A Comparision of Manugacturing Management in JIT and non-JIT Firms, *International Journal of Production Economics*, 32, 3535-364.

[11] Egels-Zandén, N., 2007. Suppliers' Compliance with MNCs' Codes of Conduct: Behind the Scenes at Chinese Toy Suppliers, *Journal of Business Ethics*, 75, 45-62.

[12] European Commission, 2002. A Business Contribution to Sustainable Development, Green Paper, Luxembourg.

[13] Frederick, W.C., 1994. From CSR1 to CSR2, *Business and Society*, 33(2), 150-164.

[14] Freeman, R.N., 1994. The Politics of Stakeholder Theory, *Business Ethics Quarterly*, 4, 409-421.

[15] French, J. L., 2010. Children's Labor Market Involvement, Household Work, and Welfare: A Brazilian Case Study, *Journal of Business Ethics*, 92, 63-78.

[16] Frynas, J.G., 2005. The False Developmental Promise of Corporate Social Responsibility:

Evidence from multinational oil companies, *International Affairs*, 81 (3), 581-594.

[17] Griffin J.J., 2000. Corporate Social Performance: Research Directions for the 21st Century, *Business and Society*, 39(4), 479-491.

[18] Jiang, B., 2009. Implementing Supplier Codes of Conduct in Global Supply Chains: Process Explanations from Theoretic and Empirical Perspectives, *Journal of Business Ethics*, 85, 77-92.

[19] Kolk, A., Van Tulder, R., 2002. Child Labor and Multinational Conduct: A Comparison of International Business and Stakeholder Codes, *Journal of Business Ethics*, 36(3), 291-306.

[20] Lotila, P., 2009. Corporate Responsiveness to Social Pressure: An Interaction-Based Model, *Journal of Business Ethics*, 94, 395-409.

[21] Mamic, I., 2005, Managing Global Supply Chain: The Sports Footwear, Apparel, and Retail Sectors, *Journal of Business Ethics*, 59, 81-100.

[22] Mares, R., 2010. The Limits of Supply Chain Responsibility: A Critical Analysis of Corporate Responsibility Instruments, *Nordic Journal of International Law*, 79(2), 193-244.

[23] Mueller, M., Gomez dos Santos, V., Securing, S., 2009. The Contribution of Environmental and Social Standards Towards Ensuring Legitimacy in Supply Chain Governance, *Journal of Business Ethics*, 89, 509-523.

[24] Ni, D., Li, K. W., Tang, X., 2010. Social Responsibility Allocation in Two-Echelon Supply Chains: Insights from Wholesale Price Contracts, *European Journal of Operational Research*, 207, 1269-1279.

[25] Parls-Poaps, H. and Rees, K., 2010, Stakeholder Forces of Socially Responsible Supply Chain Management Orientation, *Journal of Business Ethics*, 92, 305-322.

[26] Pedersen, E. R., Andersen, M., 2006. Safeguarding corporate social responsibility (CSR) in global supply chain: how codes of conduct are managed in buyer-supplier relationships, Journal of Public Affairs, 6, (3/4), 228-240.

[27] Schonberger, R.J. 1982. Japanese Manufacturing Techniques: Nine Hidden Lessons in Simplicity, Free Press, NY.

[28] Stahl, M.G. and Grigsby, D.W., 1998. *Strategic Management, Total Quality, and Global Competition*, Blackhawk Pubishing.

[29] Steiner G. and Steiner J., 1997. Business Government and Society: A Managerial Perspective, McGraw-Hills, New York.

[30] Strand, R., 2009. Corporate Responsibility in Scandinavian Supply Chain, *Journal of Business Ethics*, 85, 179-185.

[31] The ETUL, 2008. Just Doing It, *Business China*, 34 (7), 3-4.

[32] Vilanova, M., Lozano, J. M., Arenas, D., 2008. Exploring the Nature of the Relationship Between CSR and Competitiveness, *Journal of Business Ethics*, 87, 57-69.

[33] Withers, B.E., Ebrahimpour, M., and Hikmet, N., 1997. An Exploration of the Impact of TQM and JIT on ISO 9000 Registered Companies, *International Journal of Production Economics*, 53, 209-216.

[34] Witt, C.E., 1998. Trends in Supply Chain Management, *Material Handling Engineering*, January, p.14.

[35] Yu, X., 2008. Impacts of Corporate Code of Conduct on Labor Standards: A Case Study of Reebok's Athletic Footwear Supplier Factory in China, *Journal of Business Ethics*, 81, 513-529.

[36] Yu, X., 2009, From Passive Beneficiary to Active Stakeholder: Workers' Participation in CSR Movement Against Labor Abuses, *Journal of Business Ethics*, 87, 233-249.

Corporate Social Responsibility: The Case of Pharmaceutical Industry in African Countries

Morgan Germond, Ian Martins, Rachel Urato

Roger Williams University

ABSTRACT

The Pharmaceutical industry is one of the most important industries in maintaining the sustainability of human resources. This study examines the "Millennium Development Goals Report" based on extensive data compiled and led by the Department of Economic and Social Affairs of the United Nations. The research delineates how pharmaceutical companies can assist in a worldwide economic development by alleviating child mortality, contagious diseases, and other health risks that hinder such development in emerging countries.

Furthermore, this study examines the history of development of drugs and medicine in the United States as well as Europe in addition to the pharmaceutical companies' programs devised to aid the developing nations. Several programs, such as Bayer's Global Business Coalition (BCG), that are developed to target diseases like HIV and malaria in African nations are analyzed.

The final section of the study concentrates on the shortcomings of the actions by major pharmaceutical companies that hinder achieving an accepted standard of health and welfare for the African nations.

The Regulatory Capacity of a NGO-trade Union Coalition in the CSR Field

Emmanuelle Champion

Chaire de responsabilité sociale et de développement durable
Université du Québec
Montréal, Canada

ABSTRACT

In recent years, the formation of coalitions between NGOs and trade unions has developed in various areas of action, including CSR (Gallin 2001). These coalitions characteristically promote the adoption, by leading firms and public administrations, of regulatory tools designed to ensure respect for human rights and minimal environmental norms along supply chains (Palpacuer 2008, Vercher 2009). Typical among them are coalitions fighting against the proliferation of sweatshops in the developing world. This paper will discuss the conditions by which this type of coalition may acquire regulatory power in respect of CSR issues. Several analysts consider such multi-stakeholder initiatives as the foundation of a regulatory system better adapted to the globalization of economies and better positioned to newly articulate social and economic factors on a global scale (Cashore, 2003, Gendron and al., 2005, Robinson, 2007). New regulatory platforms also appear well placed to mobilize personal and institutional consumers (of product or services) by providing them with societal information or information pertaining to a variety of corporate practices (Michelleti, 2003). Few studies have yet shed light on the processes by which these regroupings develop capacities for common action. This paper will thus present a diachronic analysis of the formation of a Quebec-born NGO-trade union coalition active in the field of CSR since 2003. It will discuss the specificity of such a coalition, the challenges it faces in the development of regulatory capacities, and the obstacles it encounters in the institutionalization of regulatory tools such as codes of conduct and responsible purchasing policies. It will conclude with a discussion of how this type of regrouping could contribute to the development of a new, global system of regulation by mobilizing the works of neo-institutionalists (Scott, 1995, Campbell, 2004; Crouch, 2005) as well as Reynaud's theory of social regulation (1997).

CORPORATE SOCIAL RESPONSIBILITY: CAPITAL'S STRATEGY IN THE CONTEXT OF ECONOMIC RESTRUCTURING.

Author: Josiele Heide Azevedo – UFSC – Campus Universitário St. Trindade, Florianópolis, Santa Catarina, Brazil ZIPCODE 88040-970 - euzinhaz@hotmail.com
Co-author: Valeska Nahas Guimarães –PPGE/UFSC – Campus Universitário St. Trindade, Florianópolis, Santa Catarina, Brazil ZIPCODE 88040-970 - valeska_kenaz@yahoo.com.br

ABSTRACT

The thinking and actions of co-responsibility of employers for education, the environment, the country, have become a differentiator in its administration. As a result of Master thesis defended at Federal University of Santa Catarina, Brazil, this article discusses how to operationalize the objectives of the companies to adopt the slogan of "socially responsible" and the concept of education borne by the entrepreneurs who participated in a social education project. It is concluded that the goals of business education are the company's, focused on competitiveness and raising the educational level of the worker to ensure profits and improved corporate image.

Keywords: Work and Education, Corporate Social Responsibility, Corporate Education.

Introduction

The following article is the result of research developed at the Masters in Education held in 2008, the Federal University of Santa Catarina, Brazil [1]. The overall objective of the study aims, from the historical review of the company's option for Social Responsibility projects, search for the purpose for which companies develop educational programs for Social Responsibility.

The methodological assumption of the research was to investigate a specific research object, a vocational training project for the young with a low-income and developed by Brazilian companies, and also the analysis of dissertations on the subject.

For the theoretical analysis we used the category of "ideology" in this whole research process, from Marx understood [2] as a logical, systematic and coherent representations and norms or rules that indicate or prescribe society members what to think, value, feel and do. Thus, the category of ideology at work is presented as an indicator of a political position, to reflect on the general aspects of the real and essential, their connections and relationships, once the hegemonic ideology that preaches the study filled the fight against poverty and social injustice through corporate social responsibility in the search for a human, fairer social and environmental development.

Restructuring productive and economic restructuration of capital: Corporate Social Responsibility as business strategy

The globalization of capitalism carries with it the globalization of the world of work [3]. The work process and production appears increasingly subsumed to capital movements.

It is understood capitalism as a "civilizing" process that constantly reforms the conditions of life and work, ways of being of individuals and communities, according to its logic. To the extent that capitalism is facing crises, restructuring the relations of production and work. The elements that constitute a crisis are very complex and involve the economic, social, ideological, political dimensions and have strong repercussions in the ideal, in the subjectivity and in the values that constitute the working class.

This restructuring, or "reengineering of capital" [4], hits the surface in a phenomenal form, i.e. "restructuring without changing the basic pillars of the capitalist mode of production"[5, p.36]. Examples have been given in the organizational field, the shift of labour relations, in technology, the inclusion of so-called microelectronics-based technologies, and social, the incorporation of "ethical values" and social concerns in administrative management. For each segment there are a series of actions planned and unleashed. However, the study was delimited to observe the manifestations in the social field.

Reengineering the social segment is based on actions based on a supposed "business caution" that is engaged in the pursuit of "quality of life" of society and in the concern about its "corporate image" over consumers and the market. These actions took shape in the business fronts such as volunteering [6], the theology of the market [7], the social merchandising [8], social marketing [9] and Social Responsibility. Of these five, we focus on the latter.

Social Responsibility is entered by the business in exchange for the actions of the State, linked the responsibility of society via social policies. In the current economic context, under the aegis of neo-liberalism, the State was characterized as minimal, transferring part of their social responsibilities of public managers to society at large. This society is called, via discourses of solidarity, volunteerism and corporate social and individual responsibility, to contribute to the alleviation of social problems. However, the capitalist organization, drawing on methods used in its production process, by its own logic, distinguishes social actions that could be transformative and lasting. At its limit, appease social problems in a superficial way and limited in time.

Complements to this the fact that Social Responsibility is a tool of manipulation and social inculcation in the hands of companies, which have acted as an ideological apparatus [10]. The poor working conditions and exploitation by means of flexible accumulation are disguised by the social function of the company, using strategies such as techniques of human relations, or participation, allowing the gain relative and absolute at the same time necessary for the reproduction of capital and its over accumulation.

One of the main definitions of Corporate Social Responsibility (CSR), used by the Brazilian business community and by the Ethos Institute [11], considers it as a way of doing business that makes the company a partner and co-responsible for social development [12]. So, for a company to become socially responsible it should reconcile the different interests involved (shareholders, employees, suppliers, consumers, community, government and environmental preservation) incorporating them into the strategic planning of its activities.

As a matter of fact, a survey [13] conducted by the Institute ADVB [14] of Social Responsibility (IRES) reveals that Social Responsibility is an important strategic tool in making business decisions, for according to the same study, 97% of them use it in its management.

The corporate-citizen, for the consumer, according to research from the Institute Akatu [15] is

one that creates jobs, pays taxes and goes beyond the ethical obligations [16]. However, among the less valued actions we have actions of political transparency [17]. The practices identified by consumers delineate a paternalistic conception of CSR, philanthropic, in an emergency and with a focus on the public in need.

Thus, identifying themselves as socially responsible development support to the community where the company operates, the preservation of the environment, transparent communication, the return to shareholders; synergy with partners, customer satisfaction, the preservation of well being of employees and dependents. However, the area of education is what has gotten more attention from entrepreneurs when it comes to taking a social project.

To have a run size of investments, according to the latest census of the Group of Institutes, Foundations and Enterprises (GIFE) [18], only its members spent about $555.5 million dollars, equivalent to 20% the investments of the domestic private sector for social, cultural and environmental projects taken in a planned, systematic and monitored way [19]. The prioritized area, education, received private investment of about US$ 68.8 million in 2005 alone.

The education conducted by the company: the social educational programs for Corporate Social Responsibility and its implications

According to a survey released by GIFE [20], social investments made by companies in education projects alone represent 85%, which shows the area of education as the main focus. With globalization, the business community contributes to the strengthening of a pseudo-minimal State to assume social policies and to adopt Education for them.

This concern demonstrates the recognition of entrepreneurs for the qualification of people and is a consequence of the economic demands stimulated by the process of modernization and flexibility. However, in this movement, the business community blames the Brazilian educational system responsible for not meeting the requirements of the market, and gave several reasons for that "failure", among them, the lack of infrastructure and poor use of resources. Supposedly, this situation would prevent companies from producing better and being more competitive. Thus, entrepreneurs believe that their participation in educational policies can contribute decisively to raising the educational level of the population. The suggestions range from assistance through partnerships with schools to the establishment of schools or school systems themselves.

The new points in this process are a concern and the discourse of reform coming from bases organized by business class. It's understood that the educational relationship should be governed by a mercantile type relationship and that should mimic the market model.

Nevertheless, recommendations for improving education are just mechanisms for quality control by society and by government agencies, a system of collection and transfer of funds. That is, the implementation of criteria for improving systems of assessment and management of resources.

This speech demonstrates understanding of education as a productive investment in view of an individual income. By way of economic organizations and international finance companies, this conception is today, the ideological foundation of the new order world

education. This thought has membership of international organizations and Western governments, because it proposes a strategy of "sustainable growth" and represents an economic justification for educational expenses.

However, the corporate education is used as a means of achieving competitive advantage as well as the possibility of entering new markets. Outstandingly successful companies have brought the school into their companies. In a research carried out in a company of Santa Catarina, the food branch Sadia S/A, Cruz [21] points out that the school within the company justifies its philosophy and contributes to the shaping of workers to the needs demanded by structural changes. Moreover, companies that sponsor training programs have a competitive edge with regard to the level of training of their employees, suppliers, customers and even the communities where they operate.

Learning how to learn, communication and collaboration, creative thinking, technological expertise and global business, leadership development and career self-management are characteristics of a learning organization [22].

Educational processes are acquired by the enterprise initiative, serving private and localized interests [23]. In this movement there is a historic reversal of what is meant by education. The education model used by companies not even remotely resembles a vision of education. Training is a process quite different from instrumental and/or behavioural orders practiced within organizations. Private utilitarianism governed by the logic of profit replaces, therefore, the principles of equal opportunity, democracy and solidarity that are structured around public schools, sole unit and polytechnic from the standpoint of human emancipation. We seek to commodity education as product and consider it nonsense of school spaces, or according to Drucker [24] education can not be the monopoly of schools. It's in this sense, that the companies now organize education in and out of its physical facilities.

Piccoli's studies [25] show that the benefits with education in the company range from increased productivity to employee motivation, a decrease of 50% of its turnover. Therefore, entrepreneurs "take" education because they have a glimpse that this adoption enables competitiveness, agility and high level of adherence or appropriateness of educational programs to the demands of the productive sector.

Corporate Social Responsibility: A consortium of value to the product and the image of the "company-citizen":

The educational ideals of the company are promulgated by the adoption of a socially responsible attitude. Regardless of the CSR actions undertaken, all companies seek to receive honors from the supposedly benevolent attitude towards society.

The responsible attitude of the company is seen as a sign of corporate reputation and brand [26], and is widely reported as a strategy to add significant value to the brand as a corporate-citizen. Furthermore, CSR is now seen as a business opportunity for companies, when they take a stance considered "citizen", to receive various benefits, as evidenced Rampinelli [27]:
- in taxation, with the possibility of tax exemptions at the municipal, state and federal sponsors or directly to the projects;
- for sale and image: the strengthening and loyalty to the brand and product;
- shareholders and investors: the valuation of the company in society and the in market;

- the advertising revenue: mainly originated from the spontaneous generation of media;
- for productivity and people: the greatest commitment and motivation of employees.

Under the exemptions, according to data [28] of the Internal Revenue Service (SRF), only in 2007, the Brazilian companies that invest in Corporate Social Responsibility had incentives and tax exemptions of more than US$ 611.1 million. According to information from GIFE [29], an organization that brings together the 106 largest private social investors in Brazil with about 70% of their actions aimed at education, there was a prediction that, out of the 871,000 Brazilian firms, only 178,000 were in the regime real profit, that is, they had to pay income tax.

According to research conducted by IPEA [30] in 2006, the Itaú Social Foundation, which has assets of $205,5 million dollars, invested $18,3 million dollars. In 2007, according to Ana Beatriz Patricio, superintendent of the foundation, the forecast was investments of $22,2 million dollars with a focus on education.

Figures on tax exemptions and benefits received arising from social activities are not widely disseminated. However, the amounts invested are emphasized, as it is of interest to companies and their linkage to its responsibility for society. The companies justify the high investment in social marketing as a way to change social behavior for the sake adopted. However, a study published in 2004 by IPEA [31] with the largest companies in the southeastern Brazil, showed that 65% of multinationals admit bluntly perform social actions to improve corporate image, while only 37% of companies national admit it, too.

Thus, the company considered socially responsible, has its corporate image linked to ethics and morality, contrary to an institution that cares for the well as an end in itself, trying to mask their ultimate purpose is to generate profit. Rampinelli is ironic about this situation by saying that "by magic, all organizations are pledging their commitment to pressing community, the consumer is no longer disregarded and all businesses now are 'best to work ". He adds:

"The true corporate-citizen is full all hours of the day. One can even accept that citizenship, at bottom, is an ideal type, as defined by Max Weber (1993). Something you would like to have, but never reaches it. But well, companies that have in their DNA (with its own production process, the products it manufactures and sells) a gene that leads to destruction of the environment and the citizens (of their health, for example), or service provider companies which act in flagrant disregard for the rights of their consumers, could never be recognized as Corporate-Citizens by financing some cultural projects, sports activities or building kindergartens" [32, p.51].

How to determine if the company is ethical, if it is taking attitudes and values of our society? According to Melo Neto and Froes [33] a socially responsible company has a definite pattern of social, economic, cultural and political behaviour. For a socially responsible action it is necessary to ask themselves: what are the values of our society? What are the actions resulting from these values? These settings are not highlighted or discussed by the companies. Without pointing them out, the company can not claim that their attitude or decision meets certain "purposes and values" of society and thus, one can not conclude whether or not the company is socially responsible. And yet, one must question whether the company which complies with legislated obligations would not only be honoring its duty.

So there is subjectivity present in the definitions and limits of ethics for CSR. However, the "conciliation" with the company's ethics has become a survival strategy, and its legitimacy must be built linking the mark of a good image through social marketing. In Cattani evaluation [34], the company covered with Social Responsibility is not a pure economic agent, not just sell products, but manages its relationship with the public: its logo is its greatest value.

The Corporate Social Responsibility is one of the factors added to the fetish of the product, which has been successful in winning and retaining customers and the resulting profit for shareholders. The company takes up the ethical and religious discourse, but its goal is to keep as company. Ethical and social commitments are advertisements of the company, are presented as a virtue, but have nothing to do with it. They are reasons to stay and accumulate capital. As long as the attitude of being ethical favours their goals and not hinders them, it will be an excellent deal for it to invest.

This new posture besides authorizing the company to seek profit, it allows the company to add significant value to its product. Solidarity concerned has no moral value [35], but it has commercial value. The assignment of value to products is only possible by marketing. Besides the social value, there are other aggregators of value to products.

Thus, this is not just a matter of associating value to the product, but to infiltrate cultural ideas and iconography, so that their brands can reflect ideas and images in the culture as "extensions" of their brands [36]. Social issues related to the product have shown a growing concern for consumers in purchasing decisions. In 2000, a survey of Valor Newspaper, in partnership with the Ethos Institute for Business and Social Responsibility [37] pointed out that 24% of Brazilians already consume taking into account the positioning of the company, compared to the percentage of 46% of consumers Americans.

Actions so-called socially responsible coupled with good institutional marketing give the company credibility, honesty and trustworthiness. It is important to clarify that social responsibility can not be seen as marketing. Often actions are misunderstood; merely social isolated actions and Social Responsibility, but they are actually social marketing strategies [38]. Regardless of the marketing developed, what they aim is to spread the brand associated with the awareness of people and a sharing of principles and ideals, loyalty to their customers in order to obtain best results.

Rampinelli discusses two major criticisms of this movement: "The first is that companies, just looking to cause more public appeal and more attuned to its "calling" to social investment, no longer invest its resources in the causes considered "difficult", as juvenile offenders or children with AIDS. It sounds somewhat naive the idea that companies should invest in all social causes. The second relates to a feeling that companies would invest more resources in advertising its association with the cause than the cause itself, decreasing, for lack of motivation, the legitimacy of the investment" [39, p.69].

An article, published in the Valor Econômico Journal [40], raises this controversy citing a company that spent US$ 55,500 in a social project and, anxious to strengthen its image, has invested $ 500 thousand dollars to publicize the program media [41]. The social activities of businesses are so profitable that the stock market included the environmental and social variables in financial analysis. Such data show that marketers believe in the power of building a brand based not only on the rational and emotional, but also and primarily in ethical or

spiritual dimension. Therefore, the importance of not only promoting the product but selling an idea, need or image. This is the true cause embraced.

The idea that the relationship between profit and social action passes through the corporate responsibility, is that the tensions between production and consumption are now overcome in the due to a change of attitude motivated and fueled by a 'spiritual' conversion, for a reorientation of principles. Azambuja [42] describes the situation as follows:

"The employee, worker, servant, employee, producer of surplus value is transformed, magically, into the "collaborator", i.e., someone who loves what he does, not denying ever, always sees fairness in their dealings with employers, in which there is complete equilibrium, full understanding, mutual support and unwavering sincerity (...) And he does it within a "socially responsible" system; full of relationships in which only good is built, because it is "managed" by superior people, only producing good products, good services and good ideas. And more: a dynamic system, competent, harmonious and "proactive" and it is, therefore, entitled to urge its players - regardless of their social positions - to fraternize in a common and uninterested task in the evolution of mankind. So then ... bingo! Gone are the conflict between capital vs. labour; eternal peace reigns between production and consumption. Even the most partisan analysts with pro-administrative trends, opposed, for example, to the philosophy of Foucault, would see a hint of hypocrisy in the discourse of "social responsibility ", not by the need to be responsible, but by the inappropriateness of believing social".

From this perspective, the spread "corporate-citizen" removes the social contradictions of the scene and forces a synergy between the social and profit through CSR. The relations between workers and bosses are interwoven contradictions and resistances, from those which prevailed in the times of the Industrial Revolution, had a clash capital and labour, clearly and explicitly. However, the situation we find ourselves in today, which reflects a slow and painful process – no linear way - put labour and capital in such a situation of conflict that reached a certain point, threatening the survival of the system itself. Thus, in business discourse, ethics and company become no longer elements seemingly opposed, philanthropic and educational actions are part of the purpose of the enterprise, and thus social responsibility presents itself as one of the strategies that capitalist firms assume in the continuous and necessary process to forge cooperation between labour and capital.

Final Considerations

The Social Responsibility practiced by the Brazilian business community is highly contradictory, since it is within the objective conditions of capitalism and in a context in which it does not have, in fact, public policies, once they are poor and insufficient. It appears that the social actions carried out by companies are used for profit and competitive advantage, providing for an institutional marketing rather advantageous.

It appears that CSR is a business strategy for the ability to make improvements in the "image, performance and sustainability of the company" [43, p. 81]. Thus, the end-entity is benefited, consumers feel good to consume in an ethical company and that company employees now have more pride in belonging to its staff. In summary, there is an apparent view that all stand to gain. The recipe of success achieved by the "socially responsible" company aims at getting some benefit from the promotion of their image and reputation, to the extent that their customers are sensitized to their social actions disclosed.

There is no consensus as to the understanding and practice of CSR. However, the Brazilian social projects undertaken are aimed at "managing poverty. " In the sense that they do not aim to solve social problems, but stop them.

Corporate Social Responsibility is shown as a neutral speech, natural, free and disinterested by the companies. The social activities offered strengthen the field of volunteerism, solidarity and institutions of the Third Sector, giving repercussions in the sphere of citizenship, since they allow companies to incorporate the demands of social reproduction.

The business took on CSR as a practice intrinsic to their concerns, however, it has neglected to correct the imperfections of the State, but provide them, as alleged self-regulator. CSR has facilitated the withdrawal of the State's role, and is not only a consequence of the diminishing role of this to the social sphere. In addition, CSR downgraded the State, dismantling the public policies and transforming the rights gained historically into mere favours that, thus, prevent the development of citizenship.

Presenting the inefficiency of the State and the efficiency of the companies in solving social problems is a way to promote and enhance the decrease of the power of intervention of the State and to seek an affirmation of social power. So, a power struggle for social regulation is at stake.

Furthermore, the adhesion of enterprises to social projects in education is within the scope of social obligations to the rights and duties of the State, which has increasingly been out of public policy.

According to M. T. J. Silva [44], under the banner of Social Responsibility, companies have been broadcasting across the country, thing which is called the need to take responsibility for education as a gesture of assistance with the formation of present and future society. In this case, the educational dimension was reduced to the mere condition of moral formation. The education conducted by business legitimizes the company's philosophy and capitalist principles, contributes to the shaping of workers to the needs demanded by structural changes. Add to it the fact that they are to be granted strategies for ensuring a competitive edge.

In this perspective, CSR is part of a broader process that is part of the restructuring of capital strategy, from the crisis originated in the 1970s and serves as a mechanism for the maintenance of ideological capitalism. And, as an example of what used to happen to the sponsorship, it is one more instrument of domination and maintenance of the capitalist system that pits the concept of citizenship, making individuals dependent on charity and welfare, relegated to a forged citizenship.

When the companies get coated with socially responsible roles or 'corporate-citizens', "withdraw from the political and public arena the distributive conflicts and collective demand for citizenship and equality" [45, p.390] and convert them into a voluntary altruistic disposition on behalf of a business virtuosity when, in fact, they aim to achieve cronyism.

REFERENCES

[1] Azevedo, Josiele Heide. *Responsabilidade social empresarial e educação: estudo de caso do Projeto Pescar.* Florianópolis, Santa Catarina, Brazil, 2008. 201 f. Dissertação (Mestrado) - Universidade Federal de Santa Catarina, Centro de Ciências da Educação. Programa de Pós-

Graduação em Educação. Advisor: Valeska Nahas Guimarães e co-advisor: Lucídio Bianchetti.

[2] Marx, K. *O Capital: crítica da economia política.* Livro 1 - O processo de produção do capital. 22 ed. São Paulo: Civilização Brasileira, 2004.

[3] Ianni, O. *Sociedade global.* Rio de Janeiro: Civilização Brasileira, 1992.

[4] We refer to a movement initiated by companies in the 70's and 80's, which had its heyday in the 90's of the twentieth century, aiming to streamline its organization and operation. This movement, referred to also by Trevisan and Lameira and Antunes led to flattening the organizational structure of enterprises, the large reduction in staff, the flexibility of production and labour, outsourcing of labour, the introduction of technological and organizational innovations, among others. Sources: Trevisan, N. V. e Lameira, L. ,J. C. R. Formação do educador para pedagogia nas empresas. In: *Cadernos de Educação Especial*. v.1.n.19. Santa Maria, RS: UFSM, 2002. Antunes, R. *Adeus ao trabalho.* Ensaios sobre as metamorfoses e a centralidade do mundo do trabalho. 10 ed. São Paulo: Cortez, 2005a.

[5] Antunes, R. *Adeus ao trabalho.* Ensaios sobre as metamorfoses e a centralidade do mundo do trabalho. 10 ed. São Paulo: Cortez, 2005b.

[6] Voluntary work, according to Law No. 9608/98 (Brazil), "all unpaid activity, provided by an individual to any public entity or nonprofit private company, which has civic, cultural, educational, scientific, recreational goals or social assistance, including mutuality". Source: Brazil. Law No. 9608 of volunteering, 18 Feb. 1998.

[7] The theologizing translates into interest in occupational health for the profit of the company. Thus, investments in workers' health are planned in order to obtain profits in the final process.

[8] D. Marluce Silva, general director of Globo TV Network, ranks social merchandising as a form of planned advertising insertion, systematic and with a purposeful set of social and educational messages on television. Source: Silva, M. D. da. *Política para o Merchandising Social - Sistematização de Projeto. Rede Globo* (1997). Available at: http://observatorio.ultimosegundo.ig.com.br/artigos. Accessed: 10 de jun. de 2005.

[9] According to the marketing experts Armstrong and Kotler, social marketing is a principle by which the company believes in its advertising, consumer interests, business and society in the long term. Source: Armstrong G.; Kotler, P. *Princípios de Marketing*. São Paulo: Prentice Hall, 2004.

[10] Concept understood from Althusser, who divides the ideological apparatus into Ideological Apparatus of the State and Apparatus of the State. The first acts in a repressive manner towards society in order to maintain social order and the second focuses on the prevailing ideology, although they also act secondarily by repression, even if this crackdown is disguised or symbolic. The State is, thus, the expression, on the ground of the superstructures, of a particular ideological form of a social organization of production. Source: Althusser, L. *Ideologia e aparelhos ideológicos do Estado*. Lisboa, Portugal: Editorial Presença, 1980.

[11] The Ethos Institute for Business and Social Responsibility is a nonprofit organization, characterized as Civil Society Organization of Public Interest, whose mission is to mobilize, encourage and help companies manage their business in a socially responsible way, "making it partners in building a just and sustainable society". Source: http://www1.ethos.org.br. Accessed: 20 Sept 2006.

[12] Instituto ETHOS. Available at: http://www.ethos.org.br. Accessed: 29 de ago. de 2005.

[13] Associação dos Dirigentes de Vendas e Marketing do Brasil (ADVB). V Pesquisa Nacional sobre Responsabilidade Social nas Empresas (2004). Available at: http://www.advbfbm.org.br . Accessed: 20 de set. de 2006.

[14] Association of Sales and Marketing of Brazil (ADVB) is a nonprofit organization focused on improving in the promotion, dissemination and appreciation of marketing and sales techniques. Source: ADVB (Associação dos Dirigentes de Vendas e Marketing do Brasil). *V Pesquisa Nacional sobre Responsabilidade Social nas Empresas* (2004). Available at: http://www.advbfbm.org.br . Accessed: 20 de set. de 2006.

[14] The Akatu Institute is a non-governmental, non-profit organization established on March 15 (World Consumer Day) 2001. It aims to educate and mobilize the society for consumption awareness in building of the sustainability of life on the planet. The Akatu works on several fronts: development of activities in communities and the dissemination of concepts and information on the Internet, in publications, the news media and advertising campaigns, development of researches, and development of assessment tools and information about conscious consumption.

[15] Instituto AKATU. *Pesquisa 20005: Responsabilidade Social das Empresas, Percepção do consumidor.* Available at: http://www.akatu.org.br/media/Publicacoes/PESQUISA2005.pdf. Accessed: 29 de jan. de 2007.

[16] Akatu; Ethos. *Pesquisa n. 6 responsabilidade social empresarial: o que o consumidor consciente espera das empresas*. Available at: www.akatu.org.br. Accessed: 19 de set. de 2005.

[17] Grupo de Institutos, Fundações e Empresas (GIFE). *Censo GIFE* (2005). Disponível em http://www.gife.org.br/associados.php. Accessed: 10 de jan. de 2006.

[18] Grupo de Institutos, Fundações e Empresas (GIFE). *Censo GIFE* (2005). Disponível em http://www.gife.org.br/associados.php. Accessed: 10 de jan. de 2006.

[19] Vieira, Maria C. Iniciativa privada e espírito público (2007). *Revista Valor Econômico – SP*. Available at: http://desafios.ipea.gov.br/003/00301009.jsp?ttCD_CHAVE=2267. Accessed: abr. de 2008.

[20] Cruz, D. *A formação do 'total trabalhador Sadia': um estudo de caso sobre o processo de qualificação dos trabalhadores da Sadia S.A.* Florianópolis, 2000. 150 f. Dissertação (Mestrado) - Universidade Federal de Santa Catarina, Centro de Ciências da Saúde.

[21] Senge, P. *A quinta disciplina*: caderno de campo. Estratégias para construir uma organização que aprende. São Paulo: Qualitimark, 2000.

[22] Gentili, P. A. A. e SILVA, T. T. da (Orgs.). *Neoliberalismo, qualidade total e educação*. Visões críticas. Petrópolis, RJ: Vozes, 1994.

[23] Drucker, P. *Sociedade Pós-Capitalista*. São Paulo: Thomson Learning, 1993.

[24] Piccoli, G. S. *A educação corporativa em uma empresa de serviços estudo de caso: impacto de uma universidade corporativa em uma central de atendimento a clientes - call center*. Florianópolis, 2005. 150 f. Dissertação (Mestrado) - Universidade Federal de Santa Catarina, Centro Tecnológico. Programa de Pós-Graduação em Engenharia de Produção.

[25] Rampinelli, G. de C B. *La Responsabilidad Social Como Herramienta De La Mercadotecnia Institucional: La Realidad Brasileña*. 2006. 200 f. Dissertação (Mestrado) - Universidad Nacional Autónoma de México, México, Programa de Posgrado en Ciencias de la Administración

[26] Rampinelli, G. de C B. *La Responsabilidad Social Como Herramienta De La Mercadotecnia Institucional: La Realidad Brasileña*. 2006. 200 f. Dissertação (Mestrado) - Universidad Nacional Autónoma de México, México, Programa de Posgrado en Ciencias de la Administración

[27] Vieira, Maria C. *Iniciativa privada e espírito público* (2007). Revista Valor Econômico – SP. Available at: http://desafios.ipea.gov.br/003/00301009.jsp?ttCD_CHAVE=2267. Accessed: 12 de abr. de 2008.

[28] Vieira, Maria C. *Iniciativa privada e espírito público* (2007). Revista Valor Econômico – SP. Available at: http://desafios.ipea.gov.br/003/00301009.jsp?ttCD_CHAVE=2267. Accessed: 12 de abr. de 2008.

[29] Instituto de Pesquisa Econômica Aplicada (IPEA). *Pesquisa: A iniciativa privada e o espírito público: a evolução da ação social das empresas privadas no Brasil* (2006). Available at: www.ipea.gov.br. Accessed: 12 de abr. de 2008.

[30] Instituto de Pesquisa Econômica Aplicada (IPEA). *V Pesquisa Nacional sobre Responsabilidade Social nas Empresas* (2004). Available at: http://www.ipea.gov.br/asocial. Accessed: 20 de ago. de 2005.

[31] Rampinelli, G. de C B. *La Responsabilidad Social Como Herramienta De La Mercadotecnia Institucional: La Realidad Brasileña*. 2006. 200 f. Dissertação (Mestrado) - Universidad Nacional Autónoma de México, México, Programa de Posgrado en Ciencias de la Administración.

[32] Melo Neto, F. P. de & Froes, C. *Responsabilidade social e cidadania empresarial: a administração do terceiro setor*. Qualitymark: Rio de Janeiro, 1999.

[33] Cattani, A. D. (Org.) *A outra Economia*. Porto Alegre, RS: Veraz, 2003.

[34] Kant**,** I. *Critica da razão pura*. 5. ed. Lisboa: Fundação C. Gulbenkian, 2001.

[35] Klein, N. *Sem logo: a tirania das marcas em um planeta vendido*. Rio de Janeiro: Record, 2006.

[36] Meira, P. R. *Em que posso ajudar? A Responsabilidade Social no Varejo Brasileiro*. Socialtec. Available at: http://www.socialtec.org.br/artigos/temas.htm#responsabilidade. Accessed: 31 de ago. de 2005.

[37] Social marketing is the use of principles and marketing techniques to promote a cause, idea and/or behavior. For Kotler, social marketing is a major tool for promoting changes in behavior, attitudes and practices. But cause-related marketing, according to Rampinelli, is described as "a commercial activity, based on a collaborative relationship between business, non-profit organization and/or cause, in order to promote an image, a product or a service for mutual benefit. "Sources: Kotler, P. Marketing from A to Z: 80 concepts that every professional needs to know. Rio de janeiro: Campus, 2003. Rampinelli, G. C B. La Responsabilidad Social de la Mercadotecnia as Herramientas Institutional: La Realidad brasileña. 2006. 200 f. Thesis (MA) - Universidad Nacional Autonoma de Mexico, Mexico, Program POSTGRADUATE en Ciencias de la Administración. p.68.

[38] Rampinelli, G. de C B. *La Responsabilidad Social Como Herramienta De La Mercadotecnia Institucional: La Realidad Brasileña*. 2006. 200 f. Dissertação (Mestrado) - Universidad Nacional Autónoma de México, México, Programa de Posgrado en Ciencias de la Administración.

[39] Vieira, Maria C. Iniciativa privada e espírito público (2007*). Revista Valor Econômico – SP*. Available at: http://desafios.ipea.gov.br/003/00301009.jsp?ttCD_CHAVE=2267. Accessed: 12 de abr. de 2008.

[40] Araújo, E. T. de. *Estão "Assassinando" o Marketing Social? Uma Reflexão Sobre a Aplicabilidade deste Conceito no Brasil* (2001). Available at: www.socialtec.org.br. Acesso em: 23 de mai. de 2006.

[41] Azambuja, G. *A responsabilidade social: uma "filantropia" que a ajuda a vender.* (2005) Available at: http://www.comunicacaoempresarial.com.br/rev_artigos2GermanoAzambuja.htm. Accessed: 03 de set. 2007.

[42] Pfeifer, M. *Responsabilidade social das empresas: instrumento de consolidação hegemônica do projeto neoliberal?* Florianópolis, 2006. 176 f. Dissertação (Mestrado) - Universidade Federal de Santa Catarina, Centro Sócio-Econômico. Programa de Pós-Graduação em Serviço Social.

[43] Silva, M. T. J. *Propostas de reformulação da educação profissional do SENAI e SENAC para o século XXI.* Florianópolis, 2005. 182 f. Dissertação (Mestrado) - Universidade Federal de Santa Catarina, Centro de Ciências da Educação. Programa de Pós-graduação em Educação.

[45] Paoli, M. C. Empresas e responsabilidade social: os enredamentos da cidadania no Brasil. In: Santos, B. de S. (Org.). *Democratizar a democracia. Os caminhos da democracia participativa.* Rio de Janeiro: Civilização Brasileira, 2002. p.375-418.

A SECTOR-BASED APPROACH OF ORGANIZATIONAL CHANGE IN THE DYNAMIC OF CSR INSTITUTIONALIZATION: AN OVERVIEW OF THE LITERATURE

Urbain Kiswend-Sida Yaméogo
Ph.D student at UQAM, Ecole des Sciences de la Gestion
Chaire de responsabilité sociale et de développement durable
Pavillon des sciences de la gestion, local R-2910
315, Sainte-Catherine Est, Montréal, Québec, Canada, H2X 3X2
Phone UQAM : 514 987 3000 #2108
And at
Université Paris-Est, Institut de recherche en Gestion
Gustave Eiffel - Université Paris Est Créteil Val de Marne (UPEC)
Place de la Porte des Champs - 4, route de Choisy - 94010 Créteil, France
Email: yurbain@yahoo.fr

The literature on change and organizations has given for a long time a great emphasis on intra-organizational change (Selznick, 1949, 1957) studied from several perspectives. The new institutional approach (DiMaggio & Powell, 1991; DiMaggio, 1991 ; Fligstein, 1991, 2001) and the recent developments on "institutional work"(Lawrence & Suddaby, 2006; Zietsma & McKnight, 2009; Marti & Mair, 2009), have shown the most an interest for studying organizational fields, pertinently considered as a unit of analysis and as the locus of institutionalization and change. The sector-based change, resulting from the interaction between the actors of the organizational field and the contribution of these actors, as institutional entrepreneurs, to the field structuring give a new overview on organizational change particularly in the context of CSR emergence and institutionalization. The considered change occurs at the sectoral or societal level, and is consequently interorganizational in locus as DiMaggio and Powel (1991) argued. The development of CSR within some sectors (mining, clothing or forestry sectors, etc.) usually under scrutiny, criticism and pressure of social movement actors, has been investigated by researchers. They question the way sector-based initiatives contribute to inter-organizational sensemaking of CSR, the modelling of the commitment of the field actors and the sharing of their "best practices" and knowledge and finally the building of institutional arrangements and new reference of legitimacy. Therefore, the contribution of the literature on the inter-organizational dynamic of change induced by CSR institutionalization in organizational field has to be considered with the greatest attention. The purpose of our contribution is to make a literature review on the orientation of the last years researches on this issue. We show first the value and growing interest for sectoral or societal approach to institutional change. And second we explore the mode of CSR institutionalization and its implications in terms of change within the organizational field, particularly in the mining sector. At this end we refere to new institutional, sensemaking and social movement approaches of organizational change. The combination of these approaches gives a great potential of understanding CSR institutionalization and change within sectors and allow a multilevel and dynamic approach (Schultz & Wehmeier, 2010).

References

DiMaggio P. J. & Powell W. W. (1991). The new institutionalism in organizational analysis, The university of Chicago Press,

Fligstein N. (1991). Structural transformation of American industry: an institutional account of the cause of diversification in the largest firms, 1919-1979. in DiMaggio P. J. & Powell W. W., *The new institutionalism in organizational analysis*, The university of Chicago Press, pp. 311 –336

Lawrence T. B. & Suddaby R. (2006). Institutions and institutional work, in Clegg S., Hardy C., Lawrence T. B. & Nord W. R. (Eds.), *The SAGE handbook of organization studies* London: Sage Publications (2nd ed., pp. 215-254).

Marti & Mair, 2009, Entrepreneurship in and around institutional voids: A case study from Bangladesh, Journal of Business Venturing, 24, (5), 419-435

Schultz F. & Wehmeier S. (2010). Institutionalization of corporate social responsibility within corporate communications: Combining institutional, sensemaking and communication perspectives, *Corporate Communications: An International Journal*, 15 (1), 9-29

Selznick P. (1949) The TVA and the grass roots? Harper and Row (New York).

Selznick P. (1957). Leadership in administration, a sociological interpretation, Row, Peterson and company, New York.

Zietsma, C. & McKnight, B. (2009). Building the Iron Cage: Co-creation work in the context of competing protoinstitutions. In Lawrence, T.B., Suddaby, R. & Leca, B. (Eds.). *Institutional Work: Actors and Agency in Institutional Studies of Organizations*, Cambridge University Press, 143-177.

Zietsma, C. & Lawrence, T.B. (2010). Institutional Work in the Transformation of an Organizational Field: The Interplay of Boundary Work and Practice Work. *Administrative Science Quarterly*. 55. 189-221

REDEFINING MOBILE COMPUTING: THE IPHONE AS A MOBILECOMPUTING DEVICE *CIRCA* 2008

MARK JOHN SOMERS
NEW JERSEY INSTITUTE OF TECHNOLOGY AND
RUTGERS BUSINESS SCHOOL

ABSTRACT

This case explores the introduction of the iPhone as a disruptive technology that changed the market for smart phones. The case presents students with an analysis of the major competitors in the smart phone market and uses the concept of a business ecosystem to frame the issues that must be managed to be successful. This framework is useful in analyzing what happened during the last two years and for making predictions about who will prevail in a fragmented and unstable market.

INTRODUCTION

Upon his return to Apple a decade ago, Steve Jobs was asked about the possibility of Apple introducing a branded PDA to compete with Palm's very popular offerings. At that time, he replied that that was not what Apple was about and reiterated his commitment to the Macintosh and to modernizing its then troubled operating system. As cell phones grew in popularity worldwide, Jobs was asked if Apple had any ambitions in that area, and he reiterated the view that PDA's, cell phones and related products were best left to other companies.

At that time, however, cell phones were primarily communication devices that were best viewed as mobile phones. Over time, cell phones evolved to become progressively smaller, lighter, and less expensive, and in the process, became "smarter" as well. Increasingly sophisticated operating systems enabled cellular phones to perform the functions of PDA's and then superseded them with the addition of data services including text messages, e-mail and very basic web services. Faster processors and the addition of JAVA led to the addition of more sophisticated functions including games.

Early smart phones represented a considerable advance over the basic communication devices that they replaced, but they were still relatively limited in both usability and function; put simply, they weren't all that smart. Limitations included very small screen sizes, difficult and inefficient user interfaces, uncomfortable keyboards, and very slow data transmission. When all was said and done, most people used them primarily as enhanced phones that could send and receive text messages, and track basic contact and calendar information.

Ironically, Steve Jobs was true to his word in that the iPhone is not a cell phone. Rather, the device is viewed most accurately as a mobile computer. Although Apple executives would almost certainly bristle at this description, the iPhone can be thought of a Apple's smallest

Macintosh because it runs an almost complete version of the Macintosh OS (termed Mac OS X Mobile here) with all that implies (application support, sophisticated network connectivity, etc.).

THE SMART PHONE MARKET

Projected Growth

The market for cell phones is mature and saturated with growth expected to slow to no more than 3 percent per year. The low end of the market is subject to intense price competition and comparatively low barriers to entry resulting in a low margin, increasingly commoditized business.

In contrast, the market for smart phones is expected to grow at by at least 33 percent per year over the next five years [1]. Based on 2008 sales of approximately 128 million units [2], the smart phone market should reach at least 500 million units by 2013 as indicated in Figure 1. Growth is fastest in North America, followed by Europe and then Asia.
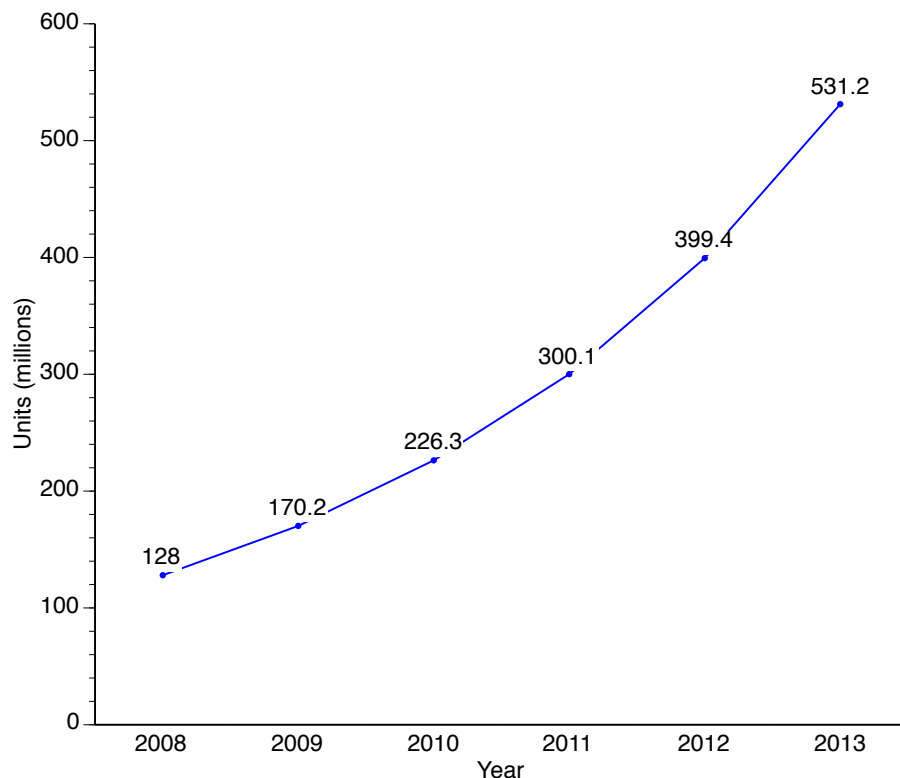
Figure 1. Projected Growth of Smart Phones Sales: 2009-2013

Unlike the market for cellular handsets which is dominated by Nokia with a 40 percent share (which has been attained with less than 3 percent of the US market), the market for smart phones remains in flux with the three largest players holding approximately equal market share in the third quarter of 2008. During this period, about 40 million smart phones were sold worldwide with Nokia holding an 18.9% share, Apple a 17% share and Research in Motion holding a 15.3 percent share of the market. Other major market participants include Palm, Microsoft, and Google [3].

TRENDS: CONSUMER MARKET FOR SMART PHONES

The Smart Phone as a Life Management & Entertainment Device

A recent study reported by the Canadian Marketing Association conducted by ARCUS identified several key trends in the smart phone market for consumers. Specifically a survey of over 1,200 individuals indicated that 95% of wireless phone users said they use at least three of the technological innovations introduced in the past five years.
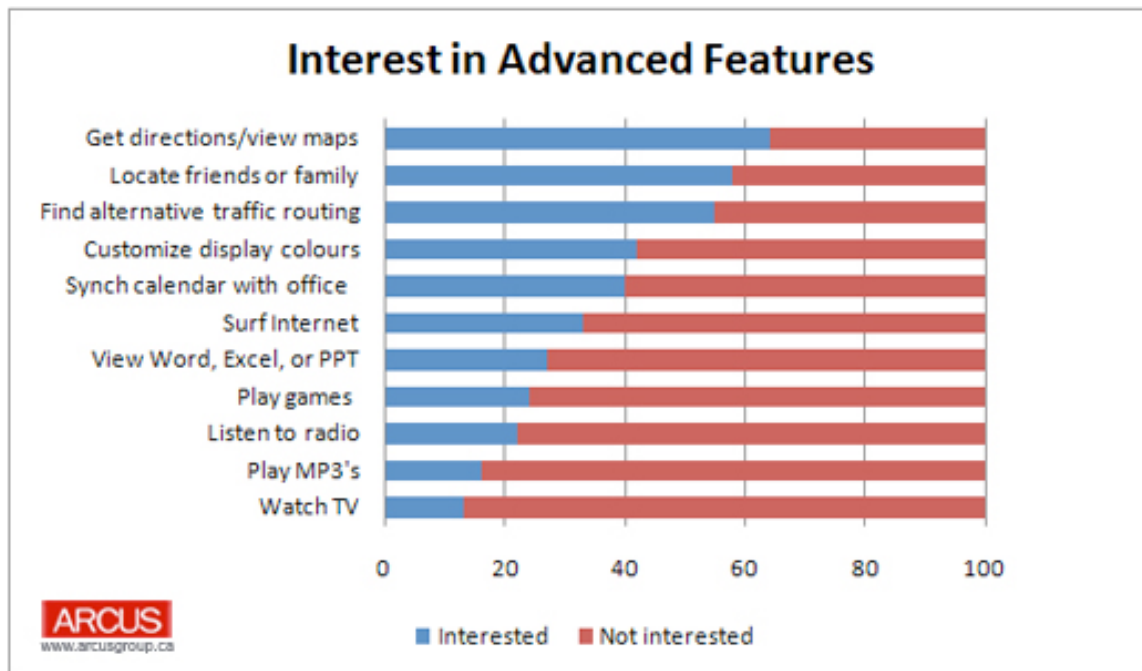
Table 1: Uses of Smart Phones Source: Canadian Marketing Association

Growth in M-commerce

The contextual, location based capabilities of smart phones are seen as a strong driver of mobile commerce. Data from Japan and South Korea, which are the most technologically advanced countries with respect to mobile commerce, show strong growth in average revenue per user. Specifically, ARPU (average revenue per user) has grown 13% in South Korea to $49.3 million (ARCUS survey) across an array of value added services while growth in North America has been restricted to messaging and data services.

Social Networking Goes Mobile

According to consultancy Informa, the number of mobile social-networking users will rise from 2.3% of global cell-phone users at the end of 2007 to as many as 23% of all mobile users by the end of 2012. The primary market segment is expected to be 18-34 year olds.

TRENDS: CORPORATE MARKET FOR SMART PHONES

InStat predicts that growth in corporate market will come from replacement of notebook computers with smart phones; that is, for many employees, a smart phone will serve as a viable substitute for heavier, bulkier, notebook computers. Data indicate that heavy users of smart phones are more likely to travel extensively as part of their jobs and are more likely to require real time connectivity to data and people.

Smart Phones as Productivity Tools

Although consumer applications for smart phones have received considerable publicity, perhaps because of the Blackberry phone, corporate smart phones are generally viewed as portable e-mail devices that perform ancillary administrative tasks such as expense tracking. This is not the case as evidenced by a growing range of software applications that bring data about business processes and business outcomes directly to smart phones over cellular networks.

Increased Concerns About Data and Network Security

As more sensitive information is transmitted and/or stored by smart phones, interest in security have increased accordingly. CIO's are concerned about a wide and growing array of issues and a recent survey indicates that 71% have increased spending on IT and data security for their mobile assets [4].

SMART PHONE PLATFORMS

The term "platform" is typically used to describe the hardware, software, and network technology of any given device. It is well suited to understanding the technology associated with smart phones and the expected gains in usability. The main differentiator among smart phones (as is the case with PC's) is the operating system software.

The various operating systems and their associated business models are presented in Table 2:

Organization	OS	Model
Apple	Mac OS X Mobile	Proprietary: Restricted Use
Palm	Palm	Proprietary: Restricted Use
RIMM	Blackberry	Proprietary:

		Restricted Use
Microsoft	Windows Mobile	Proprietary: For License
Symbian Foundation (Nokia)	Symbian	Open Source
Android Foundation (Google)	Android	Open Source
LiMo Foundation	Linux Mobile	Open Source

Table 2. Smart Phone Operating Systems

Worldwide market share for the third quarter of 2008 is presented in Figure 2.
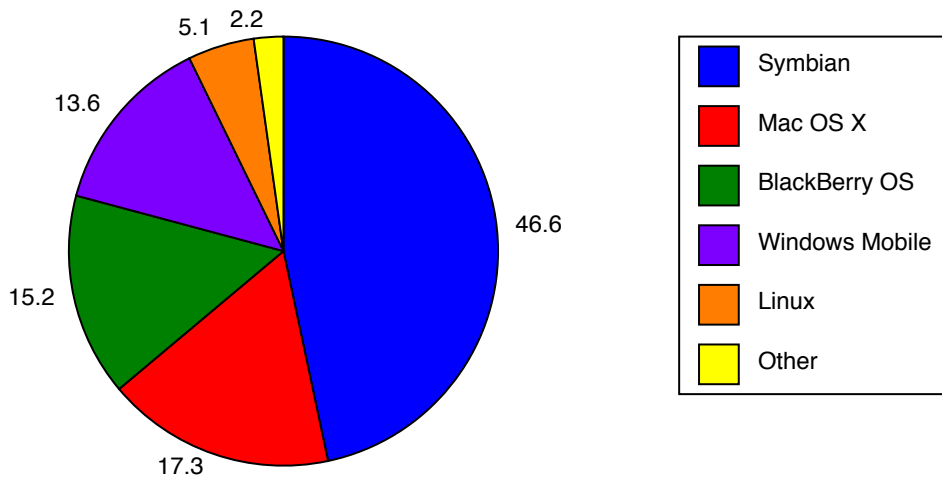
Figure 2. Worldwide Market Share of Smart Phones by OS

Apple, Palm, and Research-In-Motion have developed proprietary operating system software that is restricted to their devices; that is, it is not available for license to other handset manufacturers. controlled.

Microsoft offers its Windows Mobile operating system to any smart phone manufacturer willing to pay its licensing fees thereby retaining the business model it has used for Windows.

Finally, there are several open source and/or royalty free operating systems geared toward smart phones. In contrast to the proprietary model, open source or royalty free vendors have formed foundations with a governance structures and standards. Handset manufacturers apply for membership to the foundation and agree to accept its standards and, if accepted, are given access to the operating system at no cost.   Google is using

this model with the Android Foundation as is Nokia with its Symbian Foundation. Linux has also been ported to cell phones and is governed by the LiMo Foundation.

BUSINESS ECOSYSTEMS

Smart phones are based on connectivity and context. Connectivity drives the immediate and continual access to people and data while context refers to the ability to locate the user (and other users) in both time and space. Harnessing the advantages of connectivity and context cannot be accomplished by any one company, but rather requires a set of relationships that is most accurately viewed as a business ecosystem. James Moore defined a business ecosystem as "an economic community supported by a foundation of interacting organizations and individuals--the organisms of the business world. This economic community produces goods and services of value to customers, who are themselves members of the ecosystem. The member organizations also include suppliers, lead producers, competitors, and other stakeholders. Over time, they co-evolve their capabilities and roles, and tend to align themselves with the directions set by one or more central companies. Those companies holding leadership roles may change over time, but the function of ecosystem leader is valued by the community because it enables members to move toward shared visions to align their investments and to find mutually supportive roles." [5]

Figure 3 depicts a business ecosystem from the perspective of device manufacturers.
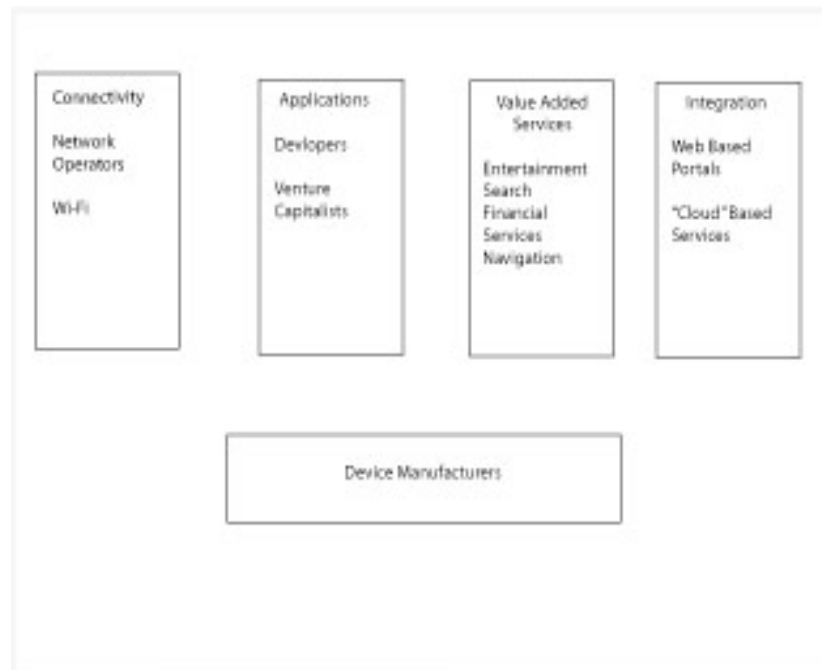
As indicated in the Figure, device manufacturers must manage a complex array of partnerships and relationships to effectively market and support their smart phones. These areas include connectivity, application software, value added services, and an integrated system that serves to coordinate the distribution of content, value added services and software. Application development and value added services can be viewed as smaller business ecosystems within a larger one.

Connectivity is obviously critical to smart phones and device manufacturers need to negotiate an array of relationships with network operators around the world. At present, virtually all connectivity comes from cellular phone network operators. Smart phones require 3G data capabilities and speeds to be used effectively and it is important for device manufacturers to form partnerships with network operators that offer secure, reliable service.

 Although WiFi is not a viable competitor to cellular networks, partnerships with WiFi providers can offer value added services that drive mobile commerce. For example, Apple has partnered with Starbucks to offer free WiFi access to its iTunes Music Store to allow iPhone users to purchase music that Starbucks is promoting in its stores.

The critical differentiating factor between smart phones and cellular handsets is that smart phones run application software. Manufacturers cannot supply the thousands of applications that are needed for smart phones to realize their value as mobile computing devices, and thus, must build and manage a developer network. Although established software houses will develop for smart phones, many applications will come from small start-up companies. A healthy developer ecosystem, thus, requires access to venture capital.

Value added services combine content, context and connectivity to enable mobile commerce. Many of these services will require specific application software and others will be based on partnerships with content providers that may or may not include network operators as well. Thus, value added services can also be viewed as a business ecosystem within a larger ecosystem.

The manner in which specific companies have structured and managed the generic business ecosystem presented in Figure 2 provides insight into their future direction. Three companies are considered in detail: Apple, Nokia, and Research-In-Motion.

Apple

Partnerships with Network Operators: Apple initially introduced the iPhone in the United States with an exclusive five-year partnership with AT&T. The terms and conditions of the partnership reflect Apple's long standing policy of maintaining a high level of control over its products and services. The first generation iPhone was sold exclusively through by AT&T or Apple, but it could activated only through Apple's iTunes Music Store. Thus, customers waited on long lines to purchase a product that they could not use until Apple authorized them to do so.

The global roll-out of the iPhone has proceeded slowly because Apple has continued its policy of exclusive partnerships with cellular network operators in specific countries and/or regions.

Apple enforces its policy of exclusivity by being the only device manufacturer than does not allow swappable SIM cards on its phones and the company has taken significant steps to make it very difficult to use an iPhone on non-authorized networks.

Developer Network: Apple's initial policy was to restrict developers to web based applications for the iPhone; that is, it was originally designed as a closed device with no outside access to the operating system. Apple abandoned this policy with the second generation iPhone, and introduced a new operating system that ran third party applications.   A software development kit for the iPhone was released well in advance and Apple actively courted established software companies to develop for the iPhone.

Apple has also partnered with an established venture capital firm in Silicon Valley to seed start-up companies. The $100 million dollar fund is focused solely on software development for the iPhone.

Apple, thus, has built an extensive developer network to support the iPhone. In less than 1 year, over 10,000 software applications have been built for the iPhone with 500 million downloads from the iTunes Store.

Value Added Services: Apple offers an extensive selection of entertainment with its partnerships with content providers for music, television and movies. Apple is the largest seller of digital music in the world and is a force in digital content. Navigation services are offered though a partnership with Google. Thus, the first mover advantage that Apple had with the iPod has been extended to the iPhone.

Integration: Apple's iTunes Music Store has evolved into a comprehensive web based portal that manages the user experience and the iPhone's operations.   Content, application software, and system software are distributed through a simple powerful interface that can be accessed directly from the iPhone. Cloud based services including shared files, contacts, and calendars among multiple devices (PC's, phones, iPods) are delivered through Apple's Mobile Me service for an annual fee.

Nokia

Partnerships with Network Operators: Nokia is world leader in both cellular handsets and smart phones and has a strong global presence with the exception of the American and Japanese markets. Accordingly, Nokia has partnerships with network operators throughout the world and with multiple network providers in the same country or region including India and China where Nokia is the market leader. Further, the Nokia-Siemens joint venture sells network technologies and gear to network operators.

Developer Network: In 2008, Nokia acquired 100 percent of Symbian and formed the Symbian Foundation to distribute the Symbian OS royalty free.   As of the second quarter of 2008, there were about 9,500 applications available for Symbian OS supported by the Symbian Developer Network.   These applications span the enterprise and consumer segments and represent products from leading software companies as well as smaller software houses. The scale of the Symbian

platform makes it difficult for developers to ignore. It is estimated that there are 226 million devices running Symbian OS of which 77 million are classified as smart phones.

Value Added Services: Nokia has launched a web based portal termed Ovi (which is the word for "door" in Finnish) which is an integrated suite of tools that allows users to share photos and coordinate calendars and contacts. Ovi is also the landing page for the Nokia Music Store which is to launch shortly. Nokia offers extensive navigation services which are built into its high end devices and is seen as a leader in this area.

Integration: Nokia's web portal and tools for integrating the user experience are much less sophisticated than are those offered by Apple. Although several reviewers have taken issue with the "control freak" nature of Apple's iTunes Music Store, it is an effective vehicle to deliver and coordinate content across all devices in the home (with the Mac as the digital hub), to ensure that application software for the iPhone is appropriate and secure (Apple can remotely remove offending applications from the iPhone) and that systems upgrades are offered seamlessly.

Nokia's web based services designed to deliver value added content and integrate content across PC's and smart phones are clearly a work in progress.   Nokia has taken a different approach from Apple which has placed all of its content and applications in one place (iTunes) while Nokia appears to be developing separate sites for games, music and maps which can be accessed through the Ovi portal or individually through the web. Further, while Apple offers a separate service (Mobile Me) to manage file sharing across devices, Nokia has chosen to include sharing as part of Ovi.

Research-In-Motion

Partnerships with Network Operators: Like Nokia, Research-In-Motion has partnered with leading cellular providers throughout he world.   However, RIMM is the only device manufacturer that runs its own network and has designed its own server software to push e-mail to its Blackberry smart phones. The added costs of this service has greatly reduced RIMM's market share in Asia and, to a lesser extent, has affected share in Europe as well. Blackberry phones are most widely used in the enterprise market in United States and Canada.

Developer Network: As of September, 2008, RIMM reported 19 million Blackberry users. The company introduced its first device in 1999 so that it took nearly a decade to reach19 million devices. In contrast, Apple sold over 10 million iPhones in one year.   This relatively slow growth rate and RIMM's focus on push e-mail for its corporate customer base has limited development of application software. There is a Blackberry Developer Network but the number of applications available for the device is unknown. Two developments are noteworthy. RIMM is moving away from encouraging developers to build browser based and JAVA based applications (an approach that Apple quickly abandoned) and the company plans to open an online applications store to compete with Apple's iTunes Music Store.   These developments suggest that RIMM wishes to rapidly increase its application and developer base.

Value Added Services: The primary value added service for Blackberry devices is a secure push e-mail system that is supported by hardware and software. Many CIO's feel that the redundancy

and security of this system justifies its additional cost. RIMM has just begun to pursue the
consumer market and the company offers little in the way of entertainment, content, and
application software.

A strong and deepening partnership with Verizon is one vehicle to offer value added services and
content using the Blackberry "push" model.   That is, Blackberry users would use the Verizon
network to access entertainment, games, and services based on a revenue sharing partnership.

 Integration: For enterprise customers, management of the end user experience is typically
governed by individual corporations. Some have written custom applications and many have
policies regarding their use and security. RIMM is, thus, in good stead in the enterprise market.
Management of the user experience in the consumer market remains in flux. The launch of an
online applications store should help spur the development and distribution of new applications
for the new touch screen Blackberry devices.   Rather than using the digital hub model, RIMM
might choose to manage upgrades to the operating system and offer value added services through
its partnership with Verizon. The Verizon network has been used to install patches to the
operating system for the Blackberry Storm device and working with Verizon might provide an
acceptable solution for the consumer market while allowing RIMM to continue to focus on its
enterprise customers.

MARKET DYNAMICS

As the growth in market for entry level and mid-priced cellular handsets slows and prices soften,
the smart phone market has gotten increased attention as a vehicle for growth and profitability
based on the higher prices these models command on the opportunity to earn revenue from
value-added services. The availability of open source operating systems such as Google's
Android OS, Linux Mobile, and Nokia's Symbian OS has lowered barriers to entry as has the
failure of a market leader to emerge. In this regard, while Nokia claims a 46% share of the smart
phone market, this number includes mid-range devices running the Symbian OS that are not true
smart phones. Nokia's share of the high-end smart phone market (e.g., true mobile computing
devices) is much smaller and the company has no presence in the United States with less than a
2% share.

Recent developments suggest an increasingly competitive market with new entrants. Garmin,
the GPS device maker, has recently announced that it plans to develop a smart phone that
incorporates the company's GPS technology. Palm's new Pre touch screen device introduced at
the 2009 CES show was very well received and it is expected to do well (although Apple is
claiming patent infringements and has threatened litigation). A resurgent Palm will certainly
change the competitive dynamics of the smart phone market.   Finally, Dell has indicated that is
seriously considering developing a smart phone and presumably to offset slowing growth in the
PC market.

The situation is somewhat similar to the early days of the PC market where there were several
manufacturers and several operating systems competing for both developers and customers.
That shakeout when through two phases with MS-DOS emerging as the dominant operating
system and the Macintosh holding a strong niche position. The second phase was characterized

by the dominance of Windows OS with 90% plus of the market for more than a decade. Microsoft's dominant position is slowly being eroded by Mac OS which has a 10 percent market share (up from 2 percent a decade ago), by trends toward web-based computing, and by Linux.

It is not clear at this point if the smart phone market will follow the same pattern. It is, however, highly unlikely that even with a 33% annual growth rate that the market can support six or more incompatible operating systems.   Thus, some companies are going to win big and others will be forced to exit the market further strengthening the position of the survivors.

REFERENCES

[1] Instat Research as reported in www.cio.com

[2] www.gartner.com

[3] www.t3.com

[4] www.cio.com

[5] Wikipedia

EXPLORING THE USE OF WIRELESS WEB SERVICES IN THE UNITED STATES

Suhong Li, Bryant University, Smithfield, RI, (401) 232-6503, sli@bryant.edu
Harold Records, Bryant University, Smithfield, RI, (401) 232-6172, hrecords@bryant.edu

ABSTRACT

This study investigates the current status of wireless web services use in the United States and empirically investigated the influence of technology acceptance model factors, social influence factors, online shopping experience and gender on the use of wireless web services. It was found that overall use of wireless web services has increased. It is also found that perceived usefulness positively impacts the use of all types of wireless web services; interpersonal influence impacts the use of communication and information services; and online shopping experience positively impacts the use of transaction, and communication/information services. In addition, the results show that male respondents have a significantly higher level of wireless web services use than females.

Keywords: Wireless web services, mobile commerce, technology acceptance model, social factor, gender

INTRODUCTION

The Internet has dramatically changed the way we study, work and conduct business today. The Internet offers many types of services including communication (email, instant messaging, social networking, etc.), information (access news, stocks, weather, etc.), entertainment (games, music, movies, etc.), and transactions (purchasing items, banking, paying bills, etc.). The most recent trend involves expanding Internet services to mobile devices such as cell phones and PDA's which has promised users "anytime, anywhere" access to information for work and personal communication [4] [8]. The mobility associated with these devices has resulted in naming this new trend mobile commerce or m-commerce. M-commerce is a natural extension of e-commerce as they share fundamental business principles and m-commerce acts as another channel through which value can be added to e-commerce [4].

Strictly speaking, m-Commerce refers only to transactions conducted through mobile devices. However, the term m-Commerce is often used to refer to all types of services offered through mobile devices, namely wireless web services or mobile Internet Services. This paper will use the broad definition of m-Commerce. We consider m-Commerce to include four general categories of wireless web services; (i) communication services, (ii) transaction services, (iii) information services and (iv) entertainment services.

Many observers predict that m-Commerce is at the forefront of the next revolution in global technology. In 2004, there were 809 million mobile commerce subscribers worldwide [30]. This number increased to 3 billion in 2007. By the end of 2006, Boston-based Celent, reports that the mobile commerce market was worth $24 billion, with Japan and South Korea accounting for nearly 60 percent of the total [28]. Today, more than half of Japan's 70 million cellular phone subscribers have Net Access via handsets. Young Japanese teenagers are keeping up with trends

in fashion by using their cell phones to purchase the latest styles [20]. China had more than 150 million mobile subscribers in 2004 [30]. The US has been slower to adopt m-commerce compared to Japan and Europe, having approximately 20.3 million mobile services subscribers in 2004. The lower adoption rate of m-Commerce in the US may be explained, in part, by the fact that US consumers are more accustomed to conducting electronic transactions using desktop/laptop computers [12]. Siau & Shen [29] suggest that small screens, low-resolution displays, and tiny multifunction keypads make mobile devices less attractive for consumers in the US. On the other hand, the higher adoption rate of mobile technology in Europe and Asia may be due to the fact that the technology is moving at such a fast pace that these countries are simply skipping wired technology and moving directly to wireless technology [30].

There is some evidence that mobile commerce in the US is starting to catch on and more people are beginning to use their cell phones to connect to the web. For example, eBay announced plans to let its subscribers use text messages on cell phones to order and pay for items [10]. Google has teamed up with phone maker Sony Ericsson to provide mobile search engines that will allow Google to display advertisements directly to cell phone users [14]. Cell phone providers such as T-Mobile have recognized the limitations of tiny keyboards and are offering Picture Messaging Services and Quick Notes to make it possible to very quickly input and send frequently repeated messages. According to ABI research, mobile commerce sales in the US will grow 100% in 2010 to $2.4 billion from $1.2 billion in 2009, That's following a 203% jump in 2009 from $396 million in 2008 [6]. According to Charlton [3], 37% of US Smartphone users have made a (non-mobile) purchase on their handsets in the last six months. 19% have purchased music on their phones, 14% have bought books, DVDs or games, while 12% have purchased movie tickets.

Many studies have been conducted to investigate how factors such as gender, cost, perceived risk, compatibility, perceived usefulness and wireless trust influence consumers' adoption of m-Commerce [11] [12] [18] [27] [29] [32] [34]. In addition, several studies have included the impact of social factors (subject norm, interpersonal influence and external influence) on the m-Commerce adoption [2] [13] [23].

Extending the previous studies, this study aims to provide an update of wireless web services use in the US and explore the factors impacting its use will be helpful in understanding the current status of wireless web services use and identifying factors facilitating its adoption and use.

LITERATURE REVIEW ON WIRELESS WEB SERVICES ADOPTION

The Technology Acceptance Model (TAM) was first proposed by Davis [5] and has since become one of the most widely accepted models of technology adoption. TAM is an adaptation of Fishbein and Ajzen's [9] theory of reasoned action (TRA), in which TRA's attitudinal determinants, derived separately for each behavior, are replaced with a set of two variables perceived ease of use and perceived usefulness [21]. TAM suggests that an individual's perceived ease of use and perceived usefulness of a particular technology determine the individual's behavioral intentions which in turn determine his or her acceptance and use of the technology. TAM posits that the impact of other external variables is fully mediated by the perceptions of ease of use and usefulness [5].

Perceived ease of use and perceived usefulness have been found to be important predictors of m-Commerce [18] [19] [32]. Siau & Shen [29] suggest that the perception of ease of use is particularly important for m-Commerce adoption because current mobile devices have small screens, low-resolution displays, and tiny multifunction keypads that make them more difficult to use and less attractive for consumers.

Many researchers suggest that TAM should be expanded to include additional variables in order to better predict technology adoption [8] [13] [23] [32. The factors suggested include price, wireless trust, attitude toward technology, gender and so on. In addition, in a review of TAM research, Lee et al. [15] suggested that more research is needed to investigate the causal linkage between social influence and IT adoption. Legris et al. [16] point to the need for including additional variables related to human and social change processes. In this study three constructs were adopted that consider social influence on the adoption and use of wireless web services: subjective norm; interpersonal influence and external influence.

Venkatesh and Davis [31] describe subjective norm as the person's perception that most people who are important to him think that he should or should not perform the behavior in question. The influence of subjective norm on technology adoption is inconclusive [15]. Few studies have considered subjective norm in the adoption of mobile commerce. Khalifa and Shen [13] found that subject norm indirectly impact a user's intention to adopt m-Commerce. The study of Pedersen [23] shows that subject norm impact a user's intention to adopt m-commerce that in turn impacts its actual use.

Battacherjee [1] suggested that a broader conceptualization of subjective norm to include both external (mass-media) and interpersonal influences. External influence refers to the influence of mass media on the person's perception on m-Commerce and interpersonal influence refers to the influence of a person's friends/colleagues on his/her perception of m-Commerce. The results of Battacherjee [1]'s study show external influence is a significant determinant of subjective norm. Pedersen [23] found that external influence and interpersonal influence significantly impact subject norm that in turn impact a person's intention to use m-Commerce.

In addition, gender has been suggested to play a role in the adoption of new technologies such as wireless web services. Rodgers and Harris [26] claim that in nearly every study that has examined gender and e-Commerce, males are typically shown to be the dominant on-line shoppers. They suggest that emotion, trust and convenience are three critical factors that influence women and men's participation in e-Commerce. Yang and Lester [33] also found that men were more involved in e-Commerce purchases than women. They attributed the fact to women being more concerned about computer anxiety, concern about money and less experience being on-line. Dittmar et al. [7] suggest that credit card security and general competence in using computers is not sufficient to explain the differences between male and female on-line shopping patterns. They suggest that women's emotional and psychological involvement in the whole shopping and buying process are important factors in explaining gender differences in e-commerce use.

On the other hand, some studies have found that gender gaps are lessening or disappearing [24]. Ray et al. [25] found that there is no significant gender difference toward computer anxiety.

Morahan-Martin and Schumacher [22] found that attitudes towards new technology, but not gender, predicted Internet and computer competencies and experiences.

In addition, since m-Commerce is a natural extension of e-Commerce, it is logical to assume that a person with Internet shopping experience will trust online transactions and will most likely to continue purchasing using a mobile device.

Based on the above discussions, this paper considers perceived ease of use, perceived usefulness, subjective norm, external influence, internal influence, Internet shopping experience and gender as influencing factors of wireless web services use.

RESEARCH METHODOLOGY

A paper-based survey was distributed to undergraduate students in several classrooms at a private university in the Northeast US. Students were asked to voluntarily participate in the survey. Items for measuring perceived ease of use and usefulness were adopted from Khalifa and Shen [13] and items for measuring subjective norm, interpersonal influence and external influence were adopted from Pedersen [23] (See Appendix A). Internet shopping experience is measured by number of year a person has shopped online. Construct validity is discussed in a section to follow.

A total of 181 usable responses were received. The profile of the adopters is summarized in Table 1. Of the sample, about 54% were male. About 27% of the respondents were under the age of 20 with 73% falling between 20 and 25 years old. The majority of the respondents (91%) were business administration majors with the balance majoring in liberal arts. Caution must be taken in explaining the results. The results may not apply to older age groups.

Regarding online shopping experience, about 24% of the respondents have less than 2 years of online shopping experience, 30% of the respondents have 2 or 3 years of online shopping experience, with the rest (46%) has over 4 years of online shopping experience.

DATA ANALYSIS AND DISCUSSION

This section will first discuss the status of wireless web services use followed by multiple regression analysis of the factors impacting wireless web services.

The Current Status of Wireless Web Services Use

Twenty one wireless web services were identified and grouped into one of four categories: communication, transaction, information and entertainment (See Appendix B). Harris et al. (2005) employed a similar classification scheme to measure wireless web services. We also added some recent services such as Facebook, Youtube and e-books that can be accessed over a mobile device.

The respondents were asked to indicate how often they have used each service through a mobile device (from every day, several times a week, once a week, once a month to never). The results are shown in Table 2. It can be seen that text messaging is the dominant service used by 98% of

the respondents on a daily basis, followed by email and Facebook that are used by 65% and 51% of the respondents on a daily base respectively. In addition, accessing news and information, weather reports and playing music are used by about one third of respondents on a daily basis.

TABLE 1. Demographic Profile of the Respondents

		Respondents	Percentage
Gender	Male	97	53.9%
	Female	83	46.1%
Age	below 20	48	26.5%
	20-25	132	73.5%
Major	Business Administration	162	90.5%
	Liberal Arts	17	9.5%
# of years you have shopped online	<2	44	24.4%
	3-4	56	31.1%
	5-6	55	30.6%
	7-8	17	9.4%
	>8	8	4.4%

TABLE 2. The Current Status of Wireless Web Services Use (in Percentage)

	Every Day	Several Times a week	Once a week	Once a month	Never
Communication Services					
Email	65.0	6.7	2.2	1.1	25.0
Text Messaging	97.8	1.7	.6	0	0
Multimedia messaging	21.2	29.6	21.8	11.2	16.2
Twitter	5.6	2.8	3.4	4.5	83.6
Facebook	50.8	11.6	4.4	3.9	29.3
LinkedIn	2.3	1.7	1.1	7.1	87.5
Transaction Services					
Online Banking	14.4	14.9	16.6	12.7	41.4
Online shopping	2.8	5.0	11.1	36.7	44.4
Pay bills	2.2	2.2	3.3	19.9	72.4
Purchase tickets	0	2.8	3.9	40.3	53.0
Information Services					
Access News and Information	39.7	19.0	11.7	5.0	24.6
Access stock quotes	16.0	6.6	14.4	11.0	51.9
Get weather report	35.9	19.3	9.9	8.8	26.0
Get map and driving directions	5.0	18.8	18.2	25.4	32.6
Get coupon and deals	3.3	3.9	8.3	18.2	66.3
Entertainment Services					
Play games	12.2	15.6	17.2	21.1	33.9
Play music	30.7	19.6	8.9	8.9	31.5
Play movies	3.9	11.6	8.8	16.6	59.1
Access e-books	1.7	2.8	6.6	8.3	80.7
Access Youtube video	16.8	18.4	14.5	13.4	36.9
Ringtones and screen savers	5.1	6.2	13.0	27.1	48.6

The values in column 2-5 of Table 2 represent-the percentage of respondents who have used each service (at least once a month), the last column in Table 2 refers to the percentage of the respondents who have not used such service. It can be seen that that email, text messaging, Facebook, access to news and information, getting weather reports, maps and driving direction*s,* playing games, playing music and access Youtube video are adopted by at least two thirds of respondents. In contrast, the results show that more than 80% of the respondents have not used Twitter, LinkedIn, paid a bill, or accessed an e-Book over their mobile devices.

Table 2 also shows that at least half of the respondents have used transaction services including online banking, online shopping and ticket purchasing. Those numbers are higher compared to previous studies. For example, Li et al. [17] found that less than 10% of the respondents have used any type of transaction services in 2006. The result of this study show the adoption of transaction services over mobile devices is growing with the introduction of 3G/4G network, decreasing price and improved wireless devices offered by vendors.

It can be concluded that communication services received the highest level of the use, followed by information and entertainment services. Transaction services received the lowest level of use.

Factors Impacting Wireless Web Services

This section will investigate factors impacting four types of wireless web services. First, the validity of the constructs will be tested through a factor analysis and a reliability test, followed by multiple regression analysis testing factors impacting the use of wireless web services.

Construct validation

As mentioned earlier, this paper includes perceived ease of use, perceived usefulness, subjective norm, external influence, internal influence, Internet shopping experience and gender as influencing factors of wireless web services adoption.

To validate ease of use, perceived usefulness, subjective norm, external influence, and internal influence (See Appendix A), a factor analysis was conducted using principal components as the means of extraction and the varimax method of rotation. After dropping UF01 (Learning to use wireless web services is easy) and II04 (Some of my friends/colleagues recommended I should try out wireless web services), five factors emerged as shown in Table 3. For simplicity, only loadings above .40 were displayed. All items loaded on their respective factors and there were no items with cross-loadings greater than .50. These results suggest that there is strong convergence and support the validity of the constructs. The last column of Table 3 shows that the Cronbach alpha coefficient of all the constructs is greater than .70, indicating good reliability of the constructs.

A similar factor analysis was conducted to all wireless web services except the ones adopted by less than 20% of the respondents. After dropping a few items, three factors emerged as shown in Table 4. All items loaded on their respective factors and there were no items with cross-loadings greater than .50. Information and Communication services loaded on a factor and will be treated as a single factor in later analysis.

TABLE 3. Factor Analysis of Constructs Impacting Wireless Web Services

	Ease of Use	Usefulness	Subject Norm	Interpersonal Influence	External Influence	α
EOU03	.846					
EOU02	.839					
EOU01	.822					.93
EOU04	.799					
EOU05	.791					
UF03		.854				
UF02		.834				
UF05		.811				.94
UF04		.783				
SN02			.866			
SN01			.829			.95
SN03			.826			
II01				.881		
II02				.833		.89
II03			.488	.624		
EI02					.810	
EI01					.732	.73
EI03			.421		.518	
Eigenvalue	9.2	2.0	1.4	1.1	.9	
% of Variance	21.8	19.2	16.8	13.3	10.6	
Cumulative % of variance	21.8	41.0	57.9	71.1	81.8	

Regression Analysis of Factors Impacting Wireless Web Services

To test for the impact of the various factors on the use of wireless web services, three linear regression analyses were conducted, using seven factors as independent variables and each type of wireless web services (transaction, entertainment and communication/information services) as a dependent variable. The results are shown in Table 5. It can be seen that transaction services are positively impacted by perceived east of use, usefulness, online shopping experience and gender. The entertainment services are positively impact by perceived ease of use, usefulness and gender, and the communication and information services are positively impacted by usefulness, interpersonal influence, online shopping experience and gender. Table 5 also shows that external influence and subject norm have no impact on any type of wireless web services.

The results indicate that gender is significant in three types of wireless web services. Further analysis shows that males have higher mean on each type of the services than female, indicating the existence of the gender gap in the use of wireless web services.

The finding also shows the importance of interpersonal influence on the adoption of communication/information services. As more than one individual are involved in a communication and the value of the technology is dependent on the degree to which the technology is used by the members of an individual's communication network. Consequently, interpersonal influence may be an important factor in the adoption of communication services.

TABLE 4. Factor Analysis of Wireless Web Services

Item	Communication/Information	Transaction	Entertainment	α
CS05	.867			
CS01	.813			
IS01	.746		.407	
IS03	.654		.515	.87
CS03	.504			
TS03		.864		
TS02		.793		
TS04		.746		.82
TS01	.475	.618		
ES03		.458	.761	
ES05			.751	
ES01			.730	.83
ES02	.427		.697	
Eigenvalue	6.5	1.4	1.2	
% of Variance	26.2	21.9	21.8	
Cumulative % of variance	26.2	48.1	69.9	

TABLE 5. Regression Analysis of Wireless Web Services

Independent Variables	Dependent Variables		
	Transaction	*Entertainment*	*Communication/Information*
Ease of Use	**.18 (.05)**	**.17 (.05)**	.10 (.26)
Usefulness	**.19 (.06)**	**.18 (.06)**	**.28 (.00)**
External Influence	.04 (.67)	-.14 (.14)	.16 (.09)
Interpersonal Influence	-.08 (.42)	.00 (1.00)	**.20 (.05)**
Subject Norm	-.07 (.49)	.08 (.43)	-.06 (.54)
Online Shopping Experience	**.17 (.02)**	.11 (.12)	**.16 (.03)**
Gender	**.21 (.00)**	**.35 (.00)**	.21 (.00)
R	.40	.48	.44
R^2	.16	.23	.19
F-statistics	4.54	7.20	5.80
Significance	**0.00**	**.00**	**.00**

CONCLUSIONS AND IMPLICATIONS

This study investigates the current status of wireless web services use in the United States and empirically investigated the influence of technology acceptance model factors (perceived ease of use and usefulness), social influence factors (subject norm, external influence, and interpersonal influence), online shopping experience and gender on the use of wireless web services. It was found that overall use of wireless web services has increased. Communication services received the highest level of use, followed by information services, entertainment services and then transaction services. Among all services, text messaging is the dominant service used by 98% of the respondents on a daily base, followed by email and Facebook that are used by 65% and 51% of the respondents daily. It is also found that perceived usefulness (technology acceptance model factors) positively impact the use of all types of wireless web services; interpersonal influence (social factors) impacts the use of communication and information services; and online shopping experience positively impacts the use of transaction, and communication/information services. In addition, the results show that a gender gap still exists in the use of wireless web services. The male respondents have a significantly higher level of wireless web services use than females.

The findings of this study show the validity of Technology Acceptance Model (TAM) in explaining wireless web services adoption. Perceived usefulness will impact the adoption of all types of wireless web services, and perceived ease of use will impact the adoption of transaction and entertainment services. Based on the regression coefficients in Table 5, it can be concluded that perceived usefulness is of more importance in impacting the use of wireless web services because perceived usefulness has higher regression coefficients than ease of use.

The results also partially support the impact of social factors on the adoption of wireless web services. Among all three social factors (subjective norm, interpersonal influence, external influence), only interpersonal influence is found to have a significant impact on the use of communication and information services. No significant impact of subjective norm and external influence on the adoption of wireless web services was evident in the specific sample used in this study which focused on university students under 25 years old. The results may be different if we expand the sample to include general population with various ages. Future research should be conducted to investigate the adoption and use of m-Commerce for other groups spanning a range of ages, occupations, cultures and locations.

The results also show that a person's online shopping experience (using a computer) is significantly associated with his/her use of transaction services over a mobile device. This finding validates our earlier statement "m-Commerce is a natural extension of e-Commerce". People who have adopted e-Commerce and who have a long history of online shopping experience will be the first ones jumping into m-Commerce and purchasing items using a mobile device.

In addition, the results of this study support the existence of gender gap in the use of wireless web services. Male respondents have a higher level of wireless web services use than females. The subjects in this study were university students attending college in the northeast United States. College students in general, may be less likely than the general population to express gender stereotypical behavior [7] and tend to be more technologically literate than the general population, more experienced with the Internet and more likely to engage in e-Commerce

transactions. Nonetheless, the fact that gender differences were observed in a university population suggests that gender differences may exist and may be more pronounced in the general population.

APPENDIX A: FACTORS IMPACTING WIRELESS WEB SERVICES

Ease of Use (EOU)
EOU1: Learning to use wireless web services is easy
EOU2: It is easy to make wireless web services do what I want them to
EOU3: My interaction with wireless web services is clear and understandable
EOU4: I find it easy to interact with wireless web services
EOU5: I find it easy to use wireless web services

Usefulness (UF)
UF01: Using wireless web services helps me save time
UF02: Wireless web services make me a better consumer
UF03: Using wireless web services improves my efficiency as a consumer
UF04: Wireless web services are useful to me as a consumer
UF05: Wireless web services increase my effectiveness as a consumer

External Influence
EI01: The media is full of reports, articles and news suggesting the use of wireless web services is a good idea
EI02: Media and advertising consistently recommend using wireless web services
EI03: In my profession it is advisable to use mobile commerce services

Interpersonal Influence
II01: Almost all of my friends/colleagues use wireless web services
II02: Almost all my friends/colleagues think using wireless web services is a good idea
II03:My friends/colleagues think that we should all use wireless web services
II04: Some of my friends/colleagues recommended I should try out wireless web services

Subjective Norm
SN01: People important to me think I should use wireless web services
SN02: People who influence my behavior think I should use wireless web services
SN03: People whose opinion I value prefer me to use wireless web services

APPENDIX B: WIRELESS WEB SERVICES

Communication Services (CS)
CS01: Email
CS02: Text messaging
CS03: Multimedia messaging
CS04: Access twitter
CS05: Access Facebook
CS06: Access LinkedIn

Transaction Services (TS)
TS01: Online banking
TS02: Online shopping
TS03: Pay bill
TS04: Purchase tickets

Information Services (IS)
IS01: Access news and information
IS02: Access stock quotes
IS03: Get weather report
IS04: Get maps and driving directions
IS05: Get coupons and deals

Entertainment Services (ES)
ES01: Play games
ES02: Play music
ES03: Play movie
ES04: Access e-book
ES05: Access youtube video
ES06: Ringtone and screen saver

REFERENCES

[1] Battacherjee, A. (2000). Acceptance of e-commerce services: the case of electronic brokerages, IEEE Transactions on Systems, Man and Cybernetics, 30 (4), pp.411-420.

[2] Bahtti, T. (2007). Exploring Factors Influencing the Adoption of Mobile Commerce, Journal of Internet Banking and Commerce, 12(3), pp.1-13.

[3] Charlton, G. (2010). Mobile Commerce Stats Round Up, EConsultancy, retrieved November 20 from http://econsultancy.com/us/blog/5870-mobile-commerce-stats-round-up.

[4] Coursaris, C., Hassanein, K., and Head, M. (2003), "M-Commerce in Canada: An interaction framework for wireless privacy", Canadian Journal of Administrative Studies (CJAS), 20(1), 54-73.

[5] Davis, F., Bagozzi, R. P., & Warshaw, P. R. (1989), "User Acceptance of Computer Technology: A Comparison of Two Theoretical Models," Management Science *35(8)*, 982-1003.

[6] Deatsch, K. (2010), "U.S. m-commerce sales to hit $2.4 billion this year, ABI Research says", Internet Retailer, retrieved November 20 from http://www.internetretailer.com/2010/02/16/u-s-m-commerce-sales-to-hit-2-4-billion-this-year-abi-researc.

[7] Dittmar, H., Long K. and Meek, R. (2004). Buying on the Internet: Gender Differences in On-Line and Conventional Buying Motivations. *Sex Roles*, 50(5/6), 423.

[8] Fang, X., Chen, S., Brzezinski, J., Xu, S. (2005), "Moderating Effects of Task Type on Wireless Technology Acceptance", Journal of Management Information Systems, 22(3), pp.123-157.

[9] Fishbein, M., & Ajzen, I., *Belief, Attitude, Intention and Behavior: An Introduction to Theory and Research.* Addison-Wesley, 1975, Reading, MA.

[10] Gohring, N. (2006). Payapl to offer new mobile services. MacWorld, retrieved April 4, 2006 from http://www.macworld.com/news/2006/03/23/paypal/index.php.

[11] Harris, P., Rettie, R., and Kwan, C. C. (2005). Adoption and Usage of M-Commerce: a Cross-Cultural Comparison of Hong Kong and the United Kingdom. Journal of Electronic Commerce Research, 6(3), 210-224.

[12] Haque, A. (2004). Mobile Commerce: Consumer Perception and It's prospect on Business Operations in Malaysia. The Journal of American Academy of Business, March 2004, 257-262.

[13] Khalifa, M., Shen, K. N. (2008). Explaining the Adoption of Transactional B2C Mobile Commerce, Journal of Enterprise Information Management, 21(2), pp.110-124.

[14] Kharif, D. (2006). Why Googles' Going Mobile. Business Week. Retrieved April 4, 2006 from http://www.businessweek.com/technology/content/mar2006/tc20060302_556420.htm?campaign_id=rss_tech

[15] Lee, Y., Kozar, K., & Larsen, K. (2003), "The Technology Acceptance Model: Past, Present and Future," *Communications of the ACM (Volume 12, Article 50)*, 752-780.

[16] Legris, P., Ingham, J., & Collerette, P.(2003), "Why Do People Use Information Technology? A Critical Review of the Technology Acceptance Model," Information and Management 40(3), 191-204.

[17] Li., S., Lin, R., Records, H. (2006), "A Comparison of Mobile Commerce Adoption between the US and China", the 3rd Sino-US E-Business Advanced Forum, Tianjin, China, June 3-5, 2006.

[18] Liu, C. and K. P. Arnet. (2002). Raising a Red Flag on Global WWW Privacy Policies. Journal of Computer Information Systems, 43(1), 117-127.

[19] Lu, J., Yu, C., & Liu. (2005). Facilitating Conditions, Wireless Trust and Adoption Intention, Journal of Computer Information Systems, Fall, 17-24.

[20] McCurry, J. (2006). "Handset Bring the mall to you in Japan's m-shopping craze. The Guardian, retrieved April 4, 2006 from http://www.guardian.co.uk/mobile/article/0,,1708426,00.html

[21] McFarland, D. J., & Hamilton, D., "Adding Contextual Specificity to the Technology Acceptance Model," Computers in Human Behavior *22*, 2006, 427-447.

[22] Morahan-Martin, J., Schunacher, P. (1999). Comparison of computer and internet competency, experiences and skills by gender (Abstract), *Proceedings of the Annual Meeting of the Society for Computers in Psychology*, 1999, Los Angeles, CA.

[23] Pedersen, P. E. (2005). Adoption of Mobile Internet Services: An Exploratory Study of Mobile Commerce Early Adopters, Journal of Organizational Computing and Electronic Commerce, 15(2), pp 203-222.

[24] Rainer, R. K., Laosethakul, K., Astone, M. K. (2003). Are gender perceptions of computing changing over time, Journal of Computer Information Systems, Summer, 108-114.

[25] Ray, C. M., Sormunen, C., Harris, T. M. (1999). Men's and women's attitudes toward computer technology: A comparison, Office System Research Journal, 17(1), 1.

[26] Rogers, S. and Harris. M. (2003). Gender and e-Commerce: An exploratory study, Journal of Advertising Research, September 2003, 322-329.

[27] Sarker, S. & Wells, J. (2003). Understanding Mobile Handheld Device Use and Adoption. Communications of the ACM, 46(12), 35-40.

[28] Schatt, D. (2006). US Warms to Mobile Commerce, Red Herring, Retrieved April 4, 2006 from http://www.redherring.com/Article.aspx?a=15728&hed=US+Warms+to+Mobile+Commerce+

[29] Siau, K., Shen, Z. (2003). Building Customer Trust in Mobile Commerce. Communications of the ACM, 46(4), 91-94.

[30] Statistics for Mobile Commerce (2005), ePay News, Retrieved April 4, 2006 from http://www.epaynews.com/statistics/mcommstats.html#24.

[31] Venkatesh, V., & Davis (2000), F., "A Theoretical Extension of the Technology AcceptanceModel: Four Longitudinal Field Studies," Management Science 46(2), 186-204.

[32] Wu, J. H., Wang, S. C. (2005) What Drives Mobile Commerce? An Empirical Evaluation of the Revised Technology Acceptance Model. Information and Management, 42, 719-729.

[33] Yang, B. and Lester, D. (2005). Gender differences in e-commerce, Applied Economics, 37, 2077-2089.

[34] Zhang, J. (2009). Exploring Drives in the Adoption of Mobile Commerce in China, The Journal of American Academy of Business, 15(1), pp.64-69.

EMERGING BUSINESS MODELS FOR MOBILE AD HOC NETWORKS AND LOCATION-BASED SERVICES

Subhankar Dhar
Department of Management Information Systems
San Jose State University
Email: subhankar.dhar@sjsu.edu
Phone : 408 924 3499

ABSTRACT

As broadband wireless access technologies mature and portable handheld devices including cell phones are equipped with more capabilities, mobile ad hoc networks are easily deployed. Various applications and services including location-based services and mobile advertising provide opportunities for new businesses. This creates additional source of revenue generation for not only service providers but also application developers, content providers and all the relevant players in the mobile communications ecosystem. This paper outlines various business models associated with mobile ad hoc networks, location-based based services and mobile advertising.

Key words: ad hoc network**,** Location-based services, mobile advertising, wireless communication, GPS

1. INTRODUCTION

The growing importance of ad hoc networks can hardly be exaggerated. An ad hoc network is a collection of nodes that form a network and communicate with each other via wireless links. The past decade has shown a phenomenal growth in wireless communications and the future of ad hoc networks looks very promising from applications standpoint.   The ad hoc nodes are small, low in power, and are inexpensive. These networks are usually multi-hop in nature and they have high redundancy in ambient conditions. Ad hoc networks are self-organizing and consist of hundreds or even thousands of nodes having limited battery power [6].

Mobile	The nodes may not be static in space and time resulting in a dynamic network topology. Nodes can move freely and independently. Also some new nodes can join the network and some nodes may leave the network.
Wireless	MANET uses wireless medium (radio, infrared, etc) to transmit and receive data. Nodes share the same media.
Self-organizing, distributed and infrastructure-less	They are self-organizing in nature. There is no centralized control which implies that network management will have to be distributed across various nodes. This makes fault detection and management quite difficult.
Multi-hop	A message from source node to destination node goes through multiple nodes because of limited transmission radius. Every node acts as a router and forwards packets from other nodes to facilitate multi-hop routing.
Scarce resources	The wireless links have limited bandwidth and variable capacity. They are also error prone. In addition, the mobile nodes have limited battery power along with limited processing power. So energy is a scare resource.
Temporary and rapidly deployable	These networks are temporary in nature. There is no base station. Whenever the nodes are within their transmission radius, they form an ad hoc network. Hence they are rapidly deployable.
Neighborhood awareness	Host connections in MANET are based on geographical distance.

Table 1: Characteristics of Mobile Ad Hoc Network (adapted from [6])

Some typical applications of mobile ad hoc networks are as follows [6].

a) *Surveillance and security:* Ad hoc networks are used for monitoring highly secured places and buildings. More and more ad hoc networks will be deployed for defense and homeland security. Ad hoc networks can also be used in minefields.

b) *Environmental and habitat monitoring:* These inexpensive devices, once deployed, can offer accurate information about temperature, detect chemicals, critical environment conditions (e.g. generate wild fire alarms); monitor certain behavior patterns like movements of some animals, etc. These networks can be deployed in agricultural fields to monitor temperature, water levels, etc. to ensure ideal conditions for vegetation.

c) *Personal area and home networking*: Ad hoc networks are quite suitable for home as well as personal area networking applications. Smart ad hoc nodes can be embedded in various home appliances, refrigerators, micro-wave ovens and can be monitored and operated remotely when connected via external networks like Internet, etc. from outside.

d) *Emergency and Health services*: Ad hoc networks will provide solutions to emergency services. In case of some disasters like flood, earthquake, ad hoc networks will be able to detect these conditions and generate alerts for proper actions. Other applications include remote integrated patient monitoring, drug administration in hospitals, etc.

e) *Military applications*: In battlefield, ad hoc networks can be deployed for communications among the soldiers in the field. Different military units are expected to communicate, cooperate with each other and within a specified area. In these kinds of low mobility environments, ad hoc network is used for communications where virtually no network infrastructure is available. They are low in cost, self-organized, self-balancing and self-healing. It is easy to scale. They are also used for monitoring forces.

f) *Ubiquitous and embedded computing applications*: With the emergence of new generations of intelligent portable mobile devices, ubiquitous computing is becoming a reality. As predicted by some researchers (Weiser, 1993), ubiquitous computers will be around us, always doing some tasks for us without our conscious effort. These machines will also react to changing environment and work accordingly. These mobile devices will form a ad hoc network and, gather various localized information and sometimes inform the users automatically.

g) *Inventory control*: Ad hoc networks can be used in inventory control, manufacturing and other applications in a warehouse.

h) *Location-based services*: Ad hoc networks when integrated with location-based information provide useful location-based services. *GPS (Global Positioning System), a* satellite-based radio navigation system, is a very effective tool in highway for exchanging traffic information and vehicle location. This system when integrated with ad hoc networks can be very effective for tracking vehicles and location discovery.

Among various applications of mobile ad hoc networks, LBS coupled with mobile advertising is becoming most promising for additional revenue generation. In the next section, we will focus on LBS solely.

2 LOCATION-BASED SERVICES
According to Schiller J. Location based services can be defined as "services that integrate a mobile device's location or position with other information so as to provide added value to a user." [12]. A LBS is a mobile information service that extends spatial and temporal information

processing capability to end users via Internet and wireless communications [7, 10,12,16,18,20,21,23]. Location-based services are the key enabler for a plethora of applications across different domains ranging from tracking and navigation systems to directory services, entertainment to emergency services, and various mobile commerce applications [14, 16]. Although, several different types of LBS are possible, we present some of the examples and their requirements in Table 2 that can be easily deployed using mobile ad hoc networks.

The potential for location-based services is evident from powerful and ubiquitous wireless devices that are growing in popularity [8]. E911 requirements for wireless cellular networks in the US have also enabled location-based functionalities using variations of triangulation, GPS, and cell-ID technologies Many surveys predict billions of dollars in revenues for mobile advertising [10,13,23]. Network operators can take full advantage of mobile ad hoc networks and are well positioned to take up a significant percent of this advertising revenue as they negotiate deals with content providers. In this paper, we discuss business models for mobile ad hoc networks, specifically location-based services and mobile advertising.

Location-Based Services	Applications
Information/directory services	• Dynamic yellow pages that automatically informs consumer of location of nearest hospitals, restaurants, shopping malls and theatre, and ATM • Nearest parking lot, drug store or gas station
Tracking and navigation services	• Tracking of children, locating lost pets, • Locating friends in a particular area • Tracking stolen vehicles, asset tracking • Dynamic navigational guidance • Voice-enabled route description
Emergency services	• Roadside assistance • Search and rescue missions • Police and fire response • Emergency medical ambulance, E911
Location-based advertising	• Wireless coupon presentation, targeted & customized ads • Marketing promotions and alerts • Customer notification and identification in the neighborhood store

Table 2. Location-based Services

2.1 Mobile Ad Hoc Networks and LBS

With the massive adoption of smart phones equipped with WiFi, wireless Internet and location technologies such as GPS, a wide range of applications enabled by mobile ad hoc networks has become a reality. Location-based services can take full advantage of mobile ad hoc networks as they can be easily deployed in places where there is no network infrastructure. Using mobile ad hoc network, a user can easily see which friends are in their neighborhood; exchange instant messages in a large stadium while watching a game, engage in multi-player games and launch various social networking applications. Hence, various location-based services along with social networking applications are being increasingly deployed [23].

Besides overcoming technological and ethical barriers, marketing location-based services has been a challenge to operators. One major challenge has been that the new applications were relying on technology that was very slow in penetrating the market. As a result, the industry now often talks about location as a means to enable services as opposed to location-based services.

3. MOBILE ADVERTISING

The mobile advertising market is poised for tremendous growth as it continues to utilize mobile ad hoc networks, location-based services along with some appealing features of portable devices and smart phones [1,3,8].

I. Portability: The devices are small in size and fit into the pocket.
II. Personalization and Instant Access: The devices are associated with the identity of the user and the applications are personalized based on the user input. The mobile devices also receive instant access from their users most of the time.
III. Mobility and Wireless Internet Connectivity: Most of the mobile devices will have Internet connectivity via wireless links.
IV. Location-aware: Most of the devices will have some built-in navigational systems like GPS.
V. Context-aware: Many applications running on the device are context-aware. For example, in case of search, the advertisements will be displayed based on user's preferences [13,19,20].

All these appealing features coupled with enterprise applications have increased the adoption rate of these devices and as a result of this; usages of these devices are growing rapidly. Hence it makes business sense to use the mobile devices as another platform for advertising, which can be customized based on the user's profile and preferences. The users can also select the types of advertisements they would like to receive on their mobile devices. This information would be saved on a server and the advertisement would be sent accordingly. During the first time set up the subscriber will have the opportunity to provide his preferences for the type of ads to be received and displayed on his mobile device. For example, after the customer logs in, he would be prompted with choices in several areas of interests such as specific type of restaurants, currently screened movies, nearby shopping malls, and so on. Several different types of advertising services and their infrastructure requirements are presented in Table 3.

Advertising Services	Description	Specific Challenges/Issues
Click-to-Call	A call back number will be provided which is hyperlinked by the advertisers. When clicked, customer will be directly connected to that particular advertiser [17]	Network carriers may charge for the call, which is an additional cost to the user.
Short Message Service (SMS) or Multimedia Messaging Service (MMS)	The advertisement provides an option where the customer can enter his number to avail the services of the advertiser. Once the number is entered, the advertiser would send the text information through SMS and any video clippings through MMS to view [10, 17].	For MMS, bandwidth requirements are significantly higher and smart phones are necessary. 3G/4G network infrastructure will be quite appropriate. MMS will cost more than SMS as network operators may charge higher.
Voice calls	Advertiser will call the customer and inform him of his services. Customer will receive a call, which will let him know about all the promotions and sales that are going on while he is shopping in a mall.	User may not want to receive calls from advertisers.
Location Finder	The advertisement provides the link to find their facility. Suppose the customer gets an advertisement from the nearest Pizza Hut, it would also provide a link to find their location.	Accurately determining the location can be challenging.

Table 3: Types of Advertising Services and Infrastructure Requirements

3.1. Emerging Business Models

With the deployment of mobile ad hoc networks, location-based services along with a wide range of social networking applications are on the rise. Various location-based services and social networking and start-ups are well positioned to capture substantial market share in this *nascent* and *growing* market of creating *smart environments* powered by mobile ad hoc networks. Hence there are substantial opportunities for revenue generation in this emerging market of mobile advertising powered by mobile ad hoc networks and various wireless technologies. Mobility, social networking and location-based services combined with context-aware advertising create opportunities for targeted marketing and revenue generation [1,2,3,9,11,13,16,22].

As the competition of voice-based mobile services becomes stiff, the network operators seek to increase their market share and generate additional revenue through data services including LBS. The network service providers play an important role and actively or passively shape the business landscape.
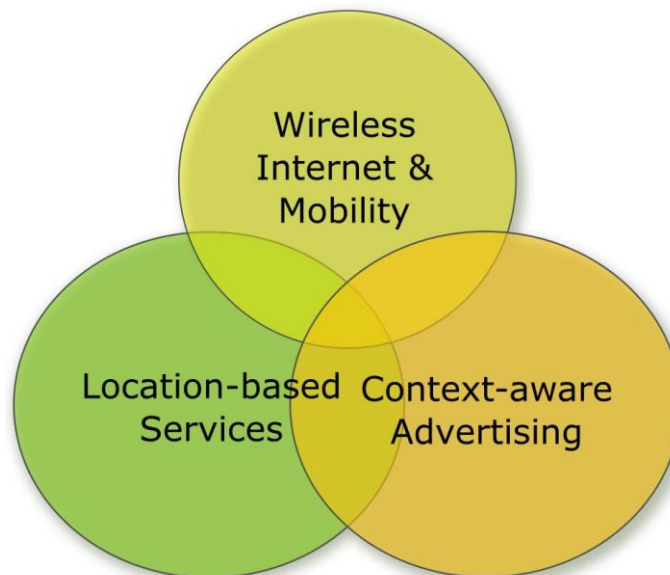
Fig 1: Convergence of Mobility, Location-based Services and Advertising

There are three types of business scenarios that evolve around the role of network service providers [9,23]. These are network service provider-dependant business scenario, network service provider-assisted business scenario, and network service provider-independent business scenario.

In the network service provider-dependant (network-centric) business scenario, the network service providers play a dominant role and collect the most of the revenue. These typically generate and own the location data and use this to provide LBS to the consumers. The network service providers use their infrastructure and marketing channel to give access to LBS thereby providing services to a select customer base. In this scenario, the network service providers keep the major portion of the revenue.

In the network service provider–assisted (service-centric) business scenario, the network service provider does not necessarily own or control the applications and manage the services. The data will be available to service providers either on a given rate or for free. The network service providers act as transporters of the data and voice. They get revenue by selling airtime usage or charging the volume of data packets transmitted.

In the network service provider-independent (device-centric) business scenario, the data does not necessarily reside on the network service provider's network. The data may come from an independent vendor. As an example, some location-based service providers may use GPS technology and use the data to provide various services to the consumers. The network service providers generate revenue by transporting the LBS data through their network. The location-based service providers may charge the consumers on pay-per-use basis or on a monthly subscription fee.

Another interesting scenario that is gaining popularity is that of OEMs creating development platforms for independent application developers. The developers sell their applications through OEM's authorized stores and share revenue with OEMs and not WSPs. A good example will be Apple's iPhone. Apple has created the Apps store that allowed independent developers to create applications and share revenues with Apple and not the WSPs.

There are a few business models that appear to be viable in the current scenario [1,11,15,21,22]. These business models have three major aspects, namely revenue sharing; hosting and advertising [7]. In revenue sharing, every party that contributes to a service takes a part of the service fee, which is based on a partnership agreement. The hosting service provider charges a fee to the LBS service provider for infrastructure and management of these services. The fee can be a fixed monthly rate or based on the usage of data. Advertising in the mobile environment is still in its infancy and is not well established like the World Wide Web. The various business models are presented in Table 4.

Model Type	Key Features	Comments
Network-carrier centric model	This business scenario is prevalent in the voice-based mobile services segment of the market where the network operator is the dominant player. The carrier typically makes specific and even exclusive deals with various service providers and content aggregators that can publish services on its network. The user pays the carrier for network access and various services, and the carrier pays service providers for delivered services. Within this model, all four gate keeping roles are owned and controlled by the network carrier, which performs the actor activities of platform operator, portal provider and network operator. Network operators can avail themselves of several recurring revenue streams under this scenario. LBS subscription fees (e.g., a monthly fee for roadside assistance service), per-usage fees for information services (e.g., directions to a restaurant), air time usage, and packet-usage for the service transaction.	This is a very common business model and in this scenario, the carrier provides end-to-end value. Revenue would also result from revenue sharing, typically a transaction fee, with marketers of location-based services. Some location-based services may also be provided free or part of a bundled service. This would help reduce churn, create differentiation, and increase market penetration.

Device centric model	The device manufacturer offers several integrated services. The device can, but does not have to be, offered bundled with a subscription from a carrier. Services can also be sent to the device via an online or offline mediator like a PC. The device manufacturer offers mobile versions of services it already provides on other devices. The device centric model is a model where the main service platform is incorporated in, or tied together with, the mobile device.	In this model, the device plays an active role in storing profile information and aggregating and brokering services. Billing and charging can be controlled by the network operator or by a third-party billing provider. A real-life example is the Apple iPhone. The introduction of this mobile device presented the industry with a number of business model innovations.
Service centric Model	The user connects via the wireless service provider and selects various services on demand. A meta-platform operator (e.g. Open Social) may allow developers to link services together and exchange information. Services outside of the meta-platform are available, but not connected. In this model, profiles are managed by individual service providers and possibly exchanged between them. The user buys those services either directly from the service provider or through the wireless network provider or through a service aggregator.	There can be a direct billing relationship if a service developer charges the end user for access to the service, but other models, e.g. advertisement-based, are possible. It is based on plans and models surrounding Google's Open Social initiative regarding open API's for social networking applications.

Table 4: A Comparison of Several Business Models

4. CHALLENGES AHEAD

Although a great deal of work has been done in mobile ad hoc networks, commercial deployment along with large scale adoption has not taken place. However, with increasing use of location-based services and mobile advertising, the wireless network carriers have to figure out how they can take advantage of mobile ad hoc networks to generate additional revenue. This poses several major challenges.

4.1 Integration of Mobile Ad Hoc Networks with Cellular Networks

One of the biggest problems that the network carriers are facing now is how to integrate cellular networks and IP-based networks with ad hoc networks. Issues like interoperability, bandwidth optimization, extended coverage, personalization are also a great concern.

4.2 Pricing and Adoption of Value-added Services

With the integration of mobile ad hoc networks with other networks like 4G, WiMax, LTE, etc, various new valued-added services can be deployed. For example, location information coupled with social networking applications creates new market opportunities for all the players in the mobile communication ecosystem. However, adoption of these services will greatly depend upon the pricing model and ease-of use.

With the increasing adoption of LBS, a large number of mobile devices that utilize geo-location are emerging. It is expected that in future, network carriers will be able to generate additional revenue by using the location data to develop customized services for their clients on

demand. By utilizing location-based mobile marketing, businesses can quickly reach out to mobile ad hoc communities and provide various social networking services. In addition, augmented reality coupled with LBS has lots of potential for mobile marketing in the social networking scenarios.

4.3 Security and Privacy

Many users are concerned about how personal data along with location information will be used to protect their privacy. As more of these personalized services are being deployed, it will be extremely important to protect consumers' privacy. In addition, security in wireless networks is always a great concern for consumers. Wide-scale adoption of these services will greatly depend on how well network operators protect privacy and security of consumers. Hence, service providers will have to make these services on a permission-based opt-in basis and assure the consumers that appropriate measures are taken to protect privacy and security.

5. CONCLUSIONS

With the rapid deployment of mobile ad hoc networks coupled with WiFi and various wireless networking technologies such as LTE, WiMax, etc, mobile advertising and other location-based services have the potential for becoming more pervasive and resulting in significant revenue for wireless network carriers, various service providers, and applications developers. Many new and interesting applications are being developed such as the design and development of large-scale real-time location-based information system [4,5,20,21]. However, many challenges must be addressed including developing acceptable standards for mobile advertising, addressing security and privacy concerns, providing end users with value-added location-based services along with performance guarantee.

REFERENCES
[1]   Pieter Ballon, Nils Walravens, Antonietta Spedalieri, Claudio Venezia, An Advertisement-based Platform Business Model for Mobile Operators, 12th International ICIN Conference, Bordeaux, France, October 2008.
[2]   Banjo, D. Charging for Mobile Content, Content Networking in the Mobile Internet, edited by S. Dixit & T. Wu, Wiley, 2004.
[3]   Rebecca Bulander, Michael Decker, Gunther Schiefer Bernhard Koelmel, Comparison of Different Approaches for Mobile Advertising, *Proceedings of the 2005 Second IEEE International Workshop on Mobile Commerce and Services (WMCS'05).*
[4]   Barbeau, S.J., Labrador, M.A., Winters, P.L., Perez, R., and Georggi, N.L. A general architecture in support of interactive, multimedia, location-based mobile applications, *IEEE Communications Magazine* 44, 11 (November 2006) 156–163.
[5]   Cai, Y and Xu, T, Design, analysis, and implementation of a large-scale real-time location-based information sharing system, *MobiSys '08: Proceeding of the 6th international conference on Mobile systems, applications, and services*, June 2008.
[6]   S. Dhar, MANET: Applications, Issues and Challenges for the Future, *International Journal of Business Data Communications and Networking,* Vol. 1, No. 2*,* April - June 2005, Pages 66-92.
[7]   Hazas M., Scott J., and Krumm J. Location-aware computing comes of age. *IEEE Computer* 37, 2 (Feb. 2004).
[8]   Hirsch, F., Kemp, J., Ilkka, J., Mobile Web Services: Architecture and Implementation, John Wiley & Sons, 2006.

[9]   Issel, K and Mrozik, J, A mobile data service framework and its business models, MoMM '08: *Proceedings of the 6th ACM International Conference on Advances in Mobile Computing and Multimedia,* November 2008, Pages 322-325.

[10]   Iris A. Junglas, I, A., Watson, R, T., Location-based services, *Communications of the ACM*, Volume 51 Issue 3, March 2008, Pages 65-69.

[11]   Koutsopoulou, M., Kaloxylos, A., Alonistioti, A., Merakos, L., A platform for charging, billing, and accounting in future mobile networks, *Computer Communications,* 30 (2007), pages 516–526.

[12]   Küpper Axel, Location-Based Services: Fundamentals and Operation, John Wiley & Sons, 2005.

[13]   Mohamed, Y.H., Gao, Z.J., & Shim, S. Wireless advertising's challenges and opportunities, IEEE Computer 36, 5 (May 2003) 30-37.

[14]   Mokbel Mohamed F and Levandoski, Justin J., Toward context and preference-aware location-based services, *Proceedings of the Eighth ACM International Workshop on Data Engineering for Wireless and Mobile Access,* June 2009, pages 25-32.

[15]   Qusay H. Mahmoud, Provisioning Context-aware Advertisements to Wireless Mobile Users, *IEEE International Conference on Multimedia and Expo*, July 2006, Pages 669-672.

[16]   Mohapatra, D and Suma, S.B, Survey of location based wireless services**,** *IEEE International Conference on Personal Wireless Communications (ICPWC)* (Jan. 2005) 358-362.

[17]   Ralph, D and Graham, P., MMS: Technologies, Usage and Business Models, Wiley, 2004.

[18]   Rao B. and L. Minakakis, Evolution of Mobile Location-based Services, *Communications of the ACM* 46, 12 (December 2003) 61-65.

[19]   Rashid O, Coulton, P, Edwards, R, Providing location based information/advertising for existing mobile phone users, *Personal and Ubiquitous Computing*, Volume 12, Number 1, January, 2008.

[20]   Schiller, J. H., and Voisard, A., Location-based services, Morgan Kaufmann Publishers, 2004.

[21]   Steed, A. Supporting mobile applications with real-time visualisation of GPS availability. *In Proceedings of Mobile HCI*. Springer, 2004.

[22]   Stefan Stieglitz, Christoph Fuchß, Christoph Lattemann, A Business Model for Mobile Ad-Hoc Communities, 2009 IEEE Conference on Commerce and Enterprise Computing (2009).

[23]   Unni R and Harmon R, Location-Based Services: Models for Strategy Development in M- Commerce, *Intl. Conf. on Technology Management for Reshaping the World*, PICMET '03 (July 2003), 416- 424.

ECONOMIES OF STYLE: A NEW SOURCE OF COMPETITIVE ADVANTAGE IN E-COMMERCE

Paul Licker, Ph. D.
School of Business Administration
Oakland University
Rochester, MI
licker@oakland.edu
1-248-370-2432

ABSTRACT

This paper introduces economies of style that afford competitive advantage. Machine production enabled economies of scale, Programming that production enabled economies of scope. Networking has enabled economies of style, effectively reducing the marginal cost of implementing the nth business model to near zero. Economies of style effect an almost limitless variation in relationships among suppliers and buyers. Subsequent control of the value network enables an organization to extract profits from the value network even without producing anything. Those profits generalize the concept of outsourcing both among suppliers, where the concept is well developed, as well as among buyers, where the concept does not yet exist. Examples from today's businesses indicate where economies of style may be achieved as e-commerce continues to grow, especially in light of the success of Web 2.0 and social computing.

Keywords: E-commerce, economy of style, Web 2.0, value network, business model

INTRODUCTION

This paper introduces a new kind of production economy to complement the venerable economy of scale and the newer economy of scope [3]. Economy of scale is well known. Through mechanization and some automation of production processes, an economy of scale is achieved when per unit production costs drop as the number of items produced rises. By substituting fixed capital investment in mechanized processes for the relatively variable cost of labor, firms economize on production costs at some level of production. These economies are achieved because the machines' initial investment is low when amortized over a large number of items produced, whereas labor costs per item are the same whether one or one thousand items are produced (ignoring learning time and operator-induced efficiencies). Economies of scale imply that scaling up production provides gains to the owner. However, the relatively high variable cost of each mechanical production process makes switching to another product, or even a variation on that product, relatively expensive, since this entails new equipment and new processes. For example, machinery that produces a 4" common nail might not be easily altered to produce, say, a 6" common nail or a 1" finishing nail.

Being able to reprogram that machinery to produce a second product or a new version of the original product can lower costs and increase profits when a second product or a new version of the original product is wanted by a market. To the extent that the machinery can quickly be

reprogrammed, economies of scope can be achieved. These economies effectively lower the fixed cost by amortizing *across products*, just as economies of scale amortize the cost of a single set of equipment *across units*. Such reprograming becomes even less expensive if the equipment is computerized, since the nature of software is "weightless" and once the expense of the software has been sunk, the more manufacturable variations or new designs that can be created, the lower the per-product cost, thus increasing profits. Machinery that can produce nails of any desired size by changing a software parameter can help a firm create products as markets change or opportunities arise without expensive retooling.

It is possible to expand on the idea of product scope to business process scope. The following definition of economy of scope is offered by InvestorWords.com: "The situation that arises when the cost of performing multiple business functions simultaneously proves more efficient than performing each business function independently"[2]. This definition is an important generalization because it helps us understand the importance of flexibility not only managing a production process, but in managing business processes in general. Economies of scale provide increased profits from more efficient production processes. Economies of scope boost profits by introducing a portfolio of such processes. However, economies of scope do not correlate with changes in a business's structure, even though they might have a major effect on the business's operations. Stated another way, business operations may change, but the business model stays fixed. The nail business is still the nail business. The business model is in effect a fixed cost that cannot be altered by reprogramming the design process.

Clearly there are potential profits to be found in new business models. A firm that produces nails might find that other businesses are (perhaps temporarily) more profitable. While it might not be possible for that firm to produce, say, glass bottles, it may be profitable to move into the marketing of fasteners or the home improvement business. In other words, being able to change the business model in an agile and effective manner might help a business survive when markets, suppliers, and legislation changes.

A way to see this is to watch firms such as IBM, which began as a clock manufacturer that moved into computers and then developed systems and applications, and now operates as a diversified consulting and outsourcing company. Each business model change afforded IBM the flexibility to test and potentially prosper in new markets. Of course there was always a danger of failure. But that danger did not come from the flexibility; such dangers reside in misdirected action afforded by flexibility.

Another example is Google, to which we will return later. For now, it is useful to note that Google does not make its money from charging for search; instead it has developed a large stream of businesses and services all of which are intended to provide a platform for advertising. Google has become a world-wide information medium that sells ad space. What has made this possible is not the search engine, although this is, of course, interesting in its own right. What sets Google apart has been its ability to turn its users into an additional source of revenue for its advertisers, by varying its business models with regularity.

This ability to turn elements of a firm's supply/buy chain into unpaid employees is, in effect, lowering the firm's costs for a given price for its products or services (regardless of any

economies of scale or scope), thereby increasing profits. We refer to this profit as a "tax" (the reason for this term will be apparent later), namely the difference between the cost of value supplied by a firm and the price it charges for this value. A firm such as Amazon.com has many ways of making money, but one of the most interesting from this point of view is that it effectively lowers its marketing costs (i.e., the cost of making a particular title attractive to a customer) by having other customers write reviews of the item (a book, for example). In the past, firms like Amazon relied on professional reviewers to create book reviews. Depending on the reviews, sales might go up or down. Successful booksellers might have some idea of which books were attractive, but the reviews certainly helped dispel customer uncertainty about the books before they bought them. Now, however, Amazon has many unpaid book reviewers whose efforts are displayed at very low cost to Amazon, thus improving the odds that any visitor to the website will in fact buy a book. This reduces Amazon's marketing costs and thus increases its profit.

ECONOMIES OF STYLE

In fact, this ability in effect to "renegotiate" a relationship such as buyer into unpaid reviewing employee constitutes a change to a business model. A business model [1] consists of three elements: what a firm will do to make money (i.e., what it will produce and sell, the revenue model), what a firm needs to do to make money (i.e., the capabilities and hence the resources it requires, the expense model) and how it will perform for its stakeholders (the value model). Changes to business plans are normally expensive and difficult to effect. They require a new revenue model or a new expense model or a new value model. Each change takes time and resources and introduces risk. In effect, each business model has a high variable cost.

However, it is possible to reduce that high variable cost through networking, by accessing and taking advantage of existing business models. If we conceive of revenue models as specifying relationships among production processes, expense models as relationships among supply and demand processes (which result in the costs of producing a product and selling it) and value models as relationships among strategic interests, then we can see the high variable costs arising from these relationships. Such relationships, say those between employer and employee or between a firm and its suppliers, are expensive and take time to develop. Porter and Millar [4] spoke about switching costs in the supply chain and such switching costs exist in other networks, too. If these costs can be lowered through some sort of flexibility, then increased profits will arise. We refer to the ability to lower these relationship costs as "economies of style"[1]. The ability to become any one of a number of businesses at relatively low cost results from economies of style. A company can then be defined as a dynamic, rather than fixed, network of relationships that are managed from a central focal point (or focal firm, to use Porter's terminology) rather than specific machinery, products, markets, or processes. Where the network is fixed, no economies of style are being achieved. Where the network is very flexible and adaptable to the firm's advantage, then economies of style are in play.

---

[1] The term "economy of style" has another common and somewhat contradictory meaning: the sparing use of stylistic forms in writing. We might, for example, speak of "Hemmingway's 'economy of style'" to refer to his direct way of writing. This is unrelated to our use of the term to mean gains in profit from the ability to adopt multiple business models.

MAKING MONEY FROM ECONOMIES OF STYLE

Taking advantage of flexibility in relationships can raise revenues or lower expenses through economies of style. The ability to raise revenues or lower expenses results in increased profits. A firm can therefore extract net income from controlling its network of relationships. In effect, an organization charges for the use of a relationship it controls and those using it are willing to pay this charge because they are achieving savings or increased revenues. Focusing on relationships rather than products, markets, and processes, we can use the term "tax" to define the net income a focal organization receives from others who use relationships the focal firm controls, just as a government taxes transactions of its residents' businesses.

The strategy, then, is to fine tune links in a dynamic network to mutual advantage of the members of that network and the focal firm controlling it. And it is a network. The term "supply chain" refers only to the flow of supplies into a firm as though this were a linear set of interlinking processes. In fact, most complex firms have multiple suppliers, some of whom supply one another, and multiple buyers, some of whom, as we have shown above, become suppliers (of labor, i.e., reviewers, to Amazon.com). Economies of style therefore result in taxes on a firm's value network.

Figure 1. F in a value network of Suppliers and Buyers
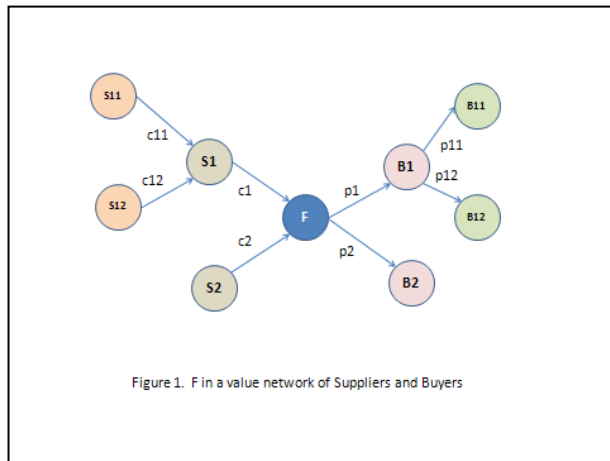
Consider the example in figure 1. In this simple network model, focal firm F has two suppliers (S1 and S2). In turn S1 has two suppliers (S11 and S12). F has two buyers (B1 and B2) and B1 has two buyers (B11 and B12). We are not, at the moment, concerned with suppliers' other buyers and buyers' other suppliers, although we will return to this situation later.

Using Porter's reasoning, F will try to *lock in* S1 and S2 at a fixed cost (c1 and c2 respectively) for a desired level of quality. Furthermore, F will want to lock in B1 and B2 similarly at fixed and favorable prices (p1 and p2). F1 will also try to *lock out* competitors, again a situation we will return to later. S1has costs it must sustain from its suppliers S11 and S12 (c11 and c12 respectively). In traditional supply chain bargaining, F will try to beat S1 down in terms of prices, but to the extent that F is helpless in helping S1 control its costs (c11 and c12), F is going to have to be content with S1's business model's calculations of c1 from c11 and c12. Stated simply, there is only so
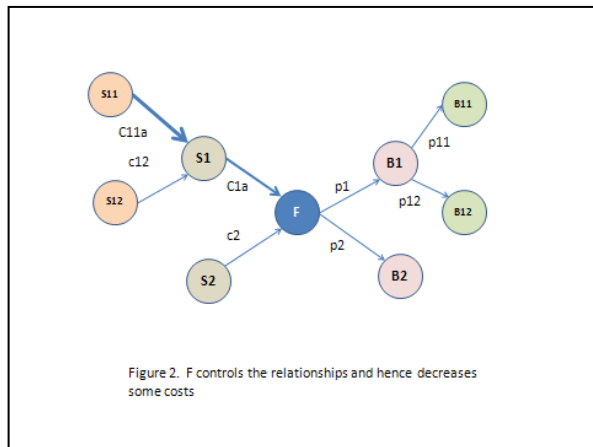
Figure 2. F controls the relationships and hence decreases some costs

much economizing F can expect from S1 before S1 find the S1-F relationship unprofitable. But if F can help S1 control costs and reduce, say, c11 to c11a<C11, then F can expect S1 to lower c1

to c1a<c1 (see figure 2). The gain (up to c1-c1a) is the "tax" that S1 should be willing to pay to remain in the relationship with F as supplier.

Figure 3. F taxes relationships on supplier
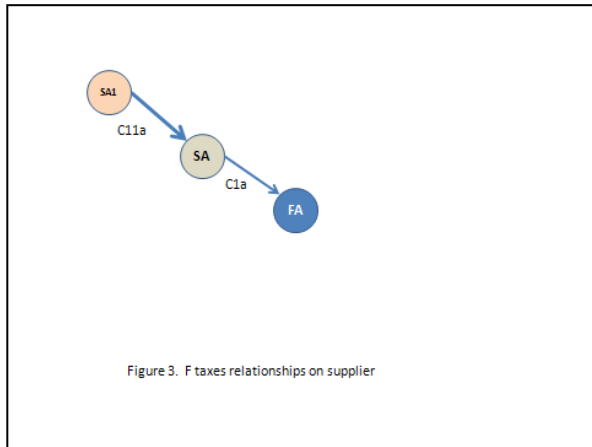
An example of this occurs in the automobile industry (see Figure 3).   One well-known manufacturer FA works with a paint suppliers. Because FA wants very high quality paint, it could submit its desires to potential suppliers and give the contract to the lowest bidder. However, FA recognizes that there is risk in this contract. Should the paint not adhere properly, for instance, FA will have inferior vehicles that won't sell well. Of course this risk can be built into the contract, but the potential loss to FA is such that few paint suppliers could withstand the loss. In effect, the risk of bad paint drives up the quality requirements beyond profitable levels for most suppliers as well as FA. However, suppose FA shares proprietary metallurgical information with a particular supplier SA, selected for its willingness to have its own paint production processes managed jointly with FA (at a price, of course). This can reduce the potential problems arising from poor adhesion, lowering SA's costs (c11 to c11a in Figure 2; this could be costs for SA's supplies, labor, or other costs) and hence both lowering the cost to FA (c1 to c1a) as well as decreasing the risk. SA pays the "tax" (up to c1-c1a) but receives the benefits (c11-c11a). The mechanism for this is F's control of the SA1-SA relationship, achieving an economy of style.

The model works equally well on the "buyer" side (see Figure 4). Consider another focal firm FM, who manufactures MRI units. Suppose FM knows not only how to manufacture MRI machines, but also how to operate them most efficiently and effectively. Suppose FM has a buyer, a clinic such as BA, who provides MRI services to its patients. BA pays F $p1 for the MRI machine and charges a patient such as BA1 $p11 for an MRI scan. BA sets this price based partly on what BA1 will pay and partly on what it costs to operate the MRI (insurance schemes complicate the picture, but do not invalidate it). If BA is willing to let FM "interfere" with its relationship with BA1 (for example, advising BA on how many hours a week the MRI machine can be used, how the machine should be operated or how patients should be introduced to the machine – especially patients prone to claustrophobia, etc.), then BA may be able to reduce its other costs or charge BA1 more for the service, in effect creating more profit for BA. Let's consider the effect of the latter. BA might therefore receive the benefit of p11a – p11. In turn BA pays a "tax" (up to p1a-p1) to FM but receives the benefit of p11a-p11. The mechanism for this is F's contribution to the BA-BA1 relationship, achieving an economy of style on the buyer side.
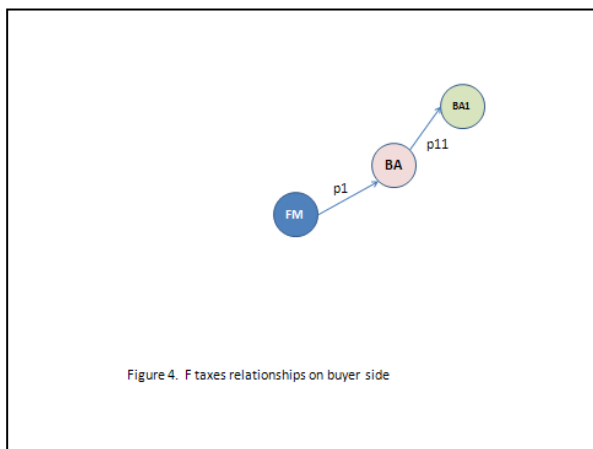
Figure 4. F taxes relationships on buyer side

The model is completely general. Whether higher prices can be charged to customers or costs from suppliers can be lowered the focal firm receives "taxes" from controlling or contributing to relationships. While such benefits are available right now with direct suppliers and buyers (indeed, that is why firms have purchasing and sales departments), achieving these benefits from higher-order relationships (suppliers' suppliers, buyers' buyers, suppliers' suppliers' suppliers, buyers' suppliers, etc.) is difficult to impossible without the networking capabilities offered through e-commerce.

EXAMPLES

In addition to the example of Amazon.com, which uses change-of-role from buyer to supplier (of reviews)[2], four other well-known examples show the power of the economy of style concept. The first two do not quite achieve the quality of "economy of style" because the number of business models is quite limited, but the third and fourth have almost limitless economies of style.

The first is On-Star™, a product available on General Motors cars (represented in Figure 5). At first glance, this product seems like GPS on steroids, providing up and down links to vehicles to give directions to the nearest video store, restaurant or gasoline station. But On-Star™ is both simpler and more sophisticated than that. It is simpler, because what it literally does is to build the customer (buyer) more permanently and more accessibly into GM's value network. Without On-Star, buyers are available, but only through relatively passive media such as mail or sometimes email or phone. With On-Star™, buyers are available whenever they are in their cars, with a priority link to GM. It is more sophisticated because of the applications that are available through On-Star™. While the above-noted GPS apps, along with a 24/7 voice link for emergencies, provide relatively
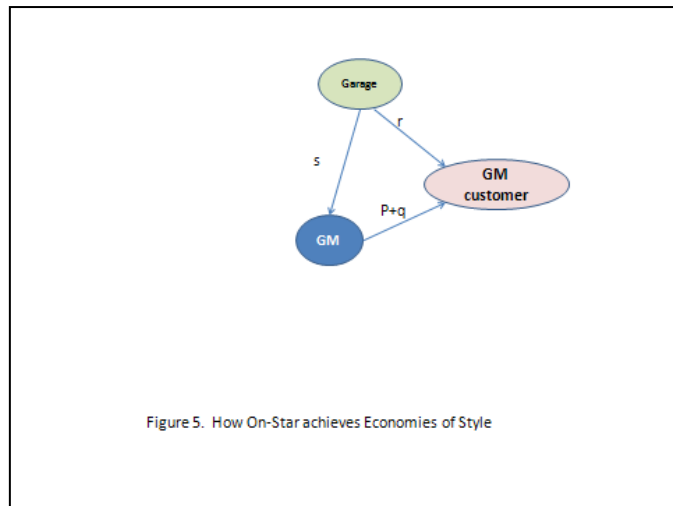
Figure 5. How On-Star achieves Economies of Style

unprofitable access to customers, the ability to upload information about the car itself to GM -- including its location, its mechanical and electrical health, and its history – provides GM with ways to tune the customer's purchasing habits. Suppose, for instance, that sensors in the car determine that the engine is potentially malfunctioning. Normally the driver might notice a red light on the dashboard signaling something – these signals are so rare that drivers may have to consult a manual to figure out what is wrong – and then take action at some time in the future. However, On-Star™ has the capability to direct the driver to the nearest open garage handling GM vehicles – or even to initiate the negotiation of a favorable fee for the owner at the garage.

---

[2] We are not going to dwell here on Amazon's "other" business, which is a form of cloud computing. While "Elastic Compute Cloud" is a byproduct of Amazon's extremely flexible business model, it is only one new business, almost the equivalent of a hosting subsidiary. Of course the Internet facilitates this particular "economy of style"

In effect, On-Star™ can create or modify a relationship between a GM buyer and a supplier (of repair services) to the buyer. In return for this, GM can charge, in advance, for the service, much like an extended warranty or breakdown service. But in this case it is GM and not someone else that determines when, how, where and who is in this relationship. GM gets to charge an additional fee, say q, for the service to the GM customer. And the garage can charge a fee, denoted as r, to the customer, but less than the customer might have paid for emergency service at a random station. It is likely, too, that the garage might pay a finder's fee, denoted as s, to GM. While all this is possible without On-Star™, it is far less likely to work smoothly and far less responsive to individual needs. The costs involved are much higher and typically customers go elsewhere to buy emergency services unrelated to their cars. This example is only ONE business model, of course; GM could use On-Star™ to create a myriad business models but hasn't yet chosen to do so.

The second example is Dell computers, whose Virtual Integration model is well known and need not be reviewed here [1]. Virtual Integration (Figure 6), in essence, is simply first-order economy of style from the supplier side to the buyer side, a network-assisted brokerage function. Virtual Integration is therefore, from an economy of style viewpoint, not very interesting other than the fact that Dell has plainly built a new business model that it can vary at will (building different kinds of computers for individual buyers) and none of the business models include Dell's manufacturing, delivery, setup, or service. In fact, each business model includes Dell only as a high-powered broker. Dell taxes the "virtual" relationship it builds between supplier (say a printer manufacturer P) and a buyer B – this is the essence of brokerage. That it can do this for a wide variety of suppliers and types of buyers is the secret of Dell's success. Dell pockets the difference between p1 and c1. The customer B gets a printer appropriate for the computer that Dell has sold, so receives increased value (i.e., B does not have to worry about whether or not this printer will work with this computer). The printer supplier gets a payment, possibly less than c1 because Dell has found the customer for P. Dell can do this for a variety of suppliers to B, but virtual integration is not a portfolio of business models, at least not yet.
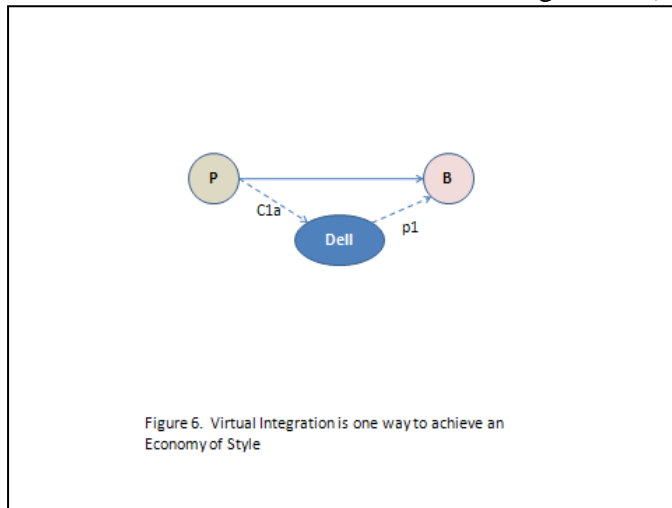
Figure 6. Virtual Integration is one way to achieve an Economy of Style

The third example is Google, mentioned above. By integrating immense sources of target data (books, maps, websites), Google has the ability to market not only information to consumers, but information **about** consumers to marketers. As with Amazon and GM, this is a small set of variations in business models. But even more interesting is the basic algorithm, which uses the popularity of search terms to enhance the value of searches to individuals who are searching. This changes customers into information suppliers to other customers who are searching for information. In addition, because Google has a great deal of information about what is being searched for, Google can sell that information, too. Figure 7 illustrates the simple reality that

Google sells customer B (an advertiser or a product developer) information about its customers'

Figure 7. How Google Sells customers to other customers
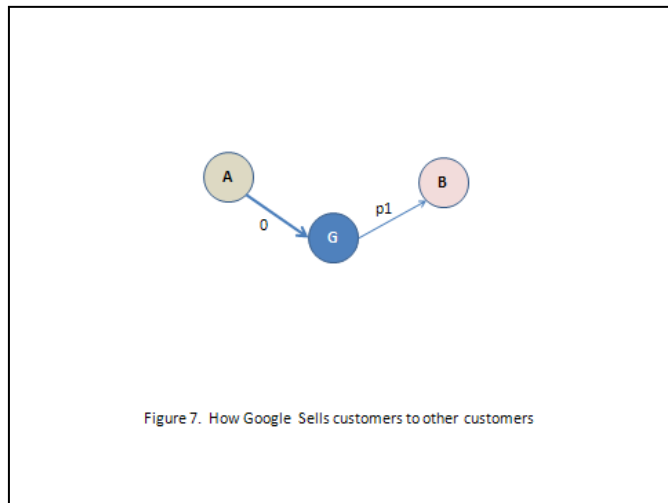
search terms (such as A seeking information on Paris Hilton). It is the activity of A (and millions of people like A) that generates the data that Google (G) sells to B. Because search terms vary across the entirety of human intellect, interest, and curiosity, Google can become a supplier of information to any industry. Interestingly, Google can do more than this, because it has the information already and through analytics can in turn manufacture other (intellectual) products, too. Google receives a "tax" of p1 from B for a service (information about what A wants to see) which is essentially provided, free of charge (the cost of operating the search is offset by the sale of advertising) by A.

For the fourth example, we turn to Facebook and the hundreds of other social networking websites. The product of these websites is, of course, its users. In that sense, a social networking website achieves both economies of scale through the use of software, built on a network that literally manufactures its product, i.e., its users. Economies of scope appear because the social networking apps, which make the products more attractive (and in effect take the original product – users – and make new products and services out of them), are easy to produce once the original platform is in place. Where social networking achieves economies of

Figure 8. Social Networking Sells Customers' Existence to Other Customers
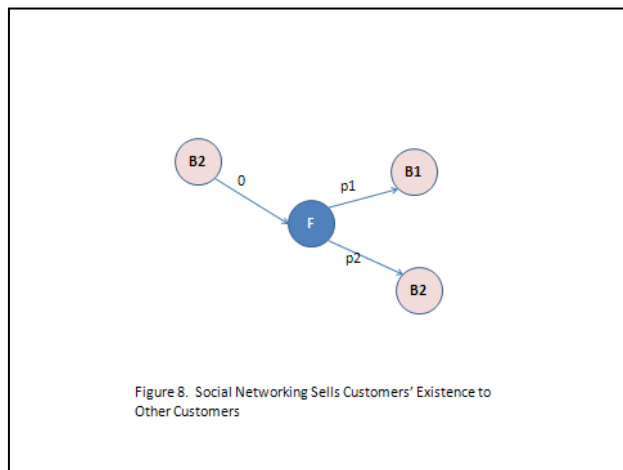
style is twofold. First, the nature of networking is establishing links. Users do this themselves, effectively bringing the site new revenue (through additional eyes viewing advertising), "taxing" every link constructed by users. Figure 8 illustrates how by bringing in B2 as a "friend", B1 supplies F with more product to sell (i.e., B2) at very low cost (again the physical costs are offset by advertising). Second, and more generally, such websites have the capability of employing users to construct apps by sequencing and combining existing facilities, tweaking capabilities and making suggestions, providing advice and counseling to others about how to use the website. While the hard coding might still be done by employees who have to be paid and while the software has to be hosted somewhere in real space, what the site "is" becomes a matter of almost infinitely variable individual choice. A well-constructed and well-thought-out social networking site can continually improve its service because users largely determine what services they want – and make them! The improved service, if perceived as higher quality, will become more popular, attract more users, and demand higher fees from advertisers. In that regard, social networking is the realization of the ultimate

virtual business: always ready to become something else that someone wants. Of course this is an ideal. In reality, there are limited combinations of apps, user-designed procedures, and data domains possible. Most users are not desirous or capable of creating, propagating, promoting or instituting new procedures or apps. And human choice is a very fickle activity, subject to forces far outside the realm of our simple models.

CONCLUSION

By describing a set of phenomena as "economies of style" we hope to invigorate the discussion about the effects of e-commerce on business. Economies of style are not new, but they have been incredibly difficult to achieve before the networking of businesses. And while most examples are still rather primitive, limited to a small set of possible business models, as Web 2.0 and beyond develop and as social networking becomes a more prevalent way of marketing, it is likely that businesses will begin to see the advantages of such chameleon-like behavior as changing business models to suit even individual customers. Economies of scale rose naturally from the mechanization of production around the time of the (first) industrial revolution. Economies of scope are a natural byproduct of the programming of production design through the use of digital computers, making product variation inexpensive – although simple scope economies came along with programmable looms in the nineteenth century. Economies of style are a logical next step in the evolution of value of the networked enterprise through mechanization and programming of business models stemming from the ability to affect multiple, innovative relationships in business processes.

There are many unanswered but researchable questions here. Comparing economies of style to economies of scale, the development of the basic economics needs mathematical foundations. For example, a firm would need to be able to calculate the savings from multiple business models just as they now can easily do this with economies of scale and scope. Because the idea of a virtual business is no longer so difficult to imagine, the idea of virtual business models, easily modified and fit to environmental circumstances becomes more feasible. How a firm can integrate this idea into a strategic plan becomes a challenge. Firms that compete on price or quality now may discover that individual variations of their businesses are competing on multiple, sometimes conflicting, qualities. When the flexibility of the product is so enhanced (because the product consists of network access to other customers), the possibility of niche markets of one (the ultimate mass customization) becomes enhanced. Currently web-hosted businesses, such as social networks, are still selling their services as social meeting places, but others are seeing them as what they are, infinitely flexible **market** places, in which the market is not for physical products, but for users of services and information. This presents the possibility of organizations that exist in the minds of the users rather than anywhere in real time or space, a true virtual business. And we are not talking "Second Life" here; we are talking about "real" virtual businesses, but businesses that program their value networks to extract profit from supplying services that arise from the relations in the network itself.

REFERENCES

[1] Applegate, L., R. Austin, and W. McFarlan. *Corporate Information Strategy* and *Management. (7th Ed.).* Boston: McGraw-Hill,2007.

[2] InvestorWords.com. http://www.investorwords.com/1654/economy_of_scope.html. Accessed Nov. 24, 2010.

[3] Panzar, J. and R. Willig. "Economies of Scope," *American Economic Review*, 71(2), 71(2), pp. 268-272, 1981.

[4] Porter, M. and V. Millar. How Information Gives you Competitive Advantage. *Harvard Business Review*. July/August Vol. 63, No. 4, pp. 149-160, 1985.

MANAGING E-MAIL COMMUNICATION PERFORMANCE

Wenhong Luo

Accountancy and Information Systems

Villanova School of Business

Villanova University

Villanova, PA 19085

wenhong.luo@villanova.edu

610-519-5592

ABSTRACT

E-mail has become the preferred means of communication for many companies to reach and interact with existing and potential customers. The rise of E-mail as a marketing communications tool is propelled by a number of its advantages over traditional media. First, the low costs and fast speed of sending and receiving email have made it possible for companies to interact with customer more frequently than ever. Second, companies can personalize marketing messages to suit the needs of individual customers. Finally, companies can effectively track and measure the responses of customers with the help of customer relationship management systems.

As companies rely more and more on E-mails to interact with customers, effective management of E-mail campaign performance has become more important. While the response rate is an obvious measure of E-mail campaign performance, E-mail campaigns have different objectives and thus need different performance measures. Take the example of a bank. E-mails may be used to send online bank statements, alert customers about login failures, promote new mortgage rates or credit cards. Clearly, response rate is an important performance measure for some campaigns but other campaigns may not even require a customer response. Furthermore, the performance of an E-mail campaign can be affected by previous and other ongoing E-mail campaigns. For instance, in order to ensure high response rate of a credit card promotion, a company may send multiple E-mails for the same promotion. This approach may cause some customers to drop from the mailing list, which would negatively affect future campaigns. Therefore, companies should manage all E-mail campaigns as a program.

The purpose of this research in progress is to develop a comprehensive framework for identifying measures of e-mail campaign performance following the response process model. A case study is used to illustrate the applicability of the framework.

OPPORTUNITIES AND CHALLENGES OF SOCIAL NETWORKING AS BUSINESS APPLICATIONS

Andrew Sweeney, Villanova University, 800 Lancaster Avenue, Villanova, PA 19085, USA
Q B. Chung, Villanova University, 800 Lancaster Avenue, Villanova, PA 19085, USA

ABSTRACT

Businesses seek to gain from the growing popularity of social networking websites, such as Facebook, LinkedIn, and Twitter. The proposed study will provide an overview of the social networking landscape. We seek to gain a better understanding of how businesses might utilize this growing medium by delving specifically into Facebook's development, platform, and growing popularity. We will discuss implementation practices of businesses currently utilizing the online social networking phenomenon in different capacities. Additionally, we will explore the challenges faced by companies in implementing a framework for social networking utilization including: boundary concerns, compliance, customer dynamics, and profit maximization.

Keywords: Social networking, Facebook, Business models

INTRODUCTION

The concept of "social networking websites" was coined sometime in the mid 2000's in response to a plethora of new websites sprouting up in the United States. In its most literal sense, the term describes the idea behind the sites, which provide an online forum for social gathering, idea sharing, communication, and the ability to connect and network with others. In their original form, social networking websites tunneled their respective focuses on particular groups: members of a given school, social groups, and particular age groups. In its evolution through the first decade of the new millennium, social networking has created an infrastructure from which to build future internet growth, reap the benefits of flowing advertising dollars, and the industry rivals the profitability possibilities of the most successful growth companies.

The first social networking website "Thefacebook" was founded by Harvard University student Mark Zuckerberg in 2004. Zuckerberg was studying psychology as an undergraduate student at Harvard in Cambridge, MA when the idea was born from his past forays in computer programming and image uploading. Zuckerberg had previously launched "Facemash" while a sophomore at Harvard, which was a forum for Harvard students to rate each other's attractiveness by using pirated images stolen from Harvard's identification database. Additionally, he launched "Coursematch" which "allowed users to view people taking their degree" to share ideas and communicate. The evolution of Zuckerberg's ideas came to a culmination in February 2004 when he launched "Thefacebook," an online database of Harvard students. The site was a new technological perspective on the seemingly archaic student directories given out by high schools, and universities previously.

The "Thefacebook" site gained significant popularity in its first months, becoming widely endorsed by Harvard students and establishing a significant user database. Zuckerberg was soon joined by early co-creators, Eduardo Saverin, Dustin Moskovitz, and Chris Hughes. In a matter of a few months, "Thefacebook" was "extended to Stanford and Yale where, like Harvard, it was widely accepted." Ultimately, "Thefacebook" would become a "national student network phenomenon" and Zuckerberg and Moskovitz would drop out of Harvard to run the business full time. "Thefacebook" would undergo changes in the next year, renaming itself "Facebook" in August 2005 upon purchasing the domain name "facebook.com" for a reported $200,000 and, thus, the modern Facebook was born.

Facebook and other social networking sites were fostered from the technology bubble of the late 1990's and early 2000's. The influx of ideas and funding to the internet revolution, E-commerce, and cloud computing provided the necessary infrastructure for these websites to thrive in the mid 2000's. While Zuckerberg and Facebook were the traditional "first movers" in the social networking model, many competitors soon followed his lead.

Early on, MySpace was Facebook's main competitor. Unlike Facebook, it did not feature rigid membership requirements and its members did not need to belong to particular organizations or colleges within the United States.2 However, as a result of Facebook's changing platform and growing popularity, MySpace has since lost ground on the social networking behemoth. Currently, it features far less members than Facebook and nearly half of its visitors are age 35 and older. The resulting perception publicly is that the site has evolved for a more mature audience, and typically serves as a popular forum for musical bands, local businesses, and corporations.

LinkedIn is quickly becoming Facebook's strongest competitor. The social networking site was founded by Silicon Valley investor Reid Hoffman in 2003 on the predilection to help working professionals benefit from posting professional information and resumes in an online forum. The website has since evolved to become an interactive social networking site, but still with a more career-minded format and structure. It has experienced robust growth in recent years amongst adults, working professionals, and recruiting agencies. In fact, LinkedIn now boasts over 60 million members. Additionally, "visitors to the site have jumped 31% from last year [2009] to 17.6 million in February [2010]." The company has differentiated itself by what is perceived as a "safer" forum for social networking. As Fortune's Jessi Hempel puts it: "Facebook is for fun. Tweets have a short shelf life. If you're serious about managing your career, the only social site that really matters is LinkedIn." [14]

Another emerging rival is the microblogging website, Twitter. The "Tweets" referenced above describe the short, 140 words-or-less messages and blogged comments of Twitter users worldwide. Twitter embodies the most cutting-edge social networking phenomena and its growing popularity is staggering. The site "grew a stunning 752 percent from 2007 to 2008 and turned its 500,000 unique monthly visitors in January 2008 into more than 4.4 million by December." According to ComScore, "Twitter is the fastest growing major website in the United States with 17 million registered users…" In addition to traditional social media users, it has successfully attracted unlikely users such as NASA and print media outlets like the Washington

Post. Additionally, according to the Global Language Monitor, the word "Twitter" was the top word in 2009.

Despite the fierce competition from a growing number of competitors, Facebook sits atop the social networking revolution. Facebook features more than 500 million active users, of which 50% log into Facebook any given day. An average user is extremely well connected with 130 friends and 80 community pages, groups and events. The site boasts 124 million visitors per month. In 2009, it surpassed Google.com as the most visited website.

ANALYSIS OF FACEBOOK

Originally, the Facebook website's simple design and clear-cut functionality allowed the targeted audience and amount of users to grow quickly. The user interface included an individual's "profile" that highlighted simple contact information, personal information, and even favorite quotes and movies. The profile page allowed users to interact via a traditional messaging system, "wall posts" or real-time blogs, and through customizable status updates and event invitations. The profile continued to evolve in the late 2000's to allow for more "friend-to-friend" interaction with fictional E-gift giving, the popular photo galleries, and enhanced messaging features like "poking."

The increased traffic has allowed the site to garner additional private funding and evolve from solely a social networking mechanism to a for-profit website. With this shift, the site has become a capable E-commerce forum and mecca of internet business activity. It has done so by expanding its original target market of young adults and defined social groups to an even broader audience. In fact, Facebook is now a global platform, with 70 translations now available on the site, with about 70% of members residing outside the United States. Additionally, two-thirds of comScore's U.S. Top 100 websites and half of comScore's Global Top 100 websites have integrated with Facebook.

As a result of increased popularity and exposure, advertising dollars have increasingly helped profitability. In August 2006, Facebook signed a three year deal with Microsoft to "provide and sell ads on their site in return for a revenue split." In July 2010, Facebook arranged an agreement with Apple to "give away 10 million free iTunes samplers to Facebook users" furthering entrenching the for profit mentality of the social networking website. Facebook underwent another key change to its platforms in 2007 when it opened its software platform to media and application developers, enabling them to "create widgets, mashups, tools and projected based around Facebook." As author Arielle Emmett notes, "It was a very innovative step to let anyone have access to the social graph and leverage it anywhere." [6] This open architecture mindset should only help the website grow and adapt to different user bases, and in the long-term, aid profitability.

Facebook's evolution to an open architecture platform reshaped the demographics of its members. In the open architecture format there are less restrictive membership qualifications, unlike the original student-only mindset. This has opened the doors for large corporations and local small business owners to create profile pages to better interact and communicate with customers. Companies using the Facebook platform can take advantage of a variety of

techniques to generate increased brand awareness, improved customer service, and larger revenues. We will briefly highlight different ways in which the Facebook platform is being used by businesses and discuss the benefits of conducting business on the social networking website.

Nielsen

Nielsen and Facebook have joined forces with joint venture "Brandlift" which serves to "prove both online and social media's worth as a brand advertising medium." The new company will help with the valuation and commoditization of social networking data and brand awareness as that has proved to be difficult at times. Social networking data is ever-changing and difficult to interpret and, therefore, difficult to monetize.

As Trevor Johnson, head of strategy and planning EMEA at Facebook notes:

> "The internet has always struggled to get brand money as it has always been about direct response because of measurement. We hope this will convince brands to move money from other branding media, such as TV, into online."

Nielsen's early tests with corporations on Facebook have proved to be successful. In a simple campaign with Starbucks to sell more muffins, 94% uplift in purchase intent was measured amongst consumers on the campaign's date. In sum, results from Nielsen studies have shown social networking branding Return on Investment "can be equal if not greater than other forms of media."

Comedy Central

Comedy Central is an excellent example of traditional media outlet's emerging acceptance of working with social networking websites. Initially, traditional outlets were reluctant to partner with social networking websites for fear of cannibalism of their own content. If advertising dollars were a main source of revenue for a television programming and broadcast company, the thought of driving traffic to other websites, or partnering with other content providers proved threatening. However, NBC has changed their tune with certain programming like Comedy Central's "The Daily Show with Jon Stewart." Erik Flanagan, executive vice president of digital media at MTV Networks Entertainment Group notes the pitfalls of neglecting social networking with today's programming, "We weren't capturing [people's] reaction in terms of our Internet properties, and we weren't putting our content where they were talking about our shows."

In using social media, NBC/Viacom is increasing awareness of its shows and allowing for increased chatter concerning its programming. The open-mindedness ultimately allows for better visibility and brand awareness, which are integral to driving TV viewership and generating advertising revenue for NBC's internet and TV programming. Flanagan adds, "We love the fact that other people can look at our content elsewhere and yet we have 10,000 doors back to our own site. It's almost like we've established a mini-embassy on other people's soil."

Johnson & Johnson

Johnson & Johnson is a stalwart of traditional consumer marketing and juggernaut in the consumer staples market segment. They have generated some of the globe's most powerful brands and products under the umbrellas of various subsidiaries. Marc Monseau, Director of Johnson & Johnson's media relations group highlighted a few notable changes to J&J's recent embrace of social networking websites and social media. He notes, "For the past three years, my colleagues and I in the Johnson & Johnson corporate communication group have been taking steps to get the company more involved with the social web." [24] The company began its focus on social media with Margaret Gurowitz's history blog, Kilmer House. Additionally, they feature a corporate blog-JNJBTW and a Youtube channel dedicated to personal health. Also, J&J has developed a corporate twitter account for important product updates and investor communications.

Johnson and Johnson also developed a traditional Facebook page as few other businesses have since done. The page features a company profile and link to their corporate website. Additionally, J&J uses the innovative social networking friend-to -friend concept to generate a dedicated internal "Facebook." The site features employee profiles and contact information. It allows employees to be grouped by department easily. Ultimately, Monseau highlights that Johnson & Johnson recognizes that "social media projects…shouldn't be one-off initiatives or campaigns, but should instead represent the beginning of something longer term and more substantial, where both parties can benefit." [24]

BENEFITS OF SOCIAL NETWORKING

As the internet has become a popular mode of commerce in the global marketplace, companies have had to increasingly change their tactic on retailing to customers. The internet and social networking sites specifically have made transactional exchange online far quicker and more targeted. The implications of social networking websites have aided E-Commerce most significantly by allowing an increased flow of information from company to consumer and vice versa.

The amount of website traffic that social networking websites generate is astounding. In his article, writer Steve Cross notes that in February of 2009 social networking usage exceeded Web-based email usage for the first time. He quotes Sophos senior technology consultant Graham Cluley; "…social network sites are 'just email, only quicker.'" [3] As we noted above, Facebook alone has become the most oft visited website on the web today. Not only are individual users and businesses increasingly representing themselves on social networking websites, but the frequency with which individuals visit is increasing. Cross adds that, in the United Kingdom, 50% of the internet users are predicted to visit social networks at least once a month by 2013. Additionally, office workers spend at least 30 minutes a day on social networking sites. The evidence suggests that if business can effectively integrate within the social networking framework, there will be ample opportunity to interact.

There are various guides and "how-to's" online and in articles highlighting the most effective ways for companies to brand and transact through social networking. Perhaps the most

important piece of information worth noting is that businesses must understand their audience. Author Mary Beth Popp highlights important first steps for a business to interact within social media including: defining a clear approach, having a clear measurement of success, and defining your message clearly and simply. [33] Carmi Levy adds in her "how-to" article for businesses that corporations must seek to build a broad base of support. She adds, "Commitment to social media must come from all levels of the organization, including C-level executives." [21] Both Popp and Levy note, corporations must have constant upkeep and commitment to the maintenance of social media interaction. As Levy states, "policy is useless unless someone is driving it…" so if a dedicated full-time position is not justifiable, "build a part-time job description for the role…"

Marketing

Inherent in a company's ability to effectively transact business through social media are the efforts of the company to market in this arena. Marketing may be the single-best opportunity for companies seeking to gain benefits from the social networking revolution as the evidence shows a growing and diverse user base of the websites is in place. Additionally, consumers' willingness to share preferences, feedback, and demographic information proves to be a fertile ground for company marketers and researchers.

Using the theory and framework developed by Psychologist Kenneth Burke, authors Casteleyn, Mottart, and Rutten describe in "How to use Facebook in your market research" how Burke's philosophy explains that "Facebook data [can] be seen as a crystal ball for future consumer intentions." [2] Burke's theory seeks to understand human motives and interaction and its underlying premise is that guilt plays a significant role in human communication. The authors seek to extrapolate Burke's research into the modern social networking realm where they attempt to understand interaction among Facebook groups, Facebook "wall" posts, and group discussion boards on the site. An interesting case study revealed a communication thread established by blogging beer drinkers surrounding the Dutch beer brand, Heineken. They found that as Heineken was being judged by Dutch beer drinkers, an unlikely competitor emerged, Duvel. This brand would typically not be identified as a direct competitor to Heineken because it is a strong pale ale as opposed to Heineken, a traditional lager. Additionally, as the thread continued, Belgium beer drinkers participated, and the rivalry between the two countries and the low acceptance level of the Dutch Beer became evident. The authors note, this information could prove to be invaluable to market researchers and companies alike, as it could "deepen [their] analysis of the negative image of Heineken in Belgium, or can incorporate such unforeseen competitors as strong pale ale in quantitative research."

A more traditional benefit of a company's use of social networking is the ability to enhance their brand. Author Christine Perey proposes that "brands and retailers collaborate with partners in the mobile ecosystem using a three prong strategy that focuses on engagement using advertising and sponsorships, permitting mobile community participants to become brand advocates, and on transactions." Perey notes that companies face a significant challenge as "they need to find partners with expertise in business enabling systems…technical enabling components…and security and brand protection services." [30] However, by better working together with the mobile community, corporations can best identify opportunity and measure success. If the

proper infrastructure is in place, companies can better segment their market by consumer preference, geographical location, and best align consumer needs with the brand image. Perey proposes specific tactics in each avenue including the purchase of advertising space on mobile phones and internet pages, the ability for particularly sanctioned customers to actively promote a given brand among the mobile user community, and the importance of emphasizing transactional money exchange within the mobile internet community.

Networking

In its most simplistic definition, social networking websites seek to connect individuals and groups together based on commonalities and interests. Success in this forum is integral to a company's health as a widely growing portion of any given customer base is using this framework to interact. Companies seek to benefit from their increased visibility and presence with current customers and prospects on the sites.

Author Ray Poynter points out in Facebook: the future of networking with customers, that companies' interaction with customers is beginning to change. He notes that traditional market research is now quicker and cheaper because of online interviewing. Poynter explains, "Facebook polling is a new way to find quick answers to simple questions." In addition to the speed with which market research can be conducted, the efficacy of the research is changing. Poynter states, "Traditional research is based on a paradigm where everything is designed before the research begins; something that often means the research struggles to fit with the realities of customers' views and experiences." [34] In the modern day realm of networking with customers, real- time views and sentiments are relayed between both customer and company in a candid, open format.

Social Networking sites have continued to become more personal and interactive among different users. Arielle Emmett notes in Networking News that the long-term viability of social networking sites is still in question, but "readers will engage with each other and share stories…that is a given." [6] The dynamics of traditional networking have been taken to the online forum where "the internet has gone through a shift from people who get information to people who get to each other." Companies are poised to take advantage of this by becoming members of the community and accessing centers of influence in the diverse user base, or becoming centers of influence themselves. Emmett suggests that successful social networking companies have already established a "presence on Facebook through advertising and branded pages that give users the opportunity to join 'fan communities' accessing news, photos, features, quizzes, and blogs, while sharing them among Facebook friends." One of the most effective ways for companies to increase their "social graph" or web of connections is to have content recommended between friends: what Emmett calls the "social filtering of news." An excellent example of broad based networking using social media was Barack Obama's campaign for President. President Obama raised close to $1 billion using "text messages, networks, wikis, and blogs to connect with legions of far-flung volunteers and supporters."

CHALLENGES TO SOCIAL NETWORKING

Social Networking websites are clearly going to be part of businesses' focus going forward. What is unclear is how exactly that relationship will look in the future. There are many hurdles to be overcome for the two worlds to co-exist effectively. Social Networking and corporations find themselves working together in two main contexts: one where employees of a corporation visit social networking sites on the company network, the other being companies' usage of the social networking websites themselves as a means for E-Commerce. In both instances many issues arise, including compliance and privacy issues between the different parties, information technology operational concerns, and the uncertain nuances of the integration of sales and marketing in the social media context.

Compliance, Privacy, Ethics

There are obvious concerns with the acceptance of social media in the corporate world, one of the largest being the sacrifice in productivity amongst workers. The idea is that if a given corporation allows its workers to use social networking websites, just ten minutes a day per employee will have significant exponential effects on outputs across an organization. Others argue that for corporations to most effectively attract and retain talent, and to truly accept the changing paradigm in social media, allowing employees' usage is a crucial first step There is a risk that by "completely denying staff access to their favourite social networking site, organization will drive their employees to find a way around the ban," causing a possible host of other issues.

In addition to productivity concerns, privacy issues may arise for corporations allowing internal access to social networking websites. There are significant implications in regards to privacy for the users of the websites. An employee's private life or individual information could be exposed to other employees at the firm or a rival firm. Not only may this alienate or distract those involved, it could prove a costly exercise for the corporation to sort out. Additionally, this may prove to be a legal matter if it involves slandering, discrimination, or a serious disagreement among employees. Ultimately, in making situations like this more transparent, the company runs the risk of creating a complicated, costly situation which can damage company culture.

In addition to individual privacy issues, companies are at risk as well. Author Steve Cross notes that valuable company data could be leaked or lost due to the open environment. [3] One of the significant hurdles for companies is that "Proprietary information captured on these sites must be monitored to guard against loss of an organization's competitive advantage." In addition, employees using the sites could create user risk by disclosing inappropriate financial data, or by inadvertently allowing access to the company's internal operating systems. Due to the increasing amount of risks, Barnes & Barnes note that "Case law related to cyberspace matters is a nascent and developing body of knowledge." [1]

Operational Risks

In addition to privacy concerns, corporations and social media have a variety of operational issues to sort out to create a cohesive co-existence. Author Steve Cross highlights two

operational risk in his top five areas of concern for social networking in a business setting; malware and phishing scams by cyber criminals, and open access offered to threats due to "lax and outdated attitudes towards passwords." [3] In fact, in a February 2009 survey of IT professionals, 20% cited the threat from malware as the main reason for restricting access to social networking websites. An easy fix, Cross notes, is to "educate individuals to choose strong passwords to prevent cybercriminals taking over online accounts which could provide an entry point to the IT infrastructure." Nancy Dupre and Frederick Barnes note that proper infrastructure must be in place if social networking is to be accepted, adding that it is, "better to lock the barn door before the horse is stolen." [1]

An additional operational concern is a more traditional one; headcount. To regulate the growing list of social media websites frequented by employers is a costly and time-consuming endeavor. Cross notes currently there are over 900 websites classified as "social networking." [3] Monitoring and regulating these websites for employee use would require constant attention. It is likely this would require a dedicated staff of IT administrators, a respective dedicated management team and additional resources. Authors Barnes and Barnes note that this new department would need to be carefully designed. [1] Companies would need to build out "a group of diverse, knowledgeable persons" which might be difficult based on varying abilities within the IT space. Additionally, Cross notes, "granularity is key…" when regulating the ins and outs of employee use of social networking sites. [3] This would require significant labor hours, and possibly a dedicated "Records and Information Management" (RIM) as Barnes & Barnes suggest. Barnes & Barnes think this is the most effective manner to create and monitor social networking sites in the corporate arena; however, only 13% of survey respondent s utilizes a records management plan.

In addition to some of the infrastructure and traditional operational risks, there are some broader implications for use across an organization. Regardless of context, it would be helpful for companies to establish clearly defined guidelines for using social networking mediums. This would provide a level playing field for all users of the sites, and those within the company that are not as well versed in this technology. Barnes & Barnes suggest it might be helpful to establish a training manual to help employees with any transition to internal social networking sites. [1] Additionally, they recommend a guide that incorporates all aspects of use of the site. Ideas such as these will help to provide a backdrop for which companies can regulate usage of the various websites, or establish policy for the company's presence on those sites.

Integrating Sales and Marketing

Another challenge corporations face in the social networking framework is the integration of their sales and marketing efforts. Traditional exchange of goods in an in-person format has waned over recent years and through the popularity of the internet. With the increased acceptance of internet retailing and the growing popularity of social networking sites, traditional exchange of monies is changing.

Companies are increasingly present on social networking websites. They are now offering products specific to the new form of web usage, for example, offering "widgets that can be downloaded to a reader's personal browser or website to quickly link back to the 'mother site.'"

In addition, media companies are offering new content, web applications, and product updates through the social networking forum. This represents a shift from a branding and product perspective, but also from the traditional sales format and even the so-called traditional E-Commerce format which was a simple customer-company relationship.

The shift to social networking sales and marketing represents not only a shifting mindset and strategy for customers but also presents a new set of logistical challenges. Author Mary Beth Popp suggests that success in the social networking realm requires a high level commitment and time-consuming website maintenance and consistent communication. [33] This is necessary in order to build trust with customers. Simply put, this trust takes time to build. As Ted Janusz adds in his article, you want to be perceived as a vendor who seeks to become a "trusted member of the community…Not because you are trying to sell stuff." [18] Additionally, the level of commitment incurs its own set of costs as it increases company overhead in the form of actual web specialists and support staff to help with the new company infrastructure. In addition to commitment, Popp notes that a company must "decide upfront how [they'll] measure [their] success, whether by the number of followers, or contacts, length of conversations, or offline leads." Measurement of success seems to be a simple task, however, it is has proved difficult for companies to implement a clear definition for success in the social networking world, and once they do so, difficult to monetize. [33]

Barnes & Barnes suggestion to have a dedicated group (RIM) to build out and plan the internal site might help streamline any integration with broader company goals. [1] They note that a well-thought out plan can help avoid costly missteps. According to Barnes & Barnes, corporations must develop a clear and defined purpose. This will help get buy-in from the organization's stakeholder group and aid adoption of the site. Additionally, boundary lines must be defined in terms of usage: will it be internal only or external as well? Also, it might be help to develop a "medium for an organization's marketing and public relation activities….for both internally and externally based users." In addition, by having a dedicated RIM group you address above concerns over the level of commitment and can have a group responsible for mining data for market research, and measuring the company's level of success on social media.

DISCUSSION AND CONCLUSION

Various concerns continue to plague the relationship between social media and the business community. These concerns limit the complete acceptance of social networking websites by corporations and businesses, and limit their effectiveness. In a society where E-Commerce rules the day, it is imperative that companies and social networking sites find a cohesive and safe way to work together. Customization of the website-employee-company relationship is the answer.

In "Social Networking: the business case" Steve Cross argues that clearly defined rules of engagement are important to the relationship. He questions, "Granularity is key: if only certain aspects of a social networking service trouble you, can you disable access to those aspects alone?" The article is written in the context of controlling various features of social networking websites so that employees can safely access them at work. Cross suggests a myriad of options and measures to be taken to allow the safe interaction of company employees and social networking sites on the given company's hardware and systems. [3] However, his analysis only concerns this

one specific type of interaction between website-company-employee. The idea can be extrapolated to a broader context and take the relationship between social networking websites and companies, and the users and employees of both, to a new level.

We propose a third context in which corporations and social media could find themselves working cohesively in the future. There is a significant need for companies to build out their internal infrastructure surrounding internal directories, intranets, and company websites. Anecdotally, the technology is typically weak, and pales in comparison to privately funded and dedicated social networking websites like Facebook. Facebook could fill a necessary gap in the marketplace by targeting large corporations that would like to streamline internal operations, increase communication across the firm, and improve connectivity in a global marketplace. We propose that Facebook builds out a business plan to provide services to corporations and aids them in the construction of their respective intranets. This could come in the form of high-end consultation, or more likely, in the form of outright sales of the Facebook platform and capabilities to the corporation. Essentially, Facebook sells its "product" (the social networking platform), much as Microsoft sells its "product" (software).

This would eliminate many of the concerns companies face at the present time in accessing the social networking site on their respective systems. With a capable and stimulating internal Facebook, companies could theoretically limit the time lost on external social networking websites while employees are at work. The employees could be inspired to look up fellow colleagues, their biographies, the workings of other business groups, etc. The proposed plan would allow for more streamlined communication. This would help limit much of the difficulty global companies' face operationally. The new platform would be interactive and engaging and its use more easily encouraged. Even the simple task of culling through company directories for the right contact would become easier and more engaging as the company could be chunked up into various working groups, special teams, or departments, much in the same way Facebook uses its "groups" technology now.

There would be cursory benefits to a customized technology platform provided by Facebook as well. It would aid in issues like employee turnover. With a fully integrated internal network and the better flow of communication across a firm, no longer would one employee become indispensable. Typically, with this employee's departure, effectiveness of a given business group goes down, awareness of the new contacts is limited, and working with that group becomes cumbersome. With a dedicated business group online, it would be easier to see whom the replacement contacts are, and learn more about them and effectively network internally and efficiently complete tasks. Additionally, a simple week of vacation for some employees proves to be a challenge as their job function is so human capital intensive and their absence detrimental to a business group. Allowing storage of information on the individual's profile page, or message board, or maybe a dedicated section of "accessible public information" would ease any given transition. This would eliminate the issue of an employee's human capital and important information being buried in their respective mail.

Another benefit is mobility within a given firm. The ability to effectively network is integral to one's future career success. Inherently, this is why social networking has gained so much popularity amongst all age groups and personality types. More importantly, the ability to

network internally is vital to one's future career success at a given company. An effective and engaging internal Facebook would allow for ease of communication and ability to easily uncover future networking opportunities. A business group could post that they are having a birthday cake for a fellow employee at 3pm and it might be an employee you know from your past or from daily job tasks. Perhaps an executive at your firm shares an alma mater and you notice that on his/her profile updates that they will be attending the alumni event this weekend, or doing a recruiting activity and you can become an active participant. This information would not become readily available otherwise. A sense of pride can be imbued on an employee's page for given job tasks, business groups, and personality characteristics.

The biggest hurdle to the employer-employee-website relationship is the cyber-risk associated with the sites' usage. A customized internal social networking site would eliminate a lot of the risks. The site could be customizable per the corporation's rules and level of comfort with certain site features. While the proposed site would be operated on the company's network, it could prevent malware/spyware because it is contained and monitored by the company's compliance and dedicated tech groups. It would operate much like company email today, and be monitored by given managers and compliance. The presumed increased efficiency would be noticeable where employees no longer need a company's intranet for corporate updates, personal profiles and lookups, and a separate email for data storage and correspondence. One dedicated internal Facebook page could consist of an employee profile where corporate updates show up, where correspondence and exchange of information between employees is available, and where biographies, individual profiles, and an employee's important data can be easily stored, all in one place.

Another hurdle with the external social networking sites that would be eliminated is the increased associated costs with the monitoring of employees' usage of the websites and the associated corporate technology spent to design, install, and maintain internal technology and intranet pages, etc. An internal Facebook, designed by Facebook, would eliminate a lot of the design, monitoring, and up-keep necessary. If certain features are eliminated, customized, or tailored, it streamlines the monitoring process. In addition, as Facebook operates as a consultant, they could presumably assign their respective contacts to the up-keep and monitoring of the site. Ultimately, with a better, faster, more expertly designed site, the company will presumably be more efficient, its employees happier and better connected, and by outsourcing in this case, costs could be lowered.

The internal Facebook site might significant aid company culture as well. A better flow of communication and efficiency tends to snowball and lead to more streamlined job functions and ease of work functions. Well informed and effective employees tend to lead to happier employees and improved morale. Additionally, an internal Facebook might allow more of a given employee's personality to show through in a safe format. In a changing workplace, employees are generationally different than they were 50 years ago, and society is much more open and communicative then ever before. Allowing an employee to post a brief video from their last vacation, or a list of favorite places to go, or an informative article they read in the paper incorporates a sense of pride and personality for their individuality. Something as simple as looking for an extra player for an after-work pickup basketball game is easier in an open forum. These types of freedoms go hand-in-hand with modern-day workplace changes of social

networking usage, flexible work hours, and less of a separation between workplace life and personal life. Additionally, it might just reduce bureaucracy and make management more accessible and open in the long-run as well. This would help quell what is one of employees' biggest complaints and biggest hurdles to overcome in the workplace: bureaucracy.

From the Facebook perspective it is an excellent proposition. Facebook can eliminate their biggest problem which has been the monetization of the social networking phenomenon. They can capitalize off their expertise in website design and their ability to capture the needs and wants of social networkers in a new way. With the broader population having accepted the social networking forum as one for communication, and exchange of ideas and monies, it is natural that it would make its way into the corporate world.

With a new business line from Facebook dedicated to customizable and consultative platform sales, Facebook would effectively be diversifying their business away from just pure advertising revenue, which can prove to be fickle in different market conditions. As competition strengthens in the external social networking site arena, it would allow Facebook to have a broader earnings base as well, and become a "first-mover" in this new business venture. Additionally, it would not require much new technological investment or expertise on their part, as the sites would likely mirror the public Facebook's format. Customization might prove to be tricky and the additional headcount required to become a dedicated business line for large corporations would be two foreseeable obstacles.

Facebook would also seek to profit from such a venture. The fundamentals look appealing as the demand is there in that current employees use the external Facebook site and other social networking sites while at work at an alarming rate, and growing. Corporations would be eager to control and limit the content provided on those given sites, and limit the risks associated with their use. With a strong demand from a customer standpoint, and popularity of the venture supported by given corporations' employees, it is seemingly a healthy proposition for Facebook. The customer base of corporations is nearly limitless, as most companies today feature intranets, from the largest corporations to small business owners. Additionally, Facebook could charge a hefty consultancy fee as a one stop shop technology provider of web design, customization, and up-keep. The broader, external site Facebook offers might benefit as well as their presence in a large corporation's intranet might produce converts to their external site from other social networking site loyalists or prior non-users.

The proposed business direction would represent a shift in Facebook's past company culture and direction. Providing customizable platforms for corporations feels very "non-Facebook" as it caters to the corporate world and requires a young and hip company to answer to shareholders and powerful users. It would also represent larger overhead in terms of additional employees, a possible call center for questions etc. This would be a change for a company that has traditionally limited its overhead and operated very lean. Ultimately, Facebook would be increasing its costs, expanding its horizons, and ultimately, driving profitability and the first to truly monetize what has become the social networking craze.

REFERENCES

[1]   Barnes, Nancy Dupre, and Frederick R. Barnes. "EQUIPPING YOUR ORGANIZATION FOR THE SOCIAL NETWORKING GAME." *Information Management (15352897)* 43.6 (2009): 28-33. *EBSCO MegaFILE*. EBSCO. Web. 21 Mar. 2010.

[2]   Casteleyn, Jordi, André Mottart, and Kris Rutten. "How to use Facebook in your market research." *International Journal of Market Research* 51.4 (2009): 439-447. *EBSCO MegaFILE*. EBSCO. Web. 21 Mar. 2010.

[3]   Cross, Steve. "Social Networking: the business case." *Engineering & Technology*. 6 June 19 June 2009. www.theiet.org/magazine.

[4]   Dholakia, Utpal M., and Emily Durham. "One Café Chain's Facebook Experiment." *Harvard Business Review* 88.2 (2010): 26. *EBSCO MegaFILE*. EBSCO. Web. 21 Mar. 2010.

[5]   Dumenco, Simon. "Be honest: What's your real Twitter and Facebook ROI?." *Advertising Age* 81.2 (2010): 14. *EBSCO MegaFILE*. EBSCO. Web. 21 Mar. 2010.

[6]   Emmett, Arielle. "Networking News." *American Journalism Review* 30.6 (2008): 40-43. *EBSCO MegaFILE*. EBSCO. Web. 21 Mar. 2010.

[7]   "1. Facebook." *Fast Company* 143 (2010): 54-110. *EBSCO MegaFILE*. EBSCO. Web. 21 Mar. 2010.

[8]   Yadav, Sid. "Facebook-The Complete Biography." As visited on 07/25/10 at facebook.com on mashable on facebook. 2006.

[9]   "Facebook's friends :Social netowrking site's success is appealing to advertisers. " *Financial Times* 18 Mar. 2010,ABI/INFORM Global, ProQuest. Web. 21 Mar. 2010.

[10]   http://www.facebook.com/press/info.php?statistics

[11]   "Facebook Gets Down To Business." *BusinessWeek* 4127 (2009): 30. *EBSCO MegaFILE*. EBSCO. Web. 21 Mar. 2010.

[12]   Gaudin, Sharon. "Business Use of Twitter, Facebook Exploding." *Macworld* 27.2 (2010): 70. *EBSCO MegaFILE*. EBSCO. Web. 21 Mar. 2010.

[13]   Hathi, Sona. "How social networking increases collaboration at IBM. (Cover story)." *Strategic Communication Management* 14.1 (2009): 32-35. *EBSCO MegaFILE*. EBSCO. Web. 21 Mar. 2010.

[14]   Hempel, Jessi. "How LinkedIn Will Fire Up Your Career." *Fortune*. 03/25/2010.

[15]   Hempel, Jessi. "Web 2.0 Is So Over. Welcome to Web 3.0." *Fortune International (Europe)* 159.1 (2009): 27. *EBSCO MegaFILE*. EBSCO. Web. 21 Mar. 2010.

[16]   Holloway, David. "Facebook and Twitter, why bother?." *Australian MacWorld* 133 (2009): 032-033. *EBSCO MegaFILE*. EBSCO. Web. 21 Mar. 2010.

[17]   Jacobs, Daniel G. "SURVIVING THE SOCIAL EXPLOSION. (Cover story)." *Landscape Management* 48.12 (2009): 10-13. *EBSCO MegaFILE*. EBSCO. Web. 21 Mar. 2010.

[18]   Janusz, Ted. "Marketing on Social Networks: Twitter, MySpace and Facebook Demystified." *Key Words* 17.4 (2009): 124-125. *EBSCO MegaFILE*. EBSCO. Web. 21 Mar. 2010.

[19]   Kaplan, Andreas M., and Michael Haenlein. "The fairyland of Second Life: Virtual social worlds and how to use them." *Business Horizons* 52.6 (2009): 563-572. *EBSCO MegaFILE*. EBSCO. Web. 21 Mar. 2010.

[20] Keegan, Theresa. "Cashing in on the sweet sounds of Twitter, LinkedIn, Facebook.." *Hudson Valley Business Journal* 19.4 (2010): 5. *EBSCO MegaFILE*. EBSCO. Web. 21 Mar. 2010.

[21] Levy, Carmi. "Facebook, Blogs and Twitter: Social Media Wins the Race." *Certification Magazine* 11.3 (2009): 20-21. *EBSCO MegaFILE*. EBSCO. Web. 21 Mar. 2010.

[22] Luke, Kristen. "Marketing the New-fashioned Way: Connect with your target market through social networking sites." *Journal of Financial Planning* (2009): 18-19. *EBSCO MegaFILE*. EBSCO. Web. 21 Mar. 2010.

[23] McEleny, Charlotte. "Nielsen combines data with Facebook to assess branding." *New Media Age* (2010): 04. *EBSCO MegaFILE*. EBSCO. Web. 21 Mar. 2010.

[24] Monseau, Marc. "Social media initiatives help build relationships." *Medical Marketing & Media* 44.8 (2009): 34. *EBSCO MegaFILE*. EBSCO. Web. 21 Mar. 2010.

[25] "MySpace or yours?: Advertising and social networks." *Strategic Direction* 24.8 (2008): 15-18. *EBSCO MegaFILE*. EBSCO. Web. 21 Mar. 2010.

[26] Needleman, Sarah E. "Entrepreneurs Question Value of Social Media --- Marketing via Facebook, Twitter Yields Results for Some, Others Say It's Overrated; 'Hype Right Now Exceeds the Reality'. " *Wall Street Journal* 16 Mar. 2010, Eastern edition: ABI/INFORM Global, ProQuest. Web. 21 Mar. 2010.

[27] Noella, Lyne. "Ten Tips to Raise Your Business Visibility Through Social Networks." *Financial Executive* 25.9 (2009): 17. *EBSCO MegaFILE*. EBSCO. Web. 21 Mar. 2010.

[28] Orton-Jones, Charles. "How Facebook can be used as a serious professional tool." *Financial Management (14719185)* (2009): 21. *EBSCO MegaFILE*. EBSCO. Web. 21 Mar. 2010.

[29] Pattison, Kermit. "How to Market Your Business With Facebook." *New York Times* 12 Nov. 2009: 1. *EBSCO MegaFILE*. EBSCO. Web. 21 Mar. 2010.

[30] Perey, Christine. "MOBILE SOCIAL NETWORKING: THE BRAND AT PLAY IN THE CIRCLE OF FRIENDS WITH MOBILE COMMUNITIES REPRESENTING A STRONG OPPORTUNITY FOR BRANDS." *International Journal of Mobile Marketing* 3.2 (2008): 75-79. *EBSCO MegaFILE*. EBSCO. Web. 21 Mar. 2010.

[31] Phillips, Sarah. "A Brief History of Facebook." *Guardian.co.uk* as visited on 08/01/2010. July 25, 2007.

[32] Pollock, Jim. "Local businesses are fans of Facebook." *Des Moines Business Record* 28.5 (2010): 16-24. *EBSCO MegaFILE*. EBSCO. Web. 21 Mar. 2010.

[33] Popp, Mary Beth. "Social Networking is the New 24/7 Virtual Cocktail Hour." *Small Business/Social Media for Business.* P5. January 1,2010. "Special Report."

[34] Poynter, Ray. "Facebook: the future of networking with customers." *International Journal of Market Research* 50.1 (2008): 11-12. *EBSCO MegaFILE*. EBSCO. Web. 21 Mar. 2010.

[35] Schultz, Don E. "Social Call." *Marketing Management* July 2007: 10+. *EBSCO MegaFILE*. EBSCO. Web. 21 Mar. 2010.

[36] Shields, Mike. "Facebook Growing Pains." *MediaWeek* 19.8 (2009): 10-12. *EBSCO MegaFILE*. EBSCO. Web. 21 Mar. 2010.

[37] Skey, S. "A Next Step in Online Customer Engagement. " *Marketing News* 15 Mar. 2010: ABI/INFORM Global, ProQuest. Web. 21 Mar. 2010.

[38] "Social Networking is the New 24/7 Virtual Cocktail Hour." *Business Journal (Central New York)* 24.1 (2010): 5-9. *EBSCO MegaFILE*. EBSCO. Web. 21 Mar. 2010.

[39]   Talbert III, William D. "What's Now." *Associations Now* 4.12 (2008): 15. *EBSCO MegaFILE*. EBSCO. Web. 21 Mar. 2010.

[40]   "Tech Today, Profits Tomorrow." *MediaWeek* 19.9 (2009): 1. *EBSCO MegaFILE*. EBSCO. Web. 21 Mar. 2010.

[41]   Tikkanen, Henrikki, et al. "Exploring virtual worlds: success factors in virtual world marketing." *Management Decision* 47.8 (2009): 1357-1381. *EBSCO MegaFILE*. EBSCO. Web. 21 Mar. 2010.

[42]   Warr, Wendy A. "Social software: fun and games, or business tools?." *Journal of Information Science* 34.4 (2008): 591-604. *EBSCO MegaFILE*. EBSCO. Web. 21 Mar. 2010.

[43]   Wilson, J. "Social networking: the business case." *Engineering & Technology (17509637)* 4.10 (2009): 54-56. *EBSCO MegaFILE*. EBSCO. Web. 21 Mar. 2010.

[44]   Woods, Anne Dabrow. "Get social!: Build an online community with social networking." *American Journal of Nursing* (2010): 6-7. *EBSCO MegaFILE*. EBSCO. Web. 21 Mar. 2010.

THE IMPACT OF AGING AND ECONOMICS ON FUTURE PENSIONS IN HUNGARY

Laura Forker, University of Massachusetts Dartmouth, lforker@umassd.edu, DIS Dep't.,
University of Massachusetts Dartmouth, 285 Old Westport Rd, N. Dartmouth, MA 02747-2300

ABSTRACT

Hungary, the most indebted eastern member of the European Union, is currently in crisis. Hungary has a growing elder population, a large pool of retired workers (3 million in a country of 10 million people), a national debt estimated to be 80% of its gross domestic product (GDP), and pension obligations that exceed 10% of its GDP. Political pressures are currently rolling back previous pension reforms. Demographic and economic changes in Hungary are expected to put additional burdens on an already strained pension system.

Keywords: Pensions, East Central Europe, Policy implications

INTRODUCTION

Hungary, the most indebted eastern member of the European Union (EU), is currently in crisis. Hungary has a growing elder population, a large pool of retired workers (3 million in a country of 10 million people), a national debt estimated to be 80% of its gross domestic product (GDP), and pension obligations that exceed 10% of its GDP [5] [8]. Although not alone in its troubles, the twin burdens of servicing a huge national debt and providing a livable income to its growing elder population are squeezing pensions and endangering the fulfillment of other financial obligations. Will Hungary be able to meet its future pension commitments, given its current demographic and economic troubles?

THE ECONOMICS OF AGING

Hungary faces many economic challenges currently and in the not-too-distant future. The fertility rate in Hungary, like elsewhere in Eastern Europe, is below the replacement rate necessary to keep its population from falling over time. Its number of inhabitants, which fell by .2% in 2010, is predicted to decline even further (-0.4%) by 2025, a rate matched by the Czech Republic, Poland, and Romania [25]. In East Central Europe (ECE), only Bulgaria, which is forecasted to lose 1% of its population over the next 24 years, faces a greater demographic threat [25]. (East Central Europe consists of Bulgaria, the Czech Republic, Hungary, Poland, Romania, and the Slovak Republic.) More than 23% of Hungarians have inadequate income to meet basic household needs, in part due to the relatively large portion of Hungarians who do not work [17] [26]. Among Hungarians age 15 and older, only half were in the labor force in 2008, the lowest number in ECE [15]. While not as high as Italy and Japan where more than one fifth of the population is 65 or older, Hungary still has the largest aged population in ECE – 16% of its total [18]. The ratio of Hungarians older than 64 years to those in the working age population (ages 15-64) was 23.5 in 2009, the highest in ECE [18]. Only 14% of Hungarians age 60-64 continue working; more than 50% of Americans in this age group are gainfully employed [18]. Two thirds

of working Hungarians retire before the official retirement age of 62; 80% of retirees leave work voluntarily [1].

Hungary's population growth rate has been declining at a steady rate, from 0.7% during 1950-1960 to 0.4% between 1970-1980; it began to contract starting with the 1980-1990 period, is currently falling at –0.2%, and is predicted to continue its downward slope to -0.55% by 2050 [26]. While not as dramatic as the population decline forecast for Bulgaria (-1% by 2050), this population contraction means that fewer Hungarians will be providing the tax income and retirement savings necessary to support larger numbers of retirees. Aggregating data collected and projected by the U.S. Census Bureau, the young population (from 0-19 years old) will decline from 23.3% of the total population in 2000 to 18.8% of the residents in 2025, and further to 17.8% of the total inhabitants in 2050 [26]. Conversely, the elder population (those 65 and older) will increase from 14.5% of the total in 2000 to 21.4% of all inhabitants in 2025 and further to 29.4% of all residents in 2050 [265].

Employing U.S. Census numbers and classifications, the total dependency ratio will rise from .61 in 2000 to .67 in 2025 and to .89 in 2050. However, these numbers may understate the burden on working Hungarians given Hungary's minimum retirement age for full pension benefits (62 for both men and women) as well as the population's general preference for early retirement. Because of this preference, the actual retirement age in Hungary is currently 58 years old [5]. Recalculating the dependency ratios with age 60 as the starting point for retirement (a better reflection of reality), the total dependency ratio for 2025 would be .84 and for 2050 would be 1.18 given current trends. These numbers mean that by 2050, there are projected to be more economic dependents in Hungary than working-age people available to support them. The old aged dependency ratio for each year, recomputed with a more realistic retirement age of 60, show this ratio rising from .50 in 2025 to .79 in 2050. Given that the government-sponsored pay-as-you-go pension plan still accounts for the majority of pensioners' incomes, the burden on the working age group will be increasingly formidable if current projections are correct.

HUNGARY'S PENSION REFORM

Prior to 1997, Hungary had a state-sponsored, unfunded pension system. Pension payouts absorbed 9.7% of gross domestic product [13]. The socialist government (1949-1989) provided adequate incomes for the retired but failed to raise pension payments proportionately when wages and particularly prices increased in the 1970s and 1980s. In the early 1990s, when Hungarian businesses began to privatize, many companies trimmed their payrolls causing unemployment to jump. An eroding link between contributions and pension benefits discouraged some workers from participating in the labor force and encouraged others to retire early. High earnings and capital incomes for those with marketable skills and premature retirement for those with few or no marketable skills accelerated income inequalities. "Between 1989 and 1997 the Gini coefficient of net earnings jumped from 0.21 to 0.32…and similar processes can be observed concerning personal incomes." [22, p. 4].

These structural problems as well as the economic contraction the country experienced when transitioning from a command to a market economy prompted the country's political leaders to introduce a major pension reform in 1997. Policymakers believed that the replacement of part of

Hungarians' public pensions with private savings would help reduce long-term pension liabilities of the state. The 1997 reform replaced about ¼ of the public pension system with a private funds. The new system had two mandatory components: a public earnings-related pension financed on a pay-as-you-go basis (a defined benefit [DB] system) and a defined contribution (DC) private pension fund financed with compulsory contributions that were calculated as a percent of one's wages. Between the two mandatory pension payments, employers paid 28% of an employee's total wage and employees paid 9.5% [12]. Investments from the private accounts were designed to be paid out as a lifetime annuity upon retirement. Individuals retiring under the reformed system had to select retirement fund investments, pay for pension fund operating costs, and assume the risk of bond and stock market volatility in their DC and voluntary accounts [13]. Hungarians who began working after June 30, 1998 were required to join the new pension system; the system was optional for those already working prior to this date. Although the pension reforms shifted greater responsibility and risk exposure to individuals, pension benefits remained tax-free. A "third pillar," a voluntary private pension account with large tax incentives, was introduced in 1994 to encourage Hungarians to save additional money for retirement. In 2003 however, much of this new personal responsibility for retirement savings was shifted back to the public sector by adding a "13th month" to retirees' pension benefits. This additional payment was included in retirees' monthly annuities.

RECENT ROLLBACKS

The recent economic downturn has caused Hungary and other ECE countries to rack up record deficits and to deplete their budgets of available money targeted to pay current retirees. In many cases, these public debts have gone beyond the limit of three percent of GDP that the EU requires for a country to receive loans and other forms of financial aid. In response, some countries such as Romania and Bulgaria have raised the age required for a working person to receive full social security benefits [9] [23]. Bulgaria has also increased the contributions that employers and employees must make to the state pension fund by 1.8 percent to help narrow the state pension fund's expanding deficit [9]. Other countries such as the Czech Republic and Romania have raised their value added tax on consumption items as a means of raising revenue [10] [23].

In August 2010, Hungary, Bulgaria, the Czech Republic, Latvia, Lithuania, Poland, Romania, Slovakia, and Sweden petitioned the European Commission to change EU accounting rules that assign money deposited in private pension funds as state spending [11]. These rules expand official budget deficits, adding to countries' public debts, and pushing many of them over the EU's three percent debt ceiling. Exceeding this debt ceiling disqualifies a country from being allowed to borrow or to adopt the euro as its official currency. Despite repeated requests, the European Union turned down the nine countries' appeal.

Responding to the EU's rejection, the Hungarian Parliament, at the behest of the country's prime minister, passed legislation in November 2010 that transfers Hungarians' savings in the private pension system to the state pension system. A worker must specifically ask to remain in the private pension system before January 31, 2011 if he or she does not want to transfer his or her funds. However, Hungarians who choose to keep their savings in the private system will be penalized by being ineligible for public pensions when they retire.

The pension reform rollback was legislated to provide money for the national pension fund which will have insufficient resources to cover pension obligations in 2011 and 2012 without the injection of more cash. The remaining pension savings will be used to reduce government debt – a sum that is currently too high to meet IMF requirements for further lending [21]. Hungary is also levying additional taxes on the banking, energy, telecommunications, and retail sectors to help plug its budget shortfalls [8]. Reducing the national debt by seizing private pension assets allows the government to avoid politically painful austerity measures. Given the country's rapidly aging population however, the pension rollback simply postpones the spending cuts that Hungary must eventually make to allow for sustainable growth in the future.

POLICY ALTERNATIVES AND RECOMMENDATIONS

Finding sufficient funds to support a growing aged population can be done in several ways: (a) raise revenue by increasing taxes, (b) reduce expenditures by limiting pension payments, (c) delay the age of eligibility for retirement benefits, (d) fund future retirement obligations in advance either through a private investment pool or a public pension system [3]. Bulgaria has increased its social security tax. Romania and Bulgaria as well as Germany, the United States, Slovenia, and other countries have raised the age at which citizens can receive full social security benefits when they retire. Two alternative means, particularly appropriate for Hungary, are to tighten a person's eligibility for pension benefits and to attract additional people into the labor force who don't currently work.

Currently, a Hungarian is eligible to receive full pension benefits after 40 years of service even if that benchmark occurs before a person reaches the retirement age of 62. In addition, many people who would be required to work under other countries' laws are able to leave the Hungarian work force based on disability claims for relatively minor injuries or handicaps. Given its unfavorable population distribution, a result of Hungary's continuously falling birth rate, tightening eligibility requirements for disability benefits and providing incentives for extended employment would help mitigate the country's falling funds for the pay-as-you-go pension system. Lowering retirement benefits for individuals who leave the work force before the official retirement age of 62 would encourage workers to work longer, thereby reducing the outflow of retirement monies and extending the amount of retirement savings. The additional revenue from Hungarians who are working instead of collecting disability or retirement benefits would lessen the country's liabilities and contribute to funds for future pensions.

Another recommendation for expanding labor force participation is training, an especially appropriate policy for workers whose skills may now be obsolete. A study conducted by the Organization for Economic Cooperation and Development (OECD) demonstrated that, even after controlling for advancements in education and gross domestic product (GDP), the extent of training alone accounted for up to 40% of the variation across European Union, North American, and other OECD countries' employment rates (correlation = 0.63) [14]. Hungary had the third lowest employment rate among the countries surveyed and the lowest training participation rate during the latter half of the 1990s. More importantly, the correlation between training and labor force participation was even stronger among the same countries studied (0.71), suggesting that investments in human capital may make employment more appealing due to expectations of

higher wages and/or improved odds of finding meaningful work [14]. Given Hungary's current employment rate of 57%, relevant training that can build Hungarians' skills to compete in the new market economy appears to be an especially appropriate policy. It could increase labor force participation, contributing to GDP growth as well as to the DB and DC funds that finance individuals' retirements.

Raising social security taxes could provide more money to support future retirees but the shrinking numbers of working Hungarians means that an increasing burden will be placed on future generations to support pensioners. An alternative policy to raise more revenue is to improve the enforcement of current collections of mandatory retirement contributions. A large informal economy exists in Hungary and other ECE countries; most workers in this unsanctioned sector pay no payroll taxes [12]. If incentives could be crafted to draw these employees into the official economy, another source of income could be realized to fund both the pay-as-you-go DB program as well as DC retirement benefits.

CONCLUSION

Broadening labor force participation through stricter requirements for disability payments, lower pensions for individuals who retire before the age of 62 (including those who have worked for 40 or more years), appropriate training for Hungarians with inadequate employment skills, and extending payroll tax collections of "undeclared workers" could all contribute to a sounder financial footing for the expected wave of retirees who will draw pensions and no longer contribute to the economy. The alternative of the current coercive transfers of mandatory private pensions to the state sponsored one as well as Hungary's dependence on foreign borrowing will only postpone the eventual economic spending cuts necessary to ensure the future pensions of those who have not yet retired. Future benefits for those who support current retirees will be even smaller and more uncertain. The financial burden on those who still work, especially those who will retire later, must be alleviated if the current pension system is to remain viable.

REFERENCES

[1] AXA Group (April 24, 2008). "Hardly 1/3 of Active Hungarians Pre-Save for Retirement," http://www.retirement-scope.axa.com

[2] Bloomberg News (October 28, 2010). "Bulgaria Delays Transfer of Private Pension Funds to State," http://www.hurriyetdailynews.com/n.php?n=bulgaria-delays-transfer-private-pension-funds-to-state-2010-10-28

[3] Bosworth, Barry and Burtless, Gary (1998). "Population Aging and Economic Performance," in *Aging Societies,* Bosworth, Barry and Burtless, Gary, eds. (Brookings Institution Press: Washington, D.C.), pp. 1-32.

[4] Central Intelligence Agency (2009). "The World Factbook – Hungary," https://www.cia.gov

[5] Forelle, Charles (March 25, 2009). "Pension Glut Lies at Heart of Crisis Wracking Hungary," *The Wall Street Journal*, pp. A1, A10.

[6] Global Pensions (December 15, 2010). "Hungarian Schemes Call on European Commission to Fight New Law", http://www.global-pensions.com/global-pensions/news/1932566/hungarian-schemes-europeancommission-fight-law

[7] Global Pensions (December 15, 2010). "Romania Constitutional Court Rules in Favour of Law Change," http://www.globalpensions.com/ global-pensions/news/1932587/romania-constitutional-court—rules-favour-law-change

[8] Groendahl, Boris Balazs, Edith (January 18, 2011). "Hungary Special Levies, Pension Move Cause for Concern, IMF Says", http://www.bloomberg.com/news/print/2011-01-18/humgary-special-levies-pension-move-cause-for-concern-imf-says.hmtl

[9] Konstantinova, Elizabeth, (October 20, 2010). "Bulgaria Raises Pension Contributions to Cover Budget Deficit," http://www.globalpensions.com/global-pensions/news/1800053/bulgaria-raises-pension-contributions-cover-budget-deficit

[10] Livove noviny (January 21, 2011). "Czech Govt Agrees on Key Principles of Pension Reform – Press," http://www.ceskennoviny.cz/news/zpravy/czech-govt-agrees-on-key-principles-of-pension-reform-press-585001

[11] Lovasz, Agnes and Neuger, James G. (December 7, 2010). "Poland Makes Little Progress on Pension Reform Deficit Rules in Brussels," http//www.bloomberg.com/news/print/2010-12-07/poland-makes-little-progress-on-pension–reform-deficit-rules.html

[12] Ministry of Health, Social and Family Affairs (2004). "Pension reform in Hungary," http://www.eszcsm.hu/eszcem

[13] National Center for Policy Analysis (1997). "Social Security Issues: Former Communist States Privatizing Pensions," http://www.ncpa.org

[14] Organization for Economic Cooperation and Development (2002). "OECD Employment Outlook 2002: The Ins and Outs of Long-Term Unemployment," http://www/oecd.org

[15] Organization for Economic Cooperation and Development (2004a). "OECD Employment Outlook 2004: How Does the United States Compare?" http://www.oecd.org

[16] Organization for Economic Cooperation and Development (2004b). "OECD Employment Outlook 2004 - Chapter 4 Striking Facts", http://www.oecd.org

[17] Organization for Economic Cooperation and Development (2004c). "OECD Urges Governments to Combine Job Strategies with Other Social Objectives," http://www.oecd.org

[18] Organization for Economic Cooperation and Development (October 2008). "Growing Unequal? Income Distribution and Poverty in OECD Countries," http://www.oecd.org

[19] Organization for Economic Cooperation and Development (2010). "OECD Factbook 2010: Economic, Environmental and Social Statistics," http://www.oecd.org

[20] Reuters ((December 13, 2010). "Defying Markets, Hungary Readies Pension Rollback," http://news.asiaone.com /print/News/Latest%2BNews/Business/Story/A1Story20101213-252356.html

[21] Simon, Zoltan (December 14, 2010)."Hungary Approves Pension Fund Overhaul as EU Eyes Debt Deal," http://www..global-pensions.com/global-pensions/news

[22] Simonovits, Andras (2002). "Hungarian Pension System: The Permanent Reform," Institute of Economics, Hungarian Academy of Sciences.

[23] Timu, Andra and Savu, Irina (December 7, 2010). "Romanian Parliament Adopts IMF-Backed Pension Law Raising Retirement Age," http://www.globalpensions.com/global-pensions/news/

[24] Tsvetelia, Tsolova, and Ivanova, Irina (December 11, 2010). "Bulgarian Parliament Approves Pension Reforms," http://www.reuters.com

[25] U.S. Census Bureau (2010). "IDB Summary Demographic Data for Hungary," http://www.census.gov

[26] U.S. Census Bureau (2010). "International Data Base Country Summary: Hungary," http://www.census.gov

[27] WHO Regional Office for Europe and the European Commission (2000). "Highlights on Health in Hungary," http://www.euro.who.int

KEY DRIVERS OF TOMORROW FOOD CHAIN.

Corinne Gendron
CRSDD UQÀM
315, rue Sainte Catherine Est
Montréal, H2X 3X2 Bureau R-2845
(514) 987-3000 poste 1400
gendron.corinne@uqam.ca

René Audet
CRSDD UQÀM
315, rue Sainte Catherine Est
Montréal, H2X 3X2 Bureau R-2910
audet.rene@courrier.uqam.ca

Johanna-Maud Egoroff (étudiante)
Candidate à la maitrise en Sc. de l'Environnement
315, rue Sainte Catherine Est
Montréal, H2X 3X2 Bureau R-2565
(514) 987-3000 poste 4183
egoroff.johanna-maud@courrier.uqam.ca

I- AGRICULTURE AS A MARKET.

1. The long history of food market disembeddedness.

Since we rely on large companies and supermarkets to feed ourselves, we often forget that, unlike many other goods, food was not provided through markets during most of human history. The commoditization of food, which has led to its current status as merchandise, has indeed been the result of a long and complex process from the age-old practice of self-production, through the subsequent trade in small markets. By bringing out the specificity of food in the current economic framework, an examination of these origins helps us understand what can be coined as "the exceptionality of agriculture," i.e. the fact that food is not like any other consumer good.

Approximately 10,000 years ago, the beginning of agriculture contributed to a deep transformation in the organization of social life, which evolved from the hunter-gatherer and nomadic forms of society to sedentariness. In 5550 BC, the Sumerians started large-scale and intensive cultivation with irrigation techniques, allowing excess production that could be stored or sold, and reducing wild food to a marginal part of the usual diet of the population. Moschini (2008, p. 331) explains this evolution as follows:

> The challenge of population pressure faced by an environment saturated with hunter-gatherers, catalysed by the changing climatic conditions at the end of the last Ice Age, led some people to attempt the cultivation of wild cereals in an effort to supplement their traditional sources of food (Cohen, 1977). Thus was agriculture born about 12,000 years

ago in the Fertile Crescent, and independently in at least five other distant locations over the next few millennia (Bellwood, 2005). A dwindling supply of large mammalian prey increased the attractiveness of early farming activities, and the ensuing Neolithic revolution saw the widespread transition of prehistoric humans from nomadic hunting-gathering bands to agriculture-based communities relying on a few successful domesticated species of plants and animals. Farming encouraged more permanent settlements, allowing the development of increasingly complex social structures and the ushering in of earlier civilizations.

Later on, numerous innovations led to increased productivity, freeing the population from other activities but also leading to specific configurations of social organization.[1] One could refer to commoditization and industrialization in relation to this period, but this would most certainly be an inappropriate anachronism given that society at that time was not what we call an "economic society," i.e. a society primarily organized on the basis of modern economic rationality (Vergès, 1989). The trade associated with the surplus was part of a more general social system where other rationalities were used as organizing principles, be they religious or political.

This was obvious in the feudal period, when the serf had the right to cultivate his parcel of land, but was also required to give up a substantial part of the harvest to his lord according to the vassal relationship. It is also interesting to note that the serf was denied the right to hunt, i.e. access to "wild food," which was specifically reserved for the lord. At this time, commercial relations were enacted through two dimensions which Braudel (1967) judicially distinguishes as small markets and great capitalism. Small markets were local spaces where, in search of variety, people could trade their surplus with that of others. But this was a marginal means of accessing food: a self-production system prevailed. For the elite, luxurious exchanges were a way to demonstrate supremacy more than to provide basic commodities, and in the Age of Discovery, agriculture also participated in the colonization process, and was a determining factor in the organization of slavery.

This is why, although new farming techniques and space reorganization allowed for the production of a surplus which could be traded, it would be misleading to refer to the commoditization of food or the industrialization of agriculture with regard to this period because agriculture and food markets were still organized according to ecological realities and embedded in social and political relations that cannot be reduced to economic transactions.

This is what Polanyi theorized as the concept of embeddedness, pointing out that food markets have not always been free markets, and that food staples have not always been considered to be commodities (2001). In Polanyi's view, while reciprocity served to sustain humane ways of sharing food and organizing society over the millennia, the withholding and redistribution of food by a centralized authority (such as in the feudal system) also played a prominent role in many societies. These modes of economic exchange were always grounded – embedded – in social relations and nature's cycles. The land and the work force which helped to fertilize it and

---

[1] The understanding of this dynamic may vary among authors, as illustrated by this comment by Moschini: "The path from these early days to the modern world eventually saw the gradual augmentation of technology in food production, which supported a population expansion that is continuing to this day. Some see that as a vicious circle and also lament that '… the transition from hunting and gathering to farming eventually resulted in more work, lower adult stature, worse nutritional condition and heavier disease burden' (Diamond, 2002, p. 700). But civilization has also led to knowledge accumulation, culminating in the industrial revolution and the opening up of increasingly new opportunities" (Moschini, 2008, p. 331-332).

cultivate its fruits, were not market values, they were social institutions. On the eve of the industrial revolution, the concept of the self-regulated market shattered these institutions and put modern society on the path towards the commoditization of agriculture.

2. The industrialization of agriculture.

The industrialization of agriculture can refer to different realities, starting with the mechanization of farming and the use of new technologies and modes of organization, through what has been called the "agricultural revolution" (as compared to the concomitant industrial revolution in England).[2] The turning point in this process was the integration of agricultural production into a new economic order organized on the basis of market rules and rationality instead of other logics, be they social or biological. This transformation of the status of agriculture as a "market" or "capitalist" activity – not merely economic but dedicated to trade with the aim of accumulation – was the result of many structural and ideological transformations which some authors have interpreted as the inexorable march of capitalism to encompass all human activities (Wallerstein, 1974).

First among the transforming factors were new techniques and technologies – including the use of mechanized tractors and fertilizers – which led to an increase in productivity and, to some extent, freed production from unpredictable natural conditions. Secondly, the organization of land, symbolized by the enclosures in England (which are generally seen as one of the major starting points of industrial capitalism), also allowed for an increase in productivity but led to another dramatic change: the exodus of former self-sufficient peasants to cities where they formed the new work force relying on markets to provide them with food. Lastly, it must not be forgotten that these transformations were anchored in the development of an ideology of industrial progress, closely associated with market dynamics as an optimal social regulation.

Thus, in Polanyi's view, the market appears to be an economic form of regulating social relations and human interaction with nature, a form that characterizes capitalism. Indeed, capitalism requires such a system based on the concept of a self-regulated market, at the expense of other economic models such as reciprocity and redistribution. Polanyi argues that at the peak of the industrial revolution, the concept of the self-regulated market shattered these other models. Trade and the search for profit became the fundamental rationale of economic activities in the new political economy of the eighteenth century, giving rise to "the market" as a social institution, different in this respect from local markets. Agriculture thus became part of the great capitalism referred to by Braudel. As a result, agriculture was subject to external pressures to become more productive and less expensive and, eventually, overrule natural cycles and processes. This evolution towards the disembeddedness of agriculture was later marked by the Green Revolution, and by the more recent process of globalization.

---

[2] Kerridge contradicts this understanding (1960). Referring to the thesis developed by Toynbee, he explains: "Toynbee rightly concludes that agricultural progress was relatively slow between 1700 and 1760. But he can hardly have been expected to have devoted much of his short working life to agricultural history, which to him was no more than a sideshow. He gives the matter short shrift and takes the obvious and easy course of turning away from his door all those facts that do not fit in with his preconceived notions that the period 1760-1843 witnessed 'an agrarian revolution [that] played as large part in the great industrial change at the end of the eighteenth century as does the revolution in manufacturing industries.'"

3. From the Green Revolution to globalization.

The Green Revolution originated from the work of scientists who developed new varieties of crops, supported by the use of fertilizers and pesticides, both leading to increased agricultural productivity in response to the growing needs created by exponential population growth.[3] The Green Revolution, as a process, appeared at the end of the 1960s, and deepened until very recently. As Huffman explains:

> Overall, the productivity data suggest that the Green Revolution is best understood not as a one-time jump in yields, occurring in the late 1960s, but rather as a long-term increase in the trend growth rate of productivity. This occurred because successive generations of modern varieties were developed, each contributing gains over previous generations.

> Evenson and Gollin (2003) show that without the Green Revolution, crop yields in developing countries would have been 20% to 24% lower (2004, p. 4).

Applying Western fertilization and pesticide techniques to agriculture in the South, the Green Revolution accompanied the conversion of the war industry into the petro-chemical industry, as best illustrated by Monsanto's evolution (Champion, 2004). But agricultural production was also transformed by the growing practice of trade that changed the conditions under which agricultural commodities were produced. Coleman et al. explain that since the end of the Second World War:

> (…) as agriculture has continued the movement from subsistence, peasant polyculture toward commercial farming, production has become more specialized. Processing companies buying from farmers have demanded more evenness in quality and standardization in size and the markets for these more specialized and standardized products are extending gradually further away from the locality in which crops or animals are grown. The greater emphasis on monocropping has come to rest on various expert systems to achieve economies of scale and higher productivity, including sophisticated farm equipment, chemicals to control pests and weeds, and to stimulate growth, and biological science to improve yields from plants and to raise output from animals. In some areas of livestock production like poultry and hogs, farmers are entering into quite specific contracts with processing companies that govern the nature and the quality of products to be delivered (2004, pp. 7-8).

These developments have led to a deep transformation of the relationship between the farmer and the consumer, which has been progressively mediated by economic agents, while agricultural production has become part of a complex economic system encompassing science and technology, mechanical instruments and worldwide distribution networks in a world marked by geopolitical tensions or conflicts and a new consumerist way of life. In this system, the self-producing farmer has progressively been replaced by a market producer whose conditions of

---

[3] The demographic rationale of the Green Revolution is well illustrated by Huffman: "as the demand for food and fiber grew during the past 300 years, because of the Earth's expanding human population and rising per capita incomes, society met this demand first by increasing the land area under cultivation and later by improving crops so that their yields were higher. Before 1900, land was abundant almost everywhere, and in the United States, new lands were brought into production as the frontier moved across the country between 1700 and 1900. (…) By 1900, the frontier was closed in the United States, and this increased the urgency of finding new methods for increasing crop yields" (2004, p. 3).

production are dictated by influential economic actors in concordance with state-led development strategies.

> The production of agricultural commodities has gradually become enmeshed in linkages that are more globally extensive today than they were a half-century ago, whether through suppliers of inputs to agricultural production or through the processing, distribution and sale of agri-food products. Trade both in agricultural commodities and in processed food products has risen over the same period. This intensified economic competition on a more global scale has accentuated already powerful tendencies toward greater economic, technological and socio-cultural integration in the triad of the North America, Europe, and East and South-East Asia (…) (Goodman, 1997, p. 672).

> These more intense regional and economic relationships have developed through the greater presence of multinational firms in the sector, and through important changes in technology used in the transportation and distribution of commodities and processed food products. Consequently, global economic relations are a more important factor in the lives of farmers and of food consumers than they were a half-century ago. (Coleman et al., 2004, p. 7).

For these authors, globalization can be seen in direct continuity with industrialization and the Green Revolution in the process of agricultural disembeddedness.

> Globalising processes have accelerated changes in the countryside in most countries, changes already in train due to mechanisation and chemical farming. More food is being produced on less land by fewer farmers than a half-century ago. Old patterns of relationships between farmers, the suppliers of their inputs and the buyers of their crops are shifting due to corporate restructuring. Political arrangements in the wealthier countries are being criticized and slowly changed in response to challenges from within and from those suffering the effects of subsidies in the Global South. Old orders are changing but there is scant evidence to date whether any new order will take their place (Coleman et al., 2004, p. 4).

However, while globalization seems indeed to have intensified the process of commoditization initiated centuries ago, recent developments might be interpreted as a suspension, and even a turning point in this process. We would argue that following its accentuation through different phases of varying length and speed depending on the country, and its final intensification towards the end of the 1980s, the process of commoditization of agriculture ended with the last attempts made under the globalization process to handle food as an ordinary good. The negotiations aiming to liberalize international agricultural markets were defeated by protectionist strategies at the WTO, confirming a suspension of the rationale behind the commoditization of food.

4. Global trade and agricultural policy towards decommoditization?

In spite of the commoditization process that has been underway for centuries, among global trade issues, the agricultural sector has always been recognized as being somewhat particular (Trebilcock and Howse, 2005; Aggarwal, 2003). Along with the services and investment sectors, it was one of the last major economic sectors planned to be liberalized after the Uruguay Round of the WTO. After World War II, the United States refused to apply GATT (General Agreement on Tariffs and Trade) principles to the agricultural sector for political and strategic reasons, and

was allowed to do so. Under the post-war reconstruction, Europe also relied heavily on subsidies and market protection in structuring its internal agri-food economy. Becoming competitors on a global scale, the UE and the US were then drawn into a "war of subsidies" during the seventies and eighties – a process which brought about increasing price distortion in the agricultural market.

Although every national situation is different, this era saw agricultural policies being constructed around three axes that would eventually become the three "pillars" of the WTO Agreement on Agriculture. One is the traditional imposition of entry tariffs and taxes, which make imported products less competitive than domestic equivalents. For some imported products considered to be sensitive, a country can impose "tariff peaks" – meaning tariffs that are many times higher than regular tariffs – as a barrier to entry. Another policy area that has the effect of distorting prices is domestic support. All kinds of tax reduction for producers, public insurance programs for farmers, subsidies for production or for the purchase of machinery, etc., can be considered as domestic support measures. Finally, the most contested policy area under the GATT and WTO is that concerning export subsidies, which constitute public support to sectors oriented toward exportation. While these can be justified as a means by which to counter the effects of entry barriers (of the type described above) in foreign markets, their impacts are being widely felt in developing countries, where the total amount of subsidies is a fraction of that of the US or the UE. Overproduction in developed countries can also be subsidized to facilitate exportation to other countries, thus creating a form of "dumping" (Drope & Hansen, 2006).

Thus, for different reasons – mainly political – agriculture has never been governed by a free market, and this has impeded the process of commoditization. This has made multilateral policy negotiations in agriculture quite paradoxical: while free trade originally meant opening up markets in the Global South to products from industrialized countries, free trade in agriculture refers to the reverse, that is, the opening of markets in the Global North to foodstuffs produced in developing countries. For the first time in the recent history of trade negotiations, the Uruguay Round tackled the liberalization of agriculture. The result, however, was rather modest. According to most experts on international trade law, the Uruguay Round Agreement on Agriculture, resulting from a bilateral bargain between the United States and the European Commission, consisted more in the legalization of protectionism in the agricultural sector than in its liberalization. (Aggarwal, 2005; Clap, 2006). Such a result was not going to be satisfactory to developing countries, and the next negotiating round would have to address the problem of agricultural protectionism more deeply.

5. North-South inequities in agricultural trade and the case for food security.

In the agricultural sector, the huge subsidies and market protection policies enforced by the European Union and the United States (along with the majority of developed states) have had the most adverse impacts on developing countries, whose advantage based on low cost agriculture is nullified by closed markets in the North. Instead of bringing real liberalization to the sector, as observed above, the Uruguay Round Agreement on Agriculture legalized European and American protectionism. This is why the Doha Round was launched in 2001 with much anticipation on the part of developing countries who were hoping for "the redressal of the inequitable nature of existing provisions of the Agreement on Agriculture" (G77, 2001). After more than two years of stalled negotiations at the WTO, it is now clear that the North-South inequities written into international trade law still constitute a great challenge to the notion of

"equity" in global agricultural trade. Simultaneously, they impede a potentially formidable driver of the food chain: food security.

Food security is an interesting concept because it is based on the notion that food is more than a commodity – it is a vector of security or insecurity. Indeed, since a proper diet and the nourishment of communities, be they rural or urban, ensures greater stability and productivity, food is an essential driver of development. This is why the "G33 on special products and the special safeguard mechanism," a coalition of developing countries that emerged at the Cancun WTO Ministerial Conference in 2003, states that the "deepening global food crisis has clearly demonstrated that indeed the principles of food security, livelihood security and rural development are relevant and must indeed be a part of the equation in discussing the optimal solutions to [the Doha] Development Round" (G33, 2008). If the Doha Round is ever concluded, multilateral trade agreements may well institutionalize mechanisms to ensure food security in developing countries, such as the possibility of designating "special products" – meaning products that are considered important for food security and rural development – that would benefit from better tariff protection, and a "special safeguard mechanism" that would allow countries to react with tariffs and quotas in the case of a sudden price drop or domestic market "flood" of competitive imported products (Lal Das, 2005). Similar mechanisms could be available for developed countries under the denomination of "sensitive products" (Trebilcock and Howse, 2005).

It is interesting to note that the content of the last negotiations occurring under the WTO, rather than being a pure translation of the commoditization rationale, indeed recognized the "exceptionality" of agriculture. Clearly, the food security issue has brought about the idea that agriculture may have functions other than simply being a commodity. Agriculture is increasingly presented as being "multifunctional," as shown by the new trends in food consumption and production. Thus, while we recognize that the globalization process is transforming agriculture through an intensification of the economic integration of its actors and activities, our understanding differs from that of Coleman et al., who see in globalization a deepening of the commoditization and disembedding process.

> In short, despite its rootedness in place and its dependence on the natural rhythms of the seasons, the life cycles of animals, and the climate, agriculture is changing rapidly in character as it becomes drawn into globalising processes in the economy, the world of politics, and culture. As an economic activity, it is losing its exceptionality and becoming one sector among others contributing to economic growth. Politically, the long-standing protective mantle of the nation-state is yielding to new forces, new rules and new constraints defined at regional and global levels. With nation-states yielding some authority, such cultural notions as farming being the backbone of the nation and that nations must be self-sufficient in food are being reconsidered and recast (Coleman et al., 2004, pp. 3-4)

It is mainly because of the exceptionality of agriculture that the trade liberalization process undertaken by the Uruguay Round failed, and that the globalization process took a new direction in accordance with this special status: a limitation of free trade primacy with regard to geo-strategic imperatives fuelled by the particular status of agriculture. This is not to say that we are entering a process of *de*commoditization of food, but that the march towards commoditization

has reached an end. This can be explained by a series of factors anchored in the nature of agriculture itself, wherein there have always been intrinsic limitations to the de-territorialization rationale of globalization, as clearly pointed out by Coleman et al.:

> (…) the immobility of land and the local circumstances of land use would seem to make farming a highly unlikely sector to be integrated into international and global processes. (…)
> With its implications for the declining constraints of physical location and its emphasis on de-territorialisation (Appadurai, 1996), globalisation would seem to refer to processes not particularly relevant for understanding agriculture. After all, farming remains an activity intimately tied to particular rural communities, their soils, their micro-climates, and their physical environments. Although buying and selling currencies might occur anywhere at any time, growing wheat or milking cows takes place in very particular localities, following regular patterns dictated by the seasons or basic cycles of life. Tied to the land and drawing upon centuries, if not millennia, of tradition and cumulated experience, farming, one of the ancient arts of human civilization, would seem to be somewhat beyond the reach of current globalising processes (Coleman et al. 2004, pp. viii and 1)

As Coleman et al. explain, despite these characteristics, agriculture has not been spared the effects of globalization. However, we would go one step further. Inspired by a comment made by these authors suggesting that continuing globalization should not be seen as the only possible outcome of the current interplay of economics, politics and technology (ibid.), we argue that the exceptionality of agriculture has actually changed the course of globalization, in part because of the struggle that its initial liberalization rationale brought about in this sector.

> The changes that we observe bring in their train a variety of avenues of contestation and struggle. The introduction of biotechnology, the growing of genetically modified organisms, and the incorporation of these into processed foods and animal feeds have given rise to field burnings, consumer boycotts, and political debates across the globe. Environmentalists have challenged "industrial farming" and its use of chemicals because of the effects on water courses and human health. Nation-states misuse global and regional trade rules to give "their" farmers an edge over competitors. Farmers continue to fight to stay on the land in most countries of the world. In short, the production and preparation of food promise to remain a prominent site of resistance to globalisation for the foreseeable future. (Coleman et al., 2004, p. 4)

Initially driven by universal free trade agreements, the economic integration that occurred during the first era of globalization is now entering a new age in which it obeys new logics and parameters. These parameters will both sanction the exceptionality and multifunctionality of agriculture and be supported by them. In this sense, agriculture plays an important role in the general restructuration of markets illustrated by the rise of responsible consumption and social responsibility strategies adopted by firms. In this new age, as the North-South tension is complexified by commercial alliances rooted in the strategic interest of nations (as is the case with GMOs), new issues such as environmental protection, greenhouse gas emissions, water supply and foreseeable shortages are deeply changing the governance challenges and the rationale behind state-led negotiations and the strategies adopted by firms in the field of agriculture.

II- NEW TRENDS: REDISCOVERY OF EMBEDDEDNESS THROUGH MULTIFUNCTIONALITY.

While agriculture has been a domain in which the human spirit has overcome nature's unpredictability and led to civilization and progress, today it seems to illustrate a total loss of control over nature's dynamic as it participates in disease, health problems and social fracture. Agriculture is no longer presented as the instrument of progress and well-being, but rather as an area involving the convergence of economic, political and social problems.

The publication of Rachel Carson's Silent Spring was a revelation for Western citizens concerning the dangers of chemicals for ecosystems (Hails, 2002, p. 685), and this revelation was followed by discoveries about the risk they present for human health as well. Considering the awakening provoked by what can be called the side effects of the Green Revolution, it is not surprising that the "Gene Revolution" has aroused suspicion. Even more recently, the environmental issue has brought up the problem of transportation, showing that the energy needed for food trade often exceeds the energy accumulated in food itself. In addition, the issue of water supply and new data about fishery shortages (ONU, 2006) have also led to concern.

Another issue that has been raised by the growing control of agricultural production by corporations relates to the use of land. Concurrent with the advent of urbanization, rural land is no longer seen as a space for social and human life, having been transformed into large dehumanized production areas in some countries, and being subject to severe conflict over usage in others. With regard to consumption, production processes have also been associated with health problems, especially obesity, heart disease and even cancer. Moreover, the consumer has progressively been disconnected from the ecological and natural dimensions of food which is increasingly packaged and transformed. Finally, food is increasingly seen as a major focal point of tension between the North and the South, as shown by the WTO negotiations, with tensions also arising between Europe and United States over the issue of GMOs.

This concentration of issues and the conflicts that have arisen over the last decade have been theorized in terms of the co-existence of four different paradigms to which countries refer in the governance of their agricultural sector: the dependent, competitive, multinational and global production paradigms (Josling, 2002). The dependent and competitive paradigms could be said to illustrate the first two economic periods of the last century. Coleman et al. define them as follows (2004, p. 94):

> A dependent paradigm organized around the core belief that agriculture fulfils basic food needs and provides national security, as well as social and political stability and rural employment and welfare, but requires government help to enable it to generate adequate incomes.

> A competitive paradigm that emphasizes agriculture as a sector that can hold its own against other sectors of the economy and that can thrive in a market economy and an international trade system (at least, where markets are permitted to operate free of distorting, dependent paradigm style policies).

In our view, we are currently experiencing competition between the two last paradigms,[4] a situation which is most likely to result in a combination of the two rather than the success of one over the other. These paradigms are also defined by Coleman et al. (2004, p. 95) based on the work of Josling (2002):

> A multifunctional paradigm organized around the belief that agriculture is an integral part of the countryside and provides non-market goods that would be under-produced without some degree of government support.

> A globalised production paradigm that situates agriculture in potentially global food supply chains, where farmers are seen as supplying land and animal management services to an integrated vertical process from input supply and technology provision through marketing of the product.

As Coleman et al. add in a note, this last globalized paradigm does not exclude local food systems, which are part of the model: "Josling (2002) called this paradigm 'globalized' agriculture to emphasize the international nature of many of the supply chains. We follow this terminology while recognising that local and regional (short) supply chains are also part of the model. In some cases, the 'driver' of the chain is the ultimate consumer: in most cases it is the retailer that defines the requirements to be met by the supply chain" (2004, p. 95, note 3). This remark leads us to the reflection that we are indeed probably experiencing the exploration of a new articulation between the dependent and competitive paradigms, an articulation which will manifest itself through a globalized production paradigm encompassing a multifunctional paradigm, and which becomes possible only by assessing the particular nature of agriculture and confirming its exceptionality. In our view, it is this articulation that is illustrated by the numerous innovations prevailing in the agricultural field, some of which are presented below.

1. The turn to quality and alternative models of consumption.

As shown by the diversification of restaurant food, the growing number of ethnic shops, the success of television shows about cooking and the content of advertising, food consumption today refers to more than simply satisfying basic needs (if indeed it ever meant so little). Buying food, in many sectors of society, is a symbolic act that goes as far as an expression of identity or a political statement. Slow food, local food, labeling, community-sustained agriculture, etc.: all these new trends mark a "turn to quality" in food consumption. The word "quality" here should not be restricted to the meaning of "better quality" or "different quality," but refers to the characteristics of food resulting notably from distinct processes. Indeed, the turn to quality relates to a search for better quality products, but quality also includes a range of symbolic motivations behind the purchase of a food product: its origin (Is it from my region, from this area?), its mode of distribution (Will I have contact with the producer? Can I find it in specialized, health or green food stores?), its mode of production (Is it antibiotic free, GMO free? Is it organic? Does it respect religious traditions – halal, kosher…), and values (Is it fair? Is this tuna dolphin safe?).

Thus, there is a great diversity of consumption practices existing under the label "turn to quality." Moreover, contradictions are numerous. For example, some authors have associated the turn to quality with progressive values such as fairness in consumer-producer relationships or

---

[4] The analysis offered by Coleman et al. presents current conflicts as a struggle between these four paradigms (2004, p. 119).

ecological protection. Yet research has shown that buying local is not always a matter of fairness, and neither is it inescapably an ecological good. In a study of food consumption in five regions of England in the wake of the sanitary crisis of the nineties, Michael Winter (2003) showed that what is sometimes interpreted as a search for quality can also appear as a fear of risk. He concludes that consumption is a complex act and that researchers should avoid false dichotomies such as capitalism on the one side and "buying local" on the other. A similar argument is made by Clare Hinrichs (2003) concerning local food distribution systems in Iowa, where localism seems to express a certain conservative reaction to globalization, a cultural defense against foreign food. Hinrichs thus argues for a broader vision of the turn to quality that could simultaneously integrate defensive conservative reactions and progressive sustainability values by putting them at the poles of a continuum where different degrees also exist.

This also implies that what is called the "embeddedness" approach to alternative agriculture should be reconsidered. According to David Goodman, the embeddedness concept favors only one aspect of social life – connectivity – and fails to see the social relations related to market dynamics (Goodman, 2003). This approach thus tends to build a caricature which portrays society as being in opposition to the market. As drivers of the food chain, however, the turn to quality in consumption practices and the initiatives of alternative agriculture seem, rather, to involve a complex mix of the market, the social relationship and values, and also new modes of regulation.

2. Environmental and sustainable agriculture certification.

One concern encompassed by the turn to quality trend is protection of the environment. Since, like other such issues, environmental protection cannot be assessed by examining the product itself, the consumer must rely on specialized distribution channels or certification programs. The Rainforest Alliance, to provide a first example, certifies shade grown products in the tropics in order to help poor farmers sustain the forest as their home, as a source of income and as a refuge for migrating birds. The "Core Northeast Values" initiative, with the collaboration of universities in the region, has developed a program to help farmers transition to ecological agriculture, which also involves certification and labeling. Producers from numerous regions are getting together to begin certifying local "terroir" products in order to protect their added value against imitation by larger competitors. Indeed, sustainable agriculture (and food) initiatives have widely adopted the certification and labeling strategy to promote their products. In doing so, they have contributed to the creation of a new regulatory framework in this sector – one that can be termed as "hybrid" (Audet, 2010).

Certification and standardization existed well before most of these examples. Kosher and halal labels originating from faith-oriented organizations may have been amongst the first of the kind. However, governments have also played a part in the emergence of this phenomenon: the Dolphin Safe label was an initiative of the American government to promote the use of a certain kind of (dolphin safe) net in tuna fishing, and the French government has long had its own system of certification and labeling for "origin appellations" in the winery, cheese and dairy, and meat products sectors, among others. Moreover, since a few sustainable agriculture labels have received a lot of attention from consumers and firms, governments have begun to introduce regulations to ensure consumer protection and foster the development of this sector. Organic agriculture may be the best example of this trend as most governments now have regulations concerning organic certification and even organic standards (Willer et al., 2008).

In a bottom-up fashion, a hybrid regulation framework has emerged from the interlocking of all these initiatives, in the domain of agriculture as well as in many other sectors. This framework involves different types of organizations attending to certification, inspection, accreditation, promotion, standard-setting, etc. – whether they originate from civil society or from government or intergovernmental agencies. Such a mix of private sector, civil society and state actors brings to the growth of the food certification business a set of very complex issues. For example, labeling confers on a product a new characteristic, one that an unlabeled equivalent product does not have. According to free market promoters, or to Global South producers trying to export in northern markets, such a characteristic can become an unfair advantage, hence introducing "distortion" in the market. This is why WTO rules on technical barriers to trade state that certification schemes should aim at international harmonization, so that the advantage will be the same for certified products – such as organic products – all around the world, and so that certification does not become a technical barrier to trade (Audet, 2010). As a result, hybrid regulatory frameworks increasingly interact with conventional national and international regulation frameworks, forcing them to adapt to the new reality of sustainable food production and distribution.

3. Fair trade and organic agriculture.

Among types of sustainable agriculture certification, organic agriculture and fair trade are probably the best documented and most widely known. While organic agriculture organizes agricultural production around techniques aimed at maintaining soil productivity and controlling pests without the use of chemicals, genetically modified organisms or other synthetic materials (antibiotics, growth hormones, etc.), fair trade, on the other hand, is a system of trade between marginalized producers in the South and consumers in the North, based on values related to equity. Both organic agriculture and fair trade use certification and labels to control the transparency of their respective food and staple chains and to make their products identifiable on market shelves. Historically, both initiatives conceived themselves as "movements," but fair trade and organic agriculture have evolved a great deal since their emergence in the sixties. They now also share the challenge of conventionalization, that is, the process by which organic agriculture is being mainstreamed and its standards weakened (De Wit & Verhoog, 2007), and by which fair trade is being extended to larger production models (such as plantations) and multinational agri-food corporations.

The conventionalization phenomenon is also somewhat challenging when it comes to analyzing sustainable agriculture, especially for those observers who tend to understand these initiatives as being in opposition to capitalism and free markets. It is in part in the literature on this subject that Polanyi's concepts of embeddedness and disembeddedness have grown roots. For example, Laura Raynolds (2000) and Elizabeth Barham (2002) have shown that the organic and fair trade movements were partly built on a critique of conventional agriculture, as both movements reject the specialization, intensification and "chemicalization" of agriculture. These authors demonstrate that productivism is a figure against which these movements define themselves, and that the "international organic and fair trade movements seek to create alternative trade circuits for items produced under more environmentally and socially sustainable conditions that simultaneously parallel and challenge the conventional global agro-food system. [...] Both movements work to re-embed production in natural and social processes and create an alternative agro-food system" (Raynolds, 2000, p. 306).

Re-embedding means that the production and trade activities in organic and fair trade food chains are supposed to create or strengthen social relations of solidarity and redefine the role of farmers in natural cycles. Fair trade and organic agriculture hence hypothetically contribute to the creation of new social and economic institutions embedded in nature and social relations. This interpretation, however, can be challenged. In fact, one of the main characteristics of these social movements is their tendency to work inside the market. Of course, they have led to the development of alternative distribution routes and networks – although conventionalization is making these networks increasingly less alternative – but they have also had to mobilize expertise on product quality and sanitary standards, shipping and handling, marketing know-how, and, above all, the business of standard-setting, evaluation and certification. This has not only been true for fair trade and organic agriculture, but also for most initiatives related to sustainable agriculture and food.

4. Social responsibility.

Partly in response to the protest movement and fair trade discourse, but more generally anchored in a global trend toward corporate social responsibility, agri-business corporations have begun to rethink the branding of their products, enhancing the social and environmental quality of the latter and sometimes initiating a more profound transformation of their strategic positioning. The case of Danone, among big players, and Liberty as an example of a smaller business, clearly illustrate how the discourse, as well as practices, development strategies and public positioning have begun to be modeled around social issues, especially in the field of agriculture. Corporate social responsibility is often defined as voluntary initiatives going beyond the law and mechanisms for dialogue with stakeholders; however, Gendron (2009) has shown that this movement indeed reveals a deep transformation in business legitimacy rationale, forcing corporations to define their mission with respect to the common good and general interest of society. Corporations thus play a new role in the shaping of social issues which shed light on their positioning and discourse like never before. One must add to this the fact that corporations are seen as being responsible for numerous contemporary problems, especially those of an ecological nature. Contrary to what is sometimes said, corporate social responsibility cannot be reduced to greenwashing or to superficial change; when these practices are encountered, they are generally only first and wrong steps in a long path towards the real integration of new parameters in production, governance and relations with communities and with consumers or economic partners in the food chain. As shown by the "awakening" of some supermarkets that were denounced by environmental groups, corporate social responsibility is no longer an option, and even in areas that are not (as yet) regulated, firms are obliged to show their contribution to solving society's challenges.[5]

These corporate strategies are inspired by and in some ways embody an expression of the exceptionality of agriculture, and they build on the social and ecological embeddedness of the products themselves. They are thus participating in the construction of the new rationale that will predominate in agricultural governance in the coming years.

---

[5] By way of illustration, we present the following case: after being ranked last in 2009 by Greenpeace with regard to the issue of fisheries, the Metro supermarket adopted a new policy to avoid the distribution of endangered species, to which the environmental group responded favorably, granting them a fifth position in 2010.

Conclusion

Considering the intensity, diversity and persistence of struggles surrounding agriculture, agriculture seems to be at the heart of the construction of a new rationale that will bring about or sustain a governance system that will move away from the ideal of liberalization. While this hypothesis seems daring today, it is not absurd to think that the evolution of ecological issues and their growing importance at the international level will compel a governing system that challenges the paradigm of trade liberalization as the only or most powerful vector of public good.

In this sense, as it seems to bring together today's social and environmental problems, agriculture might also be the most fertile ground for governance innovation as well as a window into the most important cultural, political and economic transformations of our societies. Following the arguments we have presented in this chapter, the key drivers of the food chain are not to be found in the increased commoditization of food as was the case during the last centuries, but rather in the social, ecological and even mythical reinvestment in food.

This reinvestment does not have to occur outside the market, because, indeed, it is transforming the market, which is increasingly being structured around normative parameters that cannot be reduced to the procedural imperatives of a free market. Agriculture is contributing to this transformation and to the rule of what might be called the second globalization, i.e. an international integration organized on the basis of a rationale other than market liberalization. Multifunctionality will play an important role in the legitimization of this new order in the field of agriculture, and this is all the more probable since it will serve the interests of the dominant actors, as proven by the failure of the last round of WTO negotiations.

References

Aggarwal, Rajesh, 2005, "Dynamics of Agriculture Negotiations in the World Trade Organization", *Journal of World Trade*, Vol. 39, No. 4, pp. 741-761.

Alavanja, Michael C. R., Claudine Samanic, Mustafa Dosemeci, Jay Lubin, Robert Tarone, Charles F. Lynch, Charles Knott, Kent Thomas, Jane A. Hoppin, Joseph Barker, Joseph Coble, Dale P. Sandler and Aaron Blair. 2003. "Use of Agricultural Pesticides and Prostate Cancer Risk in the Agricultural Health Study Cohort". *American Journal of Epidemiology*, Vol. 157, No. 9, pp. 800-814.

Audet, René, 2010, "L'institutionnalisation de l'agriculture biologique vue par le Sud: régulation inégalitaire vs tropicalisation", in C. Gendron, J.-G. Vaillancourt and R. Audet (eds.), *Développement durable et responsabilité sociale. De la mobilisation à l'institutionnalisation*, Montréal: Presses internationales Polytechnique.

Barham, Elizabeth, 2002, "Towards a theory of value-based labelling", *Agriculture and Human Values*, No. 19, pp. 349-360.

BRAUDEL, Fernand. 1967. *Civilisation matérielle, économie et capitalisme (XVe-XVIIIe siècle)*, Volume 1, Paris, Armand Colin, 1967.

Braudel, Fernand. 1988. La dynamique du capitalisme. Paris: Flammarion.

Gendron, C. *L'entreprise comme vecteur du progrès social: la fin ou le début d'une époque?,* CRSDD publication, "Recherche" collection, No. 01-2009, Montreal: Chaire de responsabilité sociale et de développement durable, ESG UQAM.

Champion, Emmanuelle. 2004. Analyse du discours de Monsanto 1970-2002: les dimensions sociales et environnementales dans le renouvellement de la légitimité institutionnelle de l'entreprise transnationale.

Charles, H. J. Godfray, John R. Beddington, Ian R. Crute, Lawrence Haddad, David Lawrence, James F. Muir, Jules Pretty, Sherman Robinson, Sandy M. Thomas and Camilla Toulmin. 2010. "Food Security: The Challenge of Feeding 9 Billion People". Science, Vol. 327, pp. 812-818.

Clap, Jennifer, 2006, "WTO Agriculture Negotiations: Implications for the Global South", *Third World Quarterly,* Vol. 27, No. 4, pp. 563-577.

Coleman, William D., Wyn Grant and Timothy Edward Josling. 2004. *Agricultural in the New Global Economy*. Cheltenham: Edward Elgar.

De Wit, J. and Verhoog, F., 2007, "Organic values and the conventionalization of organic agriculture", *NJAS - Wageningen Journal of Life Sciences*, Vol. 54, No. 4, pp. 449-462.

Drope, Jeffrey M. and Wendy L. Hansen, 2006, "Anti-dumping's Happy Birthday?", *The World Economy*, Vol. 29, No. 4, pp. 459-472.

Falcon, Walter P and Rosamond L. Naylor. 2005. "Rethinking Food Security for the 21st Century". *American Journal of Agricultural Economics*, Vol. 86, No. 5, pp. 1113-1127.

G33 on Special Products and the Special Safeguard Mechanism (G33), 2008, *G33 Statement Informal TNC*, Geneva, June 26th, 2008.

Goodman, David, 2003, "Editorial. The quality turn and alternative food practices: reflections and agenda", *Journal of Rural Studies*, No. 19: 1-7.

Group of 77 and China (G77), *Declaration by the Group of 77 and China on the Fourth WTO Ministerial Conference at Doha*, Qatar, October 22, 2001.

Hails, R. S. 2002. "Assessing the risks associated with new agricultural practices". *Nature,* Vol. 418, pp. 685-688.

Hinrichs, C. Clare, 2003, "The practice and politics of food system localization", Journal of Rural Studies, No. 19: 33-45.

Hinrichs, C. Clare. 2000. "Embeddedness and local food systems: notes on two types of direct agricultural market". *Journal of Rural Studies*, Vol. 16, pp. 295-303.

Huffman, Wallace E. 2004. "Production, Identity Preservation, and Labeling in a Marketplace with Genetically Modified and Non-Genetically Modified Foods". *Plant Physiology*, Vol. 134, pp. 3-10.

Kerridge, Eric. "The Agricultural Revolution Reconsidered". *Agricultural History* (Agricultural History Society) 43, No. 4 (October 1969): 463

Lal Das, Bhagirath, 2005, *The Current Negotiations in the WTO. Options, Opportunities and Risks for Developing Countries*, London/New York, Zed Books/Third World Network.

Lewandowski, I, and A.P.C. Faaij. 2006. "Steps towards the development of a certification system for sustainable bio-energy trade". Biomass and Bioenergy, Vol. 30, pp. 83-104.

Moschini, GianCarlo. 2008. "Biotechnology and the development of food markets: retrospect and prospects". *European Review of Agricultural Economics*, Vol.35, pp. 331-355.

OECD. 2008. Environmental performance of agriculture in OECD countries since 1990, pp. 84-91.

Polanyi, K., 1957. The economy as instituted process. In: Polanyi, K., Arensberg, C.M. and Pearson, H.W. (eds.), 1957. *Trade and Markets in the Early Empires*, Free Press, Glencoe, IL, pp. 243–270.

Polanyi, Karl, 2001, *The Great Transformation: The Political and Economic Origins of Our Time*, Boston, Beacon Press.

Raynolds, Laura T., 2000, "Re-embedding Global Agriculture: The International Organic and Fair Trade Movements", *Agriculture and Human Values*, Vol. 17, pp. 297-309.

Renard, Marie-Christine. 2003. "Fair trade: quality, market and conventions". Journal of Rural Studies, Vol. 19, pp. 87-96.

Taylor, Peter Leigh. 2005. "In the Market But Not of It: Fair trade Coffee and Forest Stewardship Council Certification as Market-Based Social Change". World Development, Vol. 33, No. 1, pp. 129-147.

Trebilock, Michael J. and Robert Howse, 2005, *The Regulation of International Trade* (3rd Edition), London/New York, Routledge.

Vergès, Pierre, 1989, "Représentations sociales de l'économie: une forme de connaissance", in *Les représentations sociales*. D. Jodelet (ed), Paris: Presses universitaires de France, pp. 387-405.

Wallerstein, Immanuel. 1974. *The Modern World-System I: Capitalist Agriculture and the Origins of the European World-Economy in the Sixteenth Century*. New York: Academic Press, Inc.

Wallerstein, Immanuel. 1976. *The Modern World-System: Capitalist Agriculture and the Origins of the European World-Economy in the Sixteenth Century*. New York: Academic Press, Inc.

Willer, Helga, Minou Yussefi-Menzler and Neil Sorensen (eds.), 2008, *The World of Organic Agriculture. Statistics and Emerging Trends 2008*. IFOAM, Bonn.

Winter, Michael, 2003, "Embeddedness, the new food economy and defensive localism", Journal of Rural Studies, No. 19, pp. 23-32.

Wong, David. Claire Loh and Vanessa Quintal. The impact of organic certification information on consumer's perceptions of organic products, in Spanjaard, D. and Denize, S. and Sharma, N. (eds.), Proceedings of Australian and New Zealand Marketing Academy Conference, Dec 1, 2008. Olympic Park, Sydney: University of Western Sydney.

INVESTORS OR GAMBLERS: ABOUT SELECTED ASPECTS OF FUNDING RISKY PROJECTS

WOJCIECH NASIEROWSKI

Społeczna Wyższa Szkoła Przedsiębiorczości i Zarządzania (SWSPiZ) (Academy of Management, Łódź, Poland), ul. Sienkiewicza 9, 90-113 Łódź, Poland;
tel./FAX 48 22 408 4623
Faculty of Business Administration, University of New Brunswick, Fredericton, NB, E3B 5A3 Canada, < nasierow@unb.ca > ; tel. 1 506 458 7338

ABSTRACT

Aspects of various types of risky business investments (in start-up ventures, innovations, technological development) are explored in this paper. Such investments could be considered a form of gambling. Based on a literature overview, conditions for these types of activities are discussed and related statistics are examined. These types of investments/expenditures are risky, though there is a potential for a high pay-off. It is shown that investment in such high-risk business ventures is positively related to ease of doing business, and pro-entrepreneurship attitudes, as well as with uncertainty avoidance. However, it is also positively correlated with pro-gambling tendencies. Implications of such findings are explored and suggestions for further studies are provided.

Key Words: private investment, 'business angels', venture capital, gambling, innovation

1. INTRODUCTION

Whether for the improvement of the economy, or for monetary, social, or environmental benefits, putting innovations into actual use, supporting new ventures and making investments in technological development are of paramount importance. There are numerous papers on these activities and the conclusions reached frequently emphasize that without the opportunity to actually apply the concepts, they account for almost nothing. Thus funds and some policy related framework, often described in the format of National Innovation Systems, is needed. All the participants bring different assets to the table: governments have the legislative power; research institutions and universities provide knowledge, findings and people ready to work towards innovation; while businesses and entrepreneurs can be a testing ground for new ideas. This paper explores some patterns of funding for innovative/inventive activities, and compares these activities with gambling. Both ways of spending money offer a high pay off. Yet, estimates indicate that only 1 out of 7 innovative projects result in commercial success. Some other sources claim only 12% projects succeed (e.g., [9], [14], [24], [37]). Thus, the question arises: why should one invest in risky business ventures which require active, knowledgeable involvement and an understanding of how business works, when there are other means to enjoy life or, to risk capital in order to win or lose. One can go, for example to a casino to play roulette or black-jack where there is a chance to win big, even if these chances are relatively low. In both cases – investments and gambling – there are country specific reg-

ulations and tax consequences. Also, there is an impact of attitudes about investors/gamblers relating to ease of doing business or gambling, entrepreneurial attitudes, propensity to accept risk and degree of greediness. Following such a line of reasoning one can formulate the following hypotheses:

H1:   Indicators of perceptions of the ease of doing business, entrepreneurial drive, and public and private support for technological development are highly positively correlated with investments in various forms of technological development.

H2:   Indicators of investments in innovations are highly and positively correlated with revenue from gambling activities.

Conclusions from an exploration of these hypotheses are expected to show some country specific solutions. These conclusions may suggest/recommend ways of fostering risky business activity, and/or policies about some forms of entertainment --- a comparison of business with gambling. To this end public/private investments may serve as an indicator of commercialization propensity. This is particularly the case with informal business investments, because businesses have a higher rate of success, since they know better what may have commercial value. On the other hand, however, gambling is also a private investment. The examination of these issues is warranted.

In order to address research questions posed in this paper the following outline has been adopted. First, a literature overview of the issues related to private business investments will be explored. The following sources of funds for business-type risky ventures are described:
-   Seed money, or start-up enterprise investments (the so called private or informal "3Fs" investments);
-   Sustained growth capital in the format of venture capital;
-   Corporate/state funding of innovations (here confined to R&D).

Then, gambling practices are presented. In the third part, questions related to methodology of investigation of the research questions and available data sets are discussed. Following these presentations, analysis of statistical data series is performed in an attempt to verify research hypotheses. The last section concludes the paper and outlines ideas for further studies.

2.   SOURCES OF FUNDING FOR RISKY INVESTMENTS AND GAMBLING

2.1.   PRIVATE INVESTMENTS

Private investment is a frequent and important source used to support risky ventures. Almost every new venture, from mom-and-pop convenience stores to Silicon Valley super-projects such as Google, start with an investment from the founders themselves, eventually supported by their friends, family, or 'business angels' ('3Fs' – family, friends, fools)[1]. These informal investors are vital to the start-up process" [5, p52]. Such investments are also referred to as

---

[1]   Bygrave and Hunt [7, p.17] provide the following patterns of relationships of informal investor to investee: close family – 49.4%; other relatives – 9.4%; work colleague – 7.9%; friend, neighbour – 26.4%; stranger – 6.9%.

informal investment. Such informal investments are motivated by: greed, personal satisfaction of being involved in business or helping others, tax advantages, way of life (e.g., [11, pp.16-17], [36, pp.258-263]).

Some features of private investments can be explained through references to 'business angels'. Angel Investors ('business angels'): ".... are individuals who invest in businesses looking for a higher return" [4, p.17]. A 'business angel' (angel investor) is usually a former entrepreneur or professional who provides starting or growth capital in promising ventures, and helps also with advice and contacts. Unlike venture capitalists, angel investors usually operate alone (or in very small groups) and play only an indirect role as advisors in the operations of the investee firm. They tend to follow quite a typical outline of the investment procedure [38]. Tamowicz [39, pp.17-18] has suggested investors be into three categories: economic investors - who invest mainly for profit reasons and accept the most risky initiatives; hedonists – who also expect higher returns, but also seek pleasure from doing business; and altruists – who take into account social benefits of investments as well. This group is the smallest. Another overview of motives has been presented by Mason [24, p.28], who claims that the prime motives of most investors deal with the possibility to quickly earn high returns, get satisfaction from involvement in business, gain current and future income, yet also benefit from favorable tax regulations.

The European Union income tax provides a 20% relief, up to $600,000, 40% on failed investments for providing support to innovations [4, p.17]. In Canada, Capital Gains Exemption for angel investors exists, along with a 30 % tax credit [13, p. 8]. However, this practice leads to abuse concerns due to a "70:30 risk sharing ratio between equity investors and government" [13, pp.6-8]. Most U.S. states offer a 25%-50% tax credit, plus individual caps of $2-8mln. in some cases [43, pp.4-5]. According to Mason and Harrison [25, pp.211-236], in 40% of cases such businesses experienced negative returns on their investments and in 10% of instances the return was higher than 100%.

In Europe 75,000 'business angels' have invested some €3 bln. for start-ups, and 250 networks [12, p.5]. Sohl [37, p.2-17] has estimated that there are some 300-350 thousand angel investors, who invest $30 bln/year, in about 50 thousand initiatives. Estimates by Tamowicz [39, p.23] indicate 250 thousand investors, investing $10-30 bln., in 30 thousand firms. In the UK there were 20-40 thousand investors, with some £ 0.5-1 bln., in 3-6 thousand projects [26, p.11]. Tamowicz (2007) claims this underestimates of the reality of investments made. Next, business angels experience problems in identifying projects worth supporting. The proportion of active business angels to those who may become active is 1:5 (1:3 in the 90s, 1:8 according to the European Commission) [39, p.23].

It is widely recognized that there is a funding gap for early-stage technology development and that markets for allocating risk capital to such activities are not efficient. The term 'Valley of Death' has been used to dramatize the particular challenges facing entrepreneurs engaged in the transition from invention to innovation [31]. "Early stage companies often experience a situation where they are making profits, but have a negative cash flow because of the need to invest in equipment and personnel" [23, p.76].

The solution adopted by the US government related to some up-front investment in innovative ventures is similar in terms of focus to that used by 'business angels'. The Small Business Innovation Research (SBIR) program and the related Small Business Technology Transfer (SBTT) program have been pointed out as particularly interesting in terms of commercializing R&D. Through the program, a specific percentage of federal R&D funds are reserved for small businesses (this may amount to $3 bln.). Federal agencies with external R&D obligations above $100 mln. must set aside 2.5% for SBIR projects. Ten agencies participated in FY 1999. Each year, the departments and agencies required to participate, designate R&D topics and accept proposals. They are responsible for releasing solicitations, evaluating proposals and awarding SBIR funding agreements on a competitive basis. Proposals are typically evaluated along three dimensions: agency importance (the ability to meet federal R&D needs), commercial importance (the ability to transform R&D into commercially viable products), and technology leadership in the science and technology of the applying firm (such as expertise, facilities and experience). Innovations that have been patented or, have patents pending, are not considered under the program – the focus is entirely on new innovations [23, p.66]. Evaluations of these procurement programs have been positive and they are considered to play an important role in the overall technology transfer effort. Another program is the Small Business Investment Companies (SBIC) that has been a significant source of venture capital in the U.S. [23, p.12].

The U.S. has a strong entrepreneurial culture. The U.S. ranks high when it comes to start-up experience and the number of people working in newly formed companies. It is often pointed out that Americans are more willing to take risks compared to people in other countries and that it is easier to start over again after a business failure. This culture has developed in the environment of well-developed entrepreneurial education at universities and colleges.

2.2.   VENTURE CAPITAL

Successful start-ups seek venture capital support if necessary. Venture Capital is a somewhat formalized means. Venture Capital[2] is a fund-raising technique for startup firms and small businesses with exceptional growth potential (so high gain or high loss) who are willing to exchange equity in the company in return for money to grow or expand the business [22]. DataMerge [10] states that VC investments in an enterprise are usually between $500,000 and $5 mln. and that the investor is likely to expect an annual return of 20% to 50%. Managerial and technical expertise is often provided as part of the agreement. A venture capitalist differs from an angel investor in terms of wanting greater control of a company and quicker return on investment [1].

In recent years, approximately .2% of GDP, has been invested annually in U.S. companies in the form of venture capital (by the early 90s it was approximately .06% of GDP). About £7 bln. was invested in European companies (approximately 0.05% of Europe's GDP). Whereas

---

[2]   Venture capital was once known as risk capital. This term has fallen out of usage. It may be expected that investors do not like to see the words risk and capital in close conjunction. They often tend to think that though their investments involve an element of risk, they are assured of a successful return by virtue of their knowledge and business sense.

it may be reasonably assumed that 'informal investment' may have a mainly 'entrepreneurial drive', venture capital is invested for profit reasons, and is associated with slightly lower risk, and is normally arranged within a tight legal context. In all of Europe, 549 seed stage companies received venture capital in 2008" [5, p.55]. In Sweden, venture capital is primarily provided by large institutions. Yet, a variety of government support systems has been established to compensate for the lack of private capital. Funding for R&D is coordinated through such agencies as NUTEK or VINNOVA. Technology Link Foundations was established in the mid-90s to provide early-stage funding that supplemented private investments [23]. Companies can also apply for regional investment subsidies to the Swedish Business Development Agency [32, p.2]. These subsidies are available to companies provided they support the company's needs for tangible and intangible assets such machinery and/or patents. In the U.S., "…in the last 40 years no more than about 30.000, or about one in one thousand, have ever received venture capital." When viewed in another way, only 1,170 U.S. companies received their first round of venture capital in 2008, and of those, only 330 were seed or start-up stage companies. The US Small Business Investment Company program is an example of leveraging private investments with government funding [23, p.83].

The U.S. venture capital market is the largest and most developed in the world and has played a crucial role in the formation of many new high-technology companies. According to Karlsson [23, p.53], over 80 percent of the entire world's venture capital is invested in the U.S. 'Angel investors' provide the most significant source of funding for individual technology entrepreneurs and small technology start-ups, while venture capitalists prefer to support firms that have at least proceeded beyond the product development stage. Generally, statistics regarding money invested differ substantially between sources. Price Waterhouse Coopers Money Tree Report (PW) [40, p.2] indicated that combined investment by venture capitalists and angel investors accounted for $30.7 bln. in 2007, however, there has been a historical downward trend since 2000. Comparing first quarter 2000 venture investment in business at $28.4 bln. and less than a decade later in first quarter 2008 at $7.4 bln., there has roughly been a 75% decrease in venture investment in the USA. The decrease may be related to the collapse of the dot-com industry, which was one of the main objects of such investment. Following estimates by Karlsson [23, p.10], venture investment was $94 bln. in 2000, then radically dropped to $19 bln. in 2002, as focus has turned to less risky, later stage investments. The difference in estimates of venture capital investments may also result from a different sample included in the Price Waterhouse Report and Karlsson sample, as well as from a departure from on-line type business support. The observations regarding a sharp decline in VC investments have also been confirmed by Ghalbouni & Rouzies [16]. However, the regulatory environment in the US has promoted the growth of investments in high-tech companies by 'business angels' and venture capitalists to an extent unmatched by any other county. A combination of tax reforms, banking and bankruptcy laws and pension fund regulations have facilitated private capital accumulation and increased the willingness to invest in high-risk enterprises [23, p. 83].

2.3.   CORPORATE FUNDING OF R&D

Corporate and state investment in R&D can be regarded as indicators of a tendency (propensity) to invest in risky ventures. Quotes related to direct investments in innovations are not available. However, both state and business R&D expenditures may be used as a proxy of

expenditures in risky projects (e.g., [46, Table 4.3.01 – 4.3.05]. European countries intend to increase levels of direct funding with the goal of reaching the Lisbon Strategy goals of 3% of GDP. There are also national programs through which the countries are funding new ventures, innovations, and technology development. The nature and scope of these programs vary from one country to another. One can account indirect funding, through improving business's access to information, education, infrastructure, as well as access to tax incentives for supporting innovation, yet systematic statistics regarding such issues has not been identified.

2.4.   GAMBLING

There are no broadly accepted definitions of gambling – yet, it is "playing games of chance for stakes". However, data provided in the report are restricted to the following gambling activities: "casinos, gambling machines, lotteries, charitable betting (Bingo, etc.), other betting (including horse racing)" [28] (e.g., horse betting, dogs racing, chicken- fights, dog-fights, sports-betting are not included). Gambling (or betting) is any behavior involving the risk of money or valuables on the outcome of a game, contest, or other event in which the outcome of that activity is partially or totally dependent upon chance – it is an act or undertaking of uncertain outcome [41]. Gambling can refer to casino gambling, fixed-odds gambling, non-casino gambling games, gambling on horse races, sports betting, e-gambling. Gambling may also refer to engaging in any high-risk behavior in which decisions are made based upon incomplete knowledge, e.g., high-risk stock investments, difficult and potentially costly ventures, or even personal relationships [44].

In Poland, some US $5 bln. was spent n gambling in 2009 (a 70% increase over 2008) [20]. This translates to some $131 per capita/year as compared to $16 per capita/year spend on business R&D (!!!). "In Canada gambling continues to expand. In 2002, an estimated 18.9 mln. adult Canadians wagered $11.3 bln. on everything from VLTs, lottery tickets and bingos to blackjack and slot machines in casinos. This amount was more than a four-fold increase from $2.7 bln. a decade earlier … About 62% of problem gamblers spent more than $1,000 a year in gambling, compared with only 4% of people who gambled with no problem" [8]. According to the American Gaming Association "gross gambling revenue (GGR) is the amount wagered, minus the winnings returned to players, a true measure of the economic value of gambling. GGR is the figure used to determine what a casino, racetrack, lottery or other gaming operation earn before taxes, salaries and other expenses are paid" which amounted to $92.3 bln. in 2007 [2]. In Bulgaria, taxes paid by "gambling" related companies are higher by some 20% than those by banks in 2009 -- gambling will bring some €51 mln. whereas banks €40 bln. The decline in budgetary income from gambling (as compared to €57 mln. in 2008) is attributed to the move to the e-gambling mode, and not the economic crisis [42]. In the Ukraine, hazard started to be expelled from some locations by a Parliamentary decree (some 200 000 of gambling machines were demolished). In June 2009, Putin's decree banned hazard, whereas Berlusconi opened the market for gamblers in Italy. The procedures to open a casino in Italy are much more relaxed compared to those in the US [34]. The UK and Spain seem to be the most liberal in Europe regarding regulations in the gambling sector. In the UK 'one-hand-bandits' and e-gambling are allowed anywhere subject to getting a license, which is not a complex process. E-gambling in Poland is illegal – however, it is not difficult to open a gambling domain in a country where e-gambling is allowed.

In 2006, global gambling revenue was $101.6 bln., and is estimated reach $144 bln. in 2011. Current U.S. revenue, which stands at about $57.5 bln. will grow an estimated 6.7% per year and reach $79.6 bln. The gambling markets of Europe, Middle East and Africa will grow 1.9% each year, from $25.2 bln. to $27.8 bln. The Asian market is growing even more rapidly and will grow 15.7% every year, which will transform the current $14.6 bln. market to $30.3 bln. This will make it the second-largest gambling market in the world (all based on Hopkins, 2007). Also, e-gambling is booming: a 10% increase in Europe, with an income of U.S. $16.3 bln. It is expected that the e-gambling sector will increase income in the US to $24,4 bln. by 2012 [6].

3.   METHODOLOGICAL CONSTRAINTS OF THE REPORT

The paper can be regarded as a concise literature overview on private business investment and gambling, with the main objective being to identify similarities and differences between business – hazard in terms of expenditures, legal/organizational aspects. Yet, the core is related to the examination of statistical interrelationships between the two, and hence, data used are of critical importance. Also, there is the problem in making the distinction between invention [15] and innovation [33], [30]. For the purpose of further argument, and keeping in mind that these notions are not clearly distinguished when 'money' is concerned, these notions will be used as synonyms, which can be considered somewhat incorrect. However, within the constraints of this initial study it is accepted. There is a pervasive problem with access to reliable data about private and venture investments , interpretation of these data, identification of what indeed should be measured [17], [3], [27]. These types of problems have been experienced during the study, yet no attempt to couple with them has been attempted.

Data for venture capital were taken from Bosman and Jonathan [5, pp.53 & 56] and recalculated to per/capita values. Data about private investments are not easily available[3]. Dataseries on these activities come mainly from private sources and are costly to obtain. There is no information about methods for data collection, nor, for example, can it be identified whether values are given in nominal dollar values or are recalculated to PPP values. The identification of the year the data was collected is troublesome.

---

[3]   There has been a strong temptation to include data about private investment into the examination of research hypotheses. However, data on family/friends investments are fragmented, incomplete, and thus, quite unreliable. They cannot be supported by official statistical data (also because of the "grey zone phenomenon"), and the data for 'business angels' deal with estimates or incomplete surveys (e.g., [11, pp.4-5]. There are some statistics on 'business angels' investments – e.g., provided by EBAN (European Business Angels Network), GEM (Global Entrepreneurship Monitor). These data series are very inconsistent. Some are based on sources from private institutions and subscription to these sources exceeds the financial constraints of this study. Some sources record capital invested, others capital available, and some deal with different years. The use of surveys does not guarantee that the key portion of investment has been captured. Even if they deal with the same item, the difference between sources is quite big. This renders any meaningful statistical analysis fruitless. Therefore, the examination of financial involvement by "business angels' was dropped.

Data for gambling were taken from MGF [28] and recalculated to per/capita values – since two cross-sections were provided the average is used. This source provides information about gambling revenues in Russia, whereas as of June 2009, gambling is illegal in Russia. The same source does not provide data about the situation in Bulgaria where gambling is a sort of industry with income higher than that of banks. In the case of gambling, the reality of expenditures is strongly affected by country specific regulations, thus preventing an examination of the propensity of citizens to gamble. E-gambling, along with illegal and 'home' gambling is gaining popularity, and remains a sort of a 'grey-zone'. Data are further deflated by the fact, that many gamblers are not the citizens of the given country. In the case when e-gambling is forbidden in a given country, the domains can be registered in another country, where it is legal.

Data for perceptions of investment climate are based on WCY [46] – regarding availability of venture capital (tab. 3.3.18); whether public and private ventures are supporting technological development (tab. 4.2.16), ease of doing business (tab, 2.4.13), entrepreneurship attitudes (tab. 3.4.07), wealth (tab. 1.1.22).

Under the circumstances, data used must be regarded as general estimates. They do not provide a very valid reference, and hence the results arrived at in this study must be treated with caution. The approach presented in this paper can be interpreted as an outline for further studies, to be carried out with an expectation of greater precision, when data needed for such a study become more reliable, timely, and more easily available for research purposes.

4. VERIFICATION OF RESEARCH HYPOTHESES

Data used to verify research hypotheses have been provided in Table 1. Examination of the Pearson correlation coefficient (PCC) was the prime means to verify hypotheses.

As was expected (hypothesis 1), in countries where it is expected that "it is easy to do business" there is a higher level of entrepreneurship (PCC= 0.499, $p<0.001$), funds from public/private sources to support technological development is available (PCC=0.724, $p<0.001$), and there are more VC funds available and invested (PCC=0.738, $p<0.001$). As well, in such countries expenditures higher are in R&D (PCC=0.488, $p<0.001$). Perceptions of higher levels of entrepreneurship are closely associated with the availability of funds for technological development (PCC= 0.534, $p<0.001$). More VC capital is invested in countries where private/public ventures support technological development (PCC=0.380, $p<0.001$). Opinions that private/public funds are available for technological development (PCC=0.518, $p<0.001$), that it is easy to do business there (PCC=0.675, $p<0.001$), and VC is available (PCC=0.595, $p<0.001$) are characteristic of countries with a low uncertainty index.

Table 1   Data used to verify research hypotheses

Country	A	B	C	D	E	F	G	H	I	J
Argentina	13775	9.9	4.6	4.2	2.2	1.5	103.4		84.9	86
Australia	34935	422.9	6.2	5.7	6.0	3.8		3.8	628.5	51
Austria	37913	802.6	6.8	6.5	5.1	3.9		7.6		70
Belgium	35183	556.3	6.2	5.3	4.1	4.3	421.7	14.8		94
Brazil	10041	35.9	5.0	6.7	4.0	3.3	10.0		21.7	76
Bulgaria	11923	5.0	5.2	5.2	4.4	3.5				
Canada	35669	441.6	7.1	5.7	5.4	4.6		29.2	446.0	48
China	5825	25.8	5.7	5.3	4.9	3.4	658.2	5.0	3.9	
Denmark	35677	937.0	7.6	6.2	7.1	4.3	355.8	24.9		23
Finland	34750	1165.7	8.1	5.8	6.6	4.9	138.9	24.3	304.9	59
France	33693	548.1	5.1	4.8	4.3	3.9	252.6	18.5	191.0	86
Germany	34907	714.0	6.8	5.5	4.7	4.1	261.8	14.0	136.2	65
Greece	29280	43.1	4.7	5.5	3.0	2.8	366.5	3.2	278.1	112
Hong Kong	43040	112.0	6.5	6.9	7.4	5.7	967.0		356.1	29
Hungary	18868	67.1	5.0	4.4	4.3	2.1	94.5	4.2		
Ireland	42993	517.2	6.4	5.9	5.7	4.2		36.5		35
Israel	26642	819.7	7.4	7.0	5.3	5.3	385.6	273.9		81
Italy	30057	196.6	4.1	4.9	3.1	2.5	360.6	3.9	270.3	75
Japan	33403	896.6	6.5	4.4	5.1	4.1	183.7	18.0	277.1	92
Korea	25268	454.8	6.7	6.0	2.9	3.4	138.9	18.9	99.4	85
Netherland	39215	484.7	7.2	6.0	4.6	5.0	353.1	25.5	169.8	53
N. Zealand	26985	125.2	5.1	5.5	5.5	3.9	9.5	169.0	49	
Norway	53481	652.3	7.2	5.5	5.2	4.0	267.7	32.1		50
Poland	16747	15.7	4.7	5.4	2.4	3.7		2.5		
Portugal	22781	127.6	6.1	5.0	5.3	2.9		9.1	177.1	104
Romania	13271	17.3	4.3	4.0	3.7	2.9	132.6	3.3		
Russia	15524	65.6	4.8	5.3	1.9	2.1	15.5		29.8	
Singapore	47646	428.3	7.9	5.9	7.9	5.4			558.6	8
Slovenia	27941	221.5	5.2	6.2	4.5	3.2	350.0	2.8		
SouAfrica	9887	26.6	5.4	5.2	3.8	4.2	98.8	43.5	91.2	49
Spain	31053	178.2	5.2	4.3	3.4	2.2	170.7	20.5	255.2	86
Sweden	36180	1296.2	7.6	5.5	5.2	4.8		54.3	274.0	29
Switzerland	41068	1024.0	7.2	6.2	5.9	5.0	411.1	18.5	358.4	58
UK	35347	431.5	6.0	4.8	4.4	3.6	141.4	31.8	248.7	35
USA	45017	858.7	7.1	5.8	4.7	4.8	630.3	88.2	312.1	46
AVG.	29600	421					291	29	239	

Explanations: column: A – PPP (Purchasing Power Parity) (in $US); [46, Table 1.1.22]; B – R&D expenditures per capita (in $US; [46, Table 4.3.04]; C - public and private ventures support to technological development (index) [46, Table 4.2.16]; D – entrepreneurship (is widespread in business) (index) [46, Table 3.4.07]; E – ease of doing business (index) [46, Table 2.4.13]; F – availability of venture capital (index) [46, Table 3.3.18]; G – amount of informal capital per capita (in $ US) (based on GEM [7, Figure 39]; H – venture capital investments per capita (in $ US) (based on GEM [7, Figure 42]; I – gambling expenditures per capita (in $ US) (based on MGF 28, Table 4 & 5]; J – uncertainty avoidance (index) [18].

For hypothesis 2 – gambling expenditures/revenues are closely linked to the perception of ease of doing business (PCC=0.475, p<0.001). Countries where it is easy to do business (PCC=0.518, p<0.001) and where venture capital is available (PCC=0.675, p<0.001) show low levels of uncertainty avoidance[4]. Characteristically, the higher are expenditures for gambling (PCC=0.466, p<0.001), the lower is uncertainty avoidance.

5.   OBSERVATIONS

Current results, keeping in mind the quality (precision) of available data, do not render it possible to draw definite conclusions regarding interrelationships between conditions of doing business, a selected characteristic of national culture, 'risky' investments and gambling. Statistical sources on number of ventures, amount of money invested, or gains are not easily available, and are very fragmented: these are only very broad indicators that do not facilitate drawing firm conclusions. 'Business angels' do exist and help start up ventures, even though it seems to be a minor portion of investment. Venture capital investment takes a more formalized format than that of 'business angels', and is a much higher amount. To this end statistics indicate an interesting pattern – availability of venture capital is positively correlated with the entrepreneurship drive of the society and income per capita. Thus, it would be worthwhile to further explore how to enhance private investors' participation in, and contribution to, our well being. These items are linked to the propensity of the society to gamble: high level of risk with a potential for a high pay-off. Further examination of such a pattern may have implications to formulating public policies.

The systematic sources of information for funding innovative ventures have not been identified. Sources regarding private financing are fragmented in some countries only, and methods for collection and reporting these data have not been verified. It was expected that this study would identify, in a quantitative format, some patterns related to different forms of financial support to risky ventures. Also, the outcomes of this report were expected to shed light on results of commercialization outcomes and investment in innovation upon contribution to financial/social benefits. However, with data-related methodological quandaries these objectives could not be achieved to the full extent at this stage of the study.

Interestingly, although the amount of money spent on gambling in Poland, Canada, and the U.S. are similar (as percentage of income), in Poland, 8 times less is spent for business R&D, than for gambling. In the U.S., spending for R&D is only 10 times higher than for gambling. Yet, on average, people spend about the same amount of money for R&D (investment for future benefits) as for alcohol!!! Almost the same amount is spent on gambling, restaurants, entertainment drugs and movies!!! [28,table 1]. Such an observation suggests that much more effort is needed in some countries to promote innovativeness, instead of advocating current unproductive, and probably excessive, entertainment drive.

---

[4]   Uncertainty avoidance is the extent to which a society feels threatened by uncertain and ambiguous situations and tries to avoid these situations by providing greater career stability, establishing more formal rules, not tolerating deviant ideas and behaviors, and believing in absolute truths and the attainment of expertise (adopted from G. Hofstede [18]).

Private and business investments in innovation are generally higher in richer countries, as is the ease of doing business and the entrepreneurial drive. As well, venture capital is more readily available in more rich countries. This suggests that wealth may stimulate more drive for further achievements. This observation does not contradict an assumption that even poor countries may be very innovative (not necessarily inventive).

REFERENCES

[1] About.com: Small Business Information, http://sbinformation.about.com/od/creditloans /l/bldef_venture.htm ; retrieved on Nov. 5, 2010.

[2] America Gaming, 2009 (http://www.americangaming.orgIndustry/factsheets/statist icdetail. cfv?id=7 : retrieved on Jan 18, 2010.

[3] Avdeitchikova, Sofia,Landstrom, Hans & Månsson, Nils, 2008, What do we mean when we talk about business angels? Some reflections on definitions and sampling, Venture Capital: An International Journal of Entrepreneurial Finance, Vol.10, No 4, October, pp. 371-394.

[4] Bloński Jacek, 2009, Business Angels - An Alternative Source of Financing Innovative SMEs, EBAN (http://www.unece.org/ceci/ppt_presentations/2009/ic/blonski.pdf, http://belisa.org.by/pdf /2009/Blonski.pdf - retrieved on Jan 18, 2009.

[5] Bosman Niels & Levie Jonathan, 2009, Global Entrepreneurship Monitor, Global Report, GEM-GERA.

[6] Bójka o e-hazard, http://www.forum.bukmacherskie.com/f250/bojka-o-e-hazard-37342.html : retrieved on Nov. 4, 2010.

[7] Bygrave W.D. & Hunt S.A., 2004, Global Entrepreneurship Monitor, Babson College – London Business School.

[8] CanGam, 2009, http://www.statcan.gc.ca/daily-quotidien/031212/dq031212c-eng.htm : retrieved on Jan., 18, 2010.

[9] Cooper Robert G., 1993, Project NewProd: Factors in New Product Success, European Journal of Marketing, Vol. 14 Issue 5/6, pp.277 – 292;

[10] DataMonitor, 2001, definition of VC; http://searchcio-midmarket.techtarget.com /sDefinition/ 0,,sid183_gci 519528,00.html; retrieved on Nov. 12, 2010.

[11] EBAN, 2008, Statistics compendium, EBAN Secretariat, Brussels.

[12] Ebancongress 2008, Editors: Stephanaie Floor-Glaasker, Claire Munck, Ben Lacor, Paul Janmaat, Keith Wallace, Business angel investing: Entrepreneurial Capital for the 21st Century, Breukelen, The Netherlands.

[13] Fletcher Jim & Gitelman Bruce, 2005, NAO Answers Seven Key Questions About The Innovation & Productivity Tax Credit, National Angel Organization, Toronto, pp. 4 & 6 ; http://www.Angelinvesttor .ca/userfiles/file/IPTC/Supporting%20Docs% 20for% 20IPTC/IPTC%20Seven%20Questions.pdf.

[14] Frese Robert & Sauter Vicki, 2003, Project success and failure: What is success, what is failure, and how can you improve your odds for success?; http://www.umsl.edu/~sauterv /analysis/6840_f03 _papers/frese/ retrieved on Nov. 7, 2010

[15] FM, 2002, Frascati Manual: The Measurement of Scientific and Technological Activities, OECD, Paris.

[16] Ghalbouni, Joseph & Rouziès, Dominique, 20101, The VC shakeout, Harvard Business Review; Jul/Aug2010, Vol. 88 Issue 7/8, pp. 21-23.

[17] Harrison, Richard & Mason, Colin, 2008, Sampling and data collection in business angel research, <u>Venture Capital: An International Journal of Entrepreneurial Finance</u>, Vol. 10, No 4, October, pp. 305-308.

[18] Hofstede, G., 1984, <u>Culture consequences: International differences in work-related values</u>, Beverly Hills, Sage Publications.

[19] Hopkins Jeremy, June 25, 2007, Global gambling revenue to climb to $144 billion by 2011- retrieved from http://news.777.com/2007-06/gambling-revenue-climb-to-144-billion : retrieved on Nov. 3, 2010.

[20] Ile Polacy wydaja na hazard i ile zarobi na tym państwo, RMT, 1 październik, 2009, Interia.pl – Fakty.

[21] The Free Dictionary by Farlex; http://www.thefreedictionary.com/gambling ; retrieved on Nov. 5, 2010.

[22] InvestorsWords.com; http://www.investorwords.com/5236/venture_capital.html#ixzz14 ErWxqbe ; retrieved on Nov. 5, 2010.

[23] Karlsson, 2004, <u>Commercialization of research results in the United States</u>, ITPS – Swedish Institute for Growth Policy Studies, Embassy of Sweden, Washington, DC.

[24] Mason Colin, 2002, Barriers to investment in the informal venture capital sector: in Tamowicz Piotr (2007).

[25] Mason Colin & Harrison Richard, 2002, Is it worth it? The rates of return from informal venture capital investments, <u>Journal of Business Venturing</u>, Vol 17: in Tamowicz Piotr (2007).

[26] Mason Colin, 2006, Informal sources of venture capital: in the <u>Life Cycle of entrepre-neurial ventures</u>, Vol. 2, International Handbook of Enterpreneurship, Kluwer.

[27] Mason, Colin & Harrison, Richard, 2008, Measuring business angel investment activity in the United Kingdom: a review of potential data sources, <u>Venture Capital: An International Journal of Entrepreneurial Finance</u>, Vol.10, No 4, October 2008 , pp. 309-330.

[28] MGF, 2010, <u>Mors Global Finance – Economics of Gambling</u>; www.morssglobal finance. com/ the-global-economics-of-gambling/ - retrieved on Nov. 3, 2010.

[29] Mors Global Finance – Economics of Gambling, 2010 p.2; www.morssglobalfinance. com/ the-global-economics-of-gambling/ : retrieved on Nov. 3, 2010.

[30] Nasierowski Wojciech, 2008, Mierniki i indeksy innowacyjności, <u>Ekonomika i Organizacja Przedsiębiorstwa</u>, No 2 (697), pp.19-28.

[31] NIST, 2002, An analysis of Funding for Early-Stage technology Development ; http://www.atp.nist.gov/eao/gcr02-841/gcr02-841.pdf : retrieved on Jan 16, 2010.

[32] NUTEK, 2007, Financial Incentives, Companies can apply for regional investment sub-sidies, when investing in municipalities categorized as "Development area"; http://www.isa-northamerica.org/literature/financial_incentives.pdf : retrieved on Jan 17, 2010.

[33] OM, Oslo Manual, 1999, <u>OECD proposed guidelines for collecting and interpreting technological innovation data</u>, OECD/EuroStat, Paris (KBN 1999).

[34] Rożek T.Ł., Wyspy to prawdziwy raj dla miłośników hazardu: http://www.polskatimes.pl/ fakty/ 179472,wyspy -to-prawdziwy-raj-dla-milosnikow-hazardu,id,t.html?cookie=1, retrieved on Nov.3, 2010.

[35] Salim Chahine, Salim, Filatotche, Igor, & <u>Wright</u> Mike.,2007, Venture Capitalists, Business Angels, and Performance of Entrepreneurial IPOs in the UK and Franc, <u>Journal of Business Finance & Accounting</u>, 2007-04, vol. 34, Iss. 3-4, pp.505-528.

[36] Sherb L., Rappai G., Makra Z., Terjesen S., 2007, Informal investment in transition economies: Individual characteristics and clusters, <u>Small Business Economics</u>, Vol 28, pp.257-271.

[37] Sohl Jeffrey, 2003, The US Angel and venture capital market: Recent trends and devel-opment, <u>Journal of Private Equity</u>, Vol. 6, No.2 pp.7-17.

[38] Stuart, Paul, Geoff, Whittam Geoff & Wyper, Janette, 2007, Towards a model of the business angel investment process, <u>Venture Capital: An International Journal of Entrepreneurial Finance</u>, Vol. 9, No 2, April, pp. 107-125.

[39] Tamowicz Piotr, 2007, <u>Business Angels – pomocna dłoń kapitału</u>, PARP, Warszawa.

[40] Thomson Financial, 2008, PricewaterhouseCoopers and National Venture Capital Association, <u>Money Tree Report</u> (https://www.pwcmoneytree.com/MTPublic/ns/moneytree/ file-source/exhibits/ MoneyTree%20%20Q2%202008%20final.pdf : retrieved on Jan 17, 2010.

[41] The Free Dictionary by Farlex; http://www.thefreedictionary.com/gambling ; retrieved on Nov. 5, 2010.

[42] W Bułgarii wpływy z hazardu do budżetu wyższe niż z banków**,** Source PAP, 2010-03-01; retrieved from http://biznes.gazetaprawna.pl/artykuly/402335,w_bulgarii _wplywy_z_ hazardu_do_budzetu_wyzsze_niz_z_bankow.html on Nov. 3, 2010.

[43] Williams Jeffrey, 2008, Tax Credits and Government Incentives for Angel Investing in Various States, Belmont University, Angel Capital Education Foundation; http://www.angel capitaleducation.org: retrieved on Jan 17, 2010.

[44] World.IQ; www.wordiq.com/definition/Gambling; retrieved on Nov. 3, 2010

[45] VINNOVA, 2003, <u>Commercialization of Academic Research Results</u>, VINNOVA Forum VFI 2003:1.

[46] WCY, 2009, <u>World Competitiveness Yearbook</u>, IMD, Lausanne.

IMPACT OF REDUCTION IN TRADE BARRIERS ON INTERNATIONAL ENTREPRENEURSHIP: EUROPEAN UNION DEVELOPMENTS AND THE '*GRANDE REGION*'

by Theo Stengelhofen
ICN Business School Nancy-Metz – CEREFIGE
3, place Edouard Branly
F-57070 Metz
Theo.Stengelhofen@icn-groupe.fr
+33 3 87 56 37 26

ABSTRACT

By its development to an Economic Union, the European Union has attempted to facilitate cross-border business activities within its member countries. The access to a larger market as well as a larger resource-base should encourage companies to do business outside their home country borders, and therefore push economic growth and prosperity, and overall improve European wealth.

This paper will address the international activities of entrepreneurs and small-and-medium-sized enterprises in the '*Grande Region*' between Belgium, France, Germany and Luxembourg before and after the removal of European trade and investment barriers. It will be shown that more SME's of the region have gone international, as well as an increase of 'born-global' firms originating from the region. Nevertheless, not all internationalisation efforts are market-driven, and remaining obstacles for internationalisation within the European Union will be highlighted.

KEY WORDS

International Entrepreneurship, Small- and Medium-Sized Companies, European Union, Organisational Knowledge and Learning, Corporate Networks

INTRODUCTION

By its development to an Economic Union, the European Union attempts to facilitate cross-border business activities within its member countries. The access to a larger market as well as a larger resource-base should encourage companies to do business outside their home country borders, and therefore improve economic growth and prosperity, and overall European wealth. This paper will address the international activities of entrepreneurs and small and medium-sized enterprises in the 'Grande Region[1]', a region connecting Belgium, France, Germany and Luxembourg, before and after the removal of European trade and investment barriers. A concept based on networking and knowledge acquisition will be provided to explain why more SME's of the region should have gone international, including

---

[1] The English translation of the Grande Region would be 'Greater Region'. Nevertheless as French and German are the main languages of the region, the English term is never used. In English publications is the region is sometimes referred to as Euregio SaarLorLuxRhein [23, p. 161]. For the reader's convenience the French title is applied throughout the text.

an increase of 'born-global' firms originating from the region. Entrepreneurs in the Grande Region should have a higher motivation, better knowledge and closer networks allowing them to internationalize with less problems and higher speed. Nevertheless not all internationalisation efforts are market-driven, and remaining obstacles for internationalisation even within the European Union still exist.

The deficits in comparable and dynamic research do not so far allow a complete empirical validation of the concept. A multi-country multi-industry observatory of small- and medium-sized companies including new ventures is needed. With regular repeated surveys of SME's in the region, further insights into the internationalisation process of SME's will be gained.

BACKGROUND

On November 1st 1993, the European Union was formally established by the Maastricht Treaty [13]. After enlargements in 1995, 2004, and between 2007 and 2009 the EU consists today of 27 countries with over 500 million citizens, while generating more than 20% of the world's gross domestic product. The Maastricht agreement opened to the member countries an Economic Union, ensuring the free movement of goods and services as well as capital and labour, under, since 2002, a common currency (the Euro). This Economic Union should be achieved by strong harmonisation efforts in fiscal, monetary and social policies, incl. taxation and labour law, while developing commonly accepted product and service standards. Based on these developments political and economic representatives of various border regions within the European Union started to discuss options for political, economic, social and cultural cooperation, as they understood that their future might depend more on the economic development of and with their direct neighbours than on the nation state to which they belong [23 and 22]. One of these regions is the *Grande Region*, already established in 1980. Here cooperation of the following regions has been reinforced: Wallonia in Belgium, incl. the German speaking Community, Lorraine in France, Rhineland-Palatinate and Saarland in Germany, and the Grand Duchy of Luxembourg. The Grande Region today includes 11.2 million inhabitants, 2.5% of the EU total population. Assuming the European Union initiatives have achieved their intended results, the following hypotheses should be valid:

H1: The internationalisation of SME's should be higher in the *Grande Region* than in non-border regions.

H2: The internationalisation of SME's should have improved significantly after the Maastricht Treaty.

H3: The number of 'born-globals' should be higher in the *Grande Region* than in non-border regions.

H4: Internationalisation takes place earlier in the development of SME's in the *Grande Region* than in non-border regions.

The validation of these four hypotheses is an important prerequisite for the following analysis. If they can not be validated, serious questioning of the European Union initiatives will become necessary.

INTERNATIONAL ENTREPRENEURSHIP LITERATURE

The most commonly used definition of international entrepreneurship is the one proposed by OVIATT and MCDOUGALL [21, p. 540], as: "… the discovery, enactment, evaluation, and exploitation of opportunities – across national borders – to create future goods and services". Nevertheless, taking into account recent publications on the subject, three perspectives, instead of two mentioned by OVIATT and MCDOUGALL [21], can be distinguished. First, the entrepreneurial activity in different countries is analysed and compared to identify institutional, economic and cultural factors explaining higher or lower entrepreneurial activity (as example see FREYTAG & THURIK [11]). This does not include only all comparative, but also single country studies, which allow comparisons to one's own home country. Second, focusing on personal characteristics of the entrepreneurs which are carrying out their activities in a cross-national context, the impact of Transnational Entrepreneurs (TE), Returnee Entrepreneurs (RE), and Ethnic Entrepreneurs (EE) on their home and/ or host country is discussed (see for details DRORI, HONIG & WRIGHT [5]). All these forms are based on the idea of individual migration, either as cosmopolitans (TE), immigrants (EE) or migrants returning home after a period of educational and/or professional experience (RE). Third, IE is linked to the internationalisation of small and medium sized businesses, when the time of the first international activity is considered [27]. In its beginning IE was limited to the analysis of new ventures internationalizing from its inception (see literature analysis from KEUPP & GASSMANN [16]). ZAHRA, KORRI & YU [28] enlarged the definition, while analysing the internationalisation of new ventures, to companies of six years or younger. Other authors distinguish between 'born-globals' and 'late starters' [9] or 'born globals' and 'international new ventures [3]. KEUPP & GASSMANN [16, p. 602] propose that future IE research studies should focus "… on studying small firms that venture abroad or on small firms that start exporting".

This third dimension, the internationalisation of SMEs, will be the one used in this paper, knowing that an analysis of International Entrepreneurship in the *Grande Region* could also be used for the other perspectives, i.e. comparing the situation of entrepreneurs in Belgium, France, Germany and Luxembourg, or analysing entrepreneurial activities of Germans in Luxembourg, French in Belgium and so on. In line with the above mentioned definition of OVIATT & MCDOUGALL [21], the main aspect is the focus on the recognition and exploitation of an opportunity, here, the reduction and disappearance of trade and investment barriers within the European Union. Of course, once established as an international SME, the process of further internationalisation of the company is not finished, as with growing international activities the foreign experience will demand change and development of the mother organisation.

As internationalisation is not an end in itself, the immediate or delayed internationalisation of SME's in the *Grande Region* needs to be related to performance indicators to give meaning to the entrepreneurial decision. The range of performance indicators is wide, from pure survival to profit or sales and turnover increase.

DIMITRATOS, PLAKOYIANNAKI, PITSOULAKI & TÜSELMANN [4] while analysing the internationalisation of SMEs into the lead countries of their industry, recognise the importance of pro-activeness towards international opportunities, risk attitude and innovativeness. With this result, the authors [4, p. 589] argue "… that there is a distinguishable body of SMEs, apart from international new ventures, which can be

characterised as international entrepreneurial". The analysis here even enlarges this concept, while not focusing solely on lead countries of the SME's industry, when SMEs are internationalising, all this under the assumption that internationalisation improves the performance of the firm doing so. Hypotheses five to seven summarise this idea:

H5: International SME's in the *Grande Region* should have an above average survival rate.

H6: International SME's in the *Grande Region* should grow faster than average.

H7: Early internationalisation should help SME's to stabilise and reduce the probability of failure.

CONCEPTUAL FRAMEWORK

To separate and understand different factors influencing the internationalisation decision of entrepreneurs and SMEs a model based on the work of SHANE [24], OVIATT & MCDOUGALL [21] and WRIGHT, WESTHEAD & UCBASARAN [27] is created. SHANE [24, p. 4f.] providing a model of the general entrepreneurial process, the individual-opportunity nexus, concludes that "the entrepreneurial process involves the identification and evaluation of the opportunity; the decision, whether or not to exploit it; the efforts to obtain resources; the process of organizing those resources into a new combination; and the development of a strategy for the new venture." OVIATT & MCDOUGALL [21, p. 540ff.] have a similar idea when proposing a model of forces influencing the speed of internationalisation, e.g. the exploitation of the entrepreneurial opportunity. What the authors call 'speed of internationalization' actually integrates three dimensions, the time to the first foreign market entry, the country scope and the corporate commitment to its international activities. Enabling and motivating factors are the technology removing physical trade constraints and the pressure or encouragement from competitors. Nevertheless the entrepreneur herself plays a key role in the concept. Her perception of international business opportunities, her knowledge in how to work internationally and her network relationships are the mediating and moderating forces. WRIGHT, WESTHEAD & UCBASARAN [27, p. 1015] present seven different themes with regard to the internationalisation of SME's: timing, intensity and sustainability, the mode of foreign market entry, the influence of the domestic environmental context, the leveraging of external resources, the unit of analysis, and the effect on SME performance.

Taking the different ideas of these works as complementary, this paper will focus on three steps[2] during the internationalisation of SME's:
1. the identification and evaluation of an international opportunity,
2. the decision to exploit the international opportunity, and
3. the way how to exploit the international opportunity.

These steps are influenced by individual attributes of the decision-maker, e.g. the entrepreneur, and environmental factors, either in the home or host country. Their execution will have an impact of the performance of the new venture and/ or SME.

---

[2] The term steps could be misleading, as it indicates that thing happen one after the other and not simultaneously.

The study will therefore have an investigative section, the understand, for example, the reasons and processes of SME's in the *Grande Region* to internationalise and to compare the findings with existing results.

The International Opportunity

Based on research in International Business (see for example HILL [12]), the following drivers of Globalisation can be distinguished:

- technological developments removing physical trade constraints, e.g. new transportation means for people, products and services, allowing transport over greater distances, with higher speed, and overall lower costs,
- the removal of communication constraints, either through new technologies to overcome greater distances and different time zones, and through education and training reducing language barriers and cultural differences,
- the removal/ decline of trade barriers, either through bilateral, multilateral or regional integration agreements such as the European Union, NAFTA or ASEAN, and
- the increase in information and knowledge about foreign opportunities and developments, especially through education and media publications.

These factors have not only created increasing opportunities for large companies but also for small- and medium-sized businesses. Still, even if all factors have their importance, the consequences of a removal/ decline of trade and investment barriers for the internationalisation of SME's, especially for those companies with a close border location, has not drawn much attention in the literature so far. One of the rare works has been produced by MASUEL, VAN HEMERT & DE GROOT [18] analysing Dutch entrepreneurship across the border. They find that 56 % of Dutch SMEs export goes to and 51 % of Dutch SMEs import comes from the direct neighbours of Belgium and Germany, indicating the importance of the close border opportunity. Already SHANE [24, p. 48] argues that geographical mobility is an important factor for the discovery of entrepreneurial opportunities. One could add that geographical proximity might work in the same direction. Entrepreneurs living within an hour of three other countries, which is the case in the *Grande Region*, should become aware of the ease to do business abroad much faster than entrepreneurs in regions far of any direct international exposure. The day-to-day international contacts and experience entrepreneurs can gain in a border region should improve their knowledge about international opportunities and create relationships to foreign businesses, suppliers and/ or customers.

The Exploitation Decision

The recognition of an international opportunity, here doing business in the *Grande Region*, does not automatically lead to its exploitation. It is therefore necessary to indentify the reasons why some entrepreneurs take the opportunity, while others do not. The already mentioned study by MASUEL, VAN HEMERT & DE GROOT [18] indicates some interesting findings concerning the reasons not to internationalize:

- exporting Dutch SMEs report high cost of internationalisation process, existing legislation and price of products and services as main barriers to internationalisation,
- importing Dutch SMEs report the price of products and services as main barrier to internationalisation, and

- concerning cross-border collaboration, restrictions caused by taxes and regulation, cultural and language differences, and the wish to remain independent are the main barriers.

As the survey allowed multiple responses, especially the barriers 'cultural and language differences', 'insufficient knowledge and capacities within company', 'do not know', and 'lack of information about suitable partners' draw our attention [18, p. 32ff.]. LEONIDOU [17] analysed 32 empirical studies on barriers hindering small businesses to export. Within his category of internal barriers, 'limited information to locate/ analyze markets', 'inability to contact overseas customers', 'identifying foreign business opportunities', 'accessing export distribution channels', and ' unfamiliar exporting procedures/ documentation' had a very high or high impact for export-management decisions. ISTRATE, KALLAS & STOUGH [15] using data from the Global Entrepreneurship Monitor and the World Bank to rank countries according to 'entrepreneurship activity' and 'barriers to entrepreneurship', ask how international entrepreneurs entering a foreign market would overcome the barriers. In addition to technological and institutional barriers, they also identify the organisational knowledge filter as barrier to International Entrepreneurship. This includes, for example, managerial ability, experience with international business, flexibility and openness and the firm's density in networks and alliances [15, p. 17].

Barriers for SMEs to internationalize do not only exist in the potential host country, also home country factors could make the going international difficult (see HILL [12] on home countries trade and investment barriers). Nevertheless, the research mentioned above also indicates that the same barriers are not perceived in the same way by different entrepreneurs. The final decision to take the opportunity and when depends on the motivation to go international by the entrepreneur, her intercultural background and knowledge, and the level of risk she is willing to take. In addition, available support within an international business network will give the entrepreneur the assurance that her decision to go international will be beneficial to her.

How the exploit the opportunity?

The decision to exploit the international opportunity does not indicate what option to internationalize to use: export, selling rights through licensing/ franchising, co-operative agreements, or creating own subsidiaries. In SHANE's argumentation [24, p. 161ff.] taking an entrepreneurial opportunity demands the re-organization of resources into a new combination. Unfortunately research to what mode of entry the re-organisation of resources leads is missing. The only results available emphasize the importance of organisational innovation [20] and organizational learning [1] when operating internationally. O'CASS & WEERAWARDENA [20, p. 1330f.] distinguish technological (product and process) from non-technological (managerial and marketing) innovation, but their final measurement of managerial innovation is not clear. ANDERSON & SKINNER [1] outline different learning stages in the SME's internationalisation process, but leave open what mode of entry is chosen by the company/ entrepreneur.

Even when WRIGHT, WESTHEAD & UCBASARAN [27, p. 1019] recognize the 'mode of internationalisation' a as main theme in the analysis of the internationalisation of SMEs, most of the existing studies focus on one mode of entry, especially export [17 and 26], and 'born-globals' [3, 8, and 10], or only discuss 'international operations' without further details [1, 2, 9 and 20]. The only study dealing with four types of internationalization, export, import, foreign direct investment and cross-border cooperation is the one by MASUEL, VAN

HEMERT & DE GROOT [18]. Nevertheless they also leave open when entrepreneurs choose what mode of entry.

KEY FACTORS

Integrating the OLI paradigm by DUNNING [6][3] into the analysis, the reasons for SME's to internationalise in the *Grande Region* can be based to the following key aspects/ advantages. First, ownership advantages (the 'O' in DUNNING's concept) are found in the entrepreneur's knowledge about the *Grande Region* and her network of business partners when taking her business abroad. These two factors play a key role in all three steps of the internationalisation process outlined above. This is supported by research findings, such as AUTIO, SAPINZA & ALMEIDA [2], who already emphasize the impact of organisational knowledge and learning on international growth. Also FREEMAN, EDWARDS & SCHRODER [10], analyzing Australian firms, emphasize the importance of network and alliances to overcome constraints to rapid internationalisation. Also OVIATT & MCDOUGALL [21] highlight the importance of knowledge and network influences on the speed of internationalisation. For ISENBERG [14] 'articulating a global purpose' and 'alliance building', including 'supply chain creation' are competencies global entrepreneurs need. Knowledge about the foreign business environment and its actors and functioning, and the experience in how to behave in that international environment, will increase the probability that entrepreneurs take their business abroad. The network of international contacts, potential financiers as well as business partners, suppliers and customers, might then be the way to obtain the necessary resources to execute the internationalization decision..

Second, location advantages (the 'L' in DUNNING's concepts) should exist in the *Grande Region* per se, as the region is economically prosperous, politically stable and open, and culturally closely connected throughout history[4].

Third, internationalisation advantages (the 'I' in DUNNING's concept) should come up for SMEs in the *Grande Region* as transaction costs are low (no trade and investment barriers, geographical proximity), and gains are high, especially in terms of acquired knowledge about internationalisation and well established networks.

Assuming therefore the existence of location and internationalization advantages, an empirical analysis based on a four country sample in the *Grande Region* should reveal consistent results on the importance of ownership advantages when controlling for cultural differences in entrepreneurial activity. Ownership advantages are in the control of the entrepreneur by acquiring additional knowledge or establishing new business relations. Knowing the international business environment and its actors across the border will then have a positive impact on the motivation of the entrepreneur to take his business abroad. Hypothesis eight will therefore assume that internationalisation of a SME in the *Grande Region* is the start and not the end of its internationalisation process:

H8:   Internationalisation in the Grande Region encourages further internationalisation into other regions.

---

[3] See also DUNNING & LUNDAN [7] for a revisited version of the paradigm.
[4] Just to mention that within the 19th and 20th century Lorraine belonged for some time to Germany, the Saarland to France, Luxembourg either to France or Germany, and that Belgium is part of the BENELUX.

CONCLUSION: THE NEED FOR AN INTERNATIONAL SME OBSERVATORY

SME's and entrepreneurship in border regions has not drawn attention in theory building and research so far, even if the development of the European Union has enlarged the business links between companies in the member countries. Today, more than 70 cross-border regions exist in Europe [23, p. 153], one of the largest is the *Grande Region*. This region is unique, as it is the only one connecting four countries within a circle of 100 miles. This uniqueness is the starting point to create a cross-country cross-industrial database on SME's. As an observatory in the *Grande Region* it will collect regularly data from SMEs in Belgium, France, Luxembourg and Germany on three levels. First a qualitative analysis with experts in the field, including entrepreneurs, from all four countries, second a quantitative analysis of secondary data provided by the respective Chambers of Commerce and Industry and the local public statistical offices, and third in-depth surveys with entrepreneurs (linked to the local incubators) and small and medium-sized company managers. The data should validate the presented hypotheses. With these findings the importance of internationalisation for SME growth and success will be further supported.

REFERENCES

[1] Anderson, V. & Skinner, D. *Organizational learning in practice: how do small businesses learn to operate internationally?* Human Resource Development International, 1999, 2(3), 235-258

[2] Autio, E., Sapienza, H. J. & Almeida, J. G. *Effects of Age at Entry, Knowledge Intensity, and Imitability on International Growth.* Academy of Management Journal, 2000, 43(5), 909-924

[3] Crick, D. *The internationalisation of born global and international new venture SMEs.* International Marketing Review, 2009, 26 (4/5), 453-476

[4] Dimitratos, P., Plakoyiannaki, E., Pitsoulaki, A. & Tüselmann, H. J. *The global smaller firm in international entrepreneurship.* International Business Review, 2010, 19, 589-606

[5] Drori, I., Honig, B. & Wright, M. *Transnational Entrepreneurship: An Emergent Field of Study.* Entrepreneurship: Theory & Practice, 2009, 1001-1022

[6] Dunning, J. H. *The Eclectic (OLI) Paradigm of International Production: Past, Present and Future.* International Journal of the Economics of Business, 2001, 8(2), 173-190

[7] Dunning, J. H. & Lundan, S. M. *Multinational enterprises and the global economy.* Cheltenham: Edward Elgar 2008

[8] Fan, T. & Phan, P. *International new ventures: revisiting the influences behind the 'born-global' firm.* Journal of International Business Studies, 2007, 38, 1113-1131

[9] Fletcher, D. *International entrepreneurship and the small business.* Entrepreneurship & Regional Development, 2004, 16, 289-305

[10] Freeman, S., Edwards, R. & Schroder, B. *How Smaller Born-Global Firms Use Networks and Alliances to Overcome Constraints to Rapid Internationalization.* Journal of International Marketing, 2006, 14(3), 33-63

[11] Freytag, A. & Thurik, R. *Entrepreneurship and its determinants in a cross-country setting.* Journal of Evolutionary Economics, 2007, 17, 117-131

[12] Hill, C. W.: *International Business: Competing in the Global Marketplace.* Boston et al.: McGraw-Hill, 2011, 8th Edition

[13] Holmes, P. *From Single European Act to Maastricht: the Creation of the European Union*, in Einhorn, B., Kaldor, M. & and Kavan, Z. (eds*). Citizenship and Democratic Control in Contemporary Europe*. Cheltenham: Edward Elgar. 1996, 54-68

[14] Isenberg, D. J. *The Global Entrepreneur.* Harvard Business Review, 2008, Dec. 107-111

[15] Istrate, E. C., Kallas, K. & Stough, R.R. *Entrepreneurship in an international context: expanding the concept.* International Journal of Management and Enterprise Development, 2009, 7(1), 11-27

[16] Keupp, M. M. & Gassmann, O. *The Past and the Future of International Entrepreneurship: A Review and Suggestions for Developing the Field.* Journal of Management, 2009, 35(3), 600-633

[17] Leonidou, L. C. *An Analysis of the Barriers Hindering Small Business Export Development.* Journal of Small Business Management, 2004, 42 (3), 279-302

[18] Masurel, E., van Hemert, P. & de Groot, P. *Entrepreneurship across the border: internationalisation of Dutch SMEs.* International Journal Of Entrepreneurship and Innovation Management, 2009, 10(1), 20-37

[19] Moensted, M. *Networking and Entrepreneurship in Small High-Tech European Firms: An Empirical Study.* International Journal of Management, 2010, 27(1), 16-30

[20] O'Cass, A. & Weerawardena, J. *Examining the role of international entrepreneurship, innovation and international market performance in SME internationalisation.* European Journal of Marketing, 2009, 43 (11/12), 1325-1348

[21] Oviatt, B. M. & McDougall, P. P. *Defining International Entrepreneurship and Modeling the Speed of Internationalization.* Entrepreneurship: Theory & Practice, 2005 537-553

[22] Paasi, A. *The Resurgence of the 'Region' and 'Regional Identity': Theoretical Perspectives and Empirical Observations on regional Dynamics in Europe*. Review of International Studies, Special Issue February 2009, 35, 121-146

[23] Perkmann, M. Cross-border Regions in Europe. European Urban & Regional Studies, 2003, 10(2), 153-171

[24] Shane, S. A General Theory of Entrepreneurship. Cheltenham: Edward Elgar 2003

[25] Tolstoi, D. *Network development and knowledge creation within the foreign market: A study of international entrepreneurial firms.* Entrepreneurship & Regional Development, 2010, 22 (5), 379-402

[26] Westhead, P., Wright, M. & Ucbasaran, D. *International market selection strategiues selected by 'mirco' and 'small' firms.* Omega, International Journal of Management Science, 30, 51-68

[27] Wright, M., Westhead, P & Ucbasaran, D. *Internationalisation of Small and Medium-sized Enterprises (SMEs) and International Entrepreneurship: A Critique and Policy Implications.* Regional Studies, 2007, 41(7), 1013-1029

[28] Zahra, S. A., Korri, J. S. & Yu, J. *Cognition and international entrepreneurship: implications for research on international opportunity recognition and exploitation.* International Business Review, 2005, 14, 129-146

GLOBAL CURRENCY MOVEMENT AND GENERALIZED RULE INDUCTION

A. G. Malliaris

Loyola University Chicago, 1 E Pearson, Chicago IL 60611, 312-915-6063, tmallia@luc.edu

M. E. Malliaris

Loyola University Chicago, 1 E Pearson, Chicago IL 60611, 312-915-7064, mmallia@luc.edu

ABSTRACT

This paper uses one form of association analysis, Generalized Rule Induction (GRI), to investigate the relationships among directional movement of eight major currencies priced relative to the dollar. Data over ten years was studied to see if there are movement rules among these currencies that might be stable over time. Eighteen rules were discovered and discussed that appear to have some robustness over both the training and validation sets. For example, when the Australian Dollar and Japanese Yen move in a similar direction, the Euro frequently does the same. These results suggest that co-movement among some specific markets exists over relatively long periods of time and does not exist among others.

Keywords: generalized rule induction, foreign exchange, currencies, data mining

1. INTRODUCTION

In the search for diversity of assets in which to place funds, an investor prefers to have some investments that do not move together. That is, they do not go up and down simultaneously. However, if it is possible to isolate some assets that do have similar patterns of movement, then this knowledge can be used in two ways: to make money if one has moved and the other has not moved yet, or, to build a more diverse portfolio by not including both in the basket. Similarly, if no pattern of co-movement can be found over time, then the investor is more assured that a pair of specific markets move in a less synchronized fashion.

This paper investigates ten years of currency price data movement, using a methodology called Generalized Rule Induction (GRI). GRI was created in 1992 by Smyth and Goodman [13] as an algorithm that could be used for the induction of rules from a set of examples. Rather than using an expert (common, but labor intensive, at that time) to obtain rules, they wanted to create an algorithm that could automatically acquire rules from data, where that data existed. They also wanted their technique to generate rules relating not only two columns of data, but possibly multiple columns. Their focus was on reducing a large dataset to a small set of rules found within that data.

In today's internet environment, the acquisition of data is much easier that when this technique was first envisioned. With such ease of data attainment, the number of rule induction methods and problems for their application has increased. These methods continue to be popular for approaching problems in finance, for example, Batyrshin et al [2] [3] use time series databases, Albanis and Batchelor [1] employ rule induction to look at stock selection, Wang et al [15] study

forecasting time series, and Tseng [14] discusses co-movement in international stocks and Bossomaier et al [5] detail a simulation of trust in wealth management.

This research uses rule induction to look at international currency market data over a ten-year period. In today's world, since it is as easy to trade globally as to trade locally, the growth in the size and complexity of international financial markets has been one of the most striking aspects of the world economy over the last decade. This process of financial globalization and its possible effects has been described by Lane and Milesi-Ferretti [11] [12] and Devereux and Sutherland [9]. Campbell et al.[8] finds that the Australian dollar and the Canadian dollar are positively correlated with local-currency returns on equity markets around the world, including their own domestic markets. At the other extreme, the euro and the Swiss franc are negatively correlated with world stock returns and their own domestic stock returns. The Japanese yen, the British pound, and the US dollar fall in the middle, with the yen and the pound more similar to the Australian and Canadian dollars, and the US dollar more similar to the euro and the Swiss franc.

With the increasing globalization, certain currencies have received great significance the past few decades. This global significance translates into a search for the pricing of these currencies. The challenge becomes even greater since currencies are priced one in terms of another. One may view the issue of pricing currencies as a comparison of all economic and financial fundamentals between two nations. Empirical evidence of chaotic dynamics in financial data such as stock market indexes, foreign currencies, macroeconomic time series and several others have been performed by various researchers such as Brock, Scheinkman, Dechert and LeBaron, [7] and Brock and Malliaris[6]. However, there is very little empirical work done to study nonlinear chaotic determinism in currency markets.

One way of looking at the interplay of currencies by themselves is to study the direction of movement in each market per day. That is, are there certain major markets that move up or down together sufficiently often for us to form a conclusion about their inter-relationships? This paper uses Generalized Rule Induction to investigate the relationships among directional movement of eight major currencies around the globe. Data over ten years was studied to see if there are movement rules among these currencies that might be stable over time.

2. DATA AND METHODOLOGY

We began with daily cash closing prices for the Australian Dollar, British Pound, Brazilian Real, Canadian Dollar, Euro, Japanese Yen, Mexican Peso, and Swiss Franc with respect to the US Dollar. That is, the data reflects the amount of each foreign currency that could be purchased with 1 US dollar that day. Though the foreign exchange market is considered to be a 24-hour market, closing prices can be quoted for individual markets in pairs. Thus, when the market in Tokyo closes, the value of the Yen to the Dollar can be established for that day. All values in this data set are in these units of the foreign currency per US dollar. The data sample covers the time period from January 2000 through July 2009 and was downloaded from Bloomberg.

The relative movement in these currencies can be seen in Figure 1. In order to view them all in a similar scale, the Mexican Peso has been multiplied by 10 and the Japanese Yen by 100 for the

graph. There are a total of 2,491 observations for prices for each of the eight daily closing prices. These prices were split into two disjoint sets for training and validation. Data from January 1 2000 to June 30 2008 was used as the training set (2215 rows), with the remainder, from July 1 2008 to July 21 2009, used as the validation set (276 rows). To study the simultaneous market movements, all data was transformed into "Up" or "Down" by comparing the value of the currency at time t with its value on the previous day.

FIGURE 1. Currency Prices in units of 1 US Dollar, Peso & Yen scaled
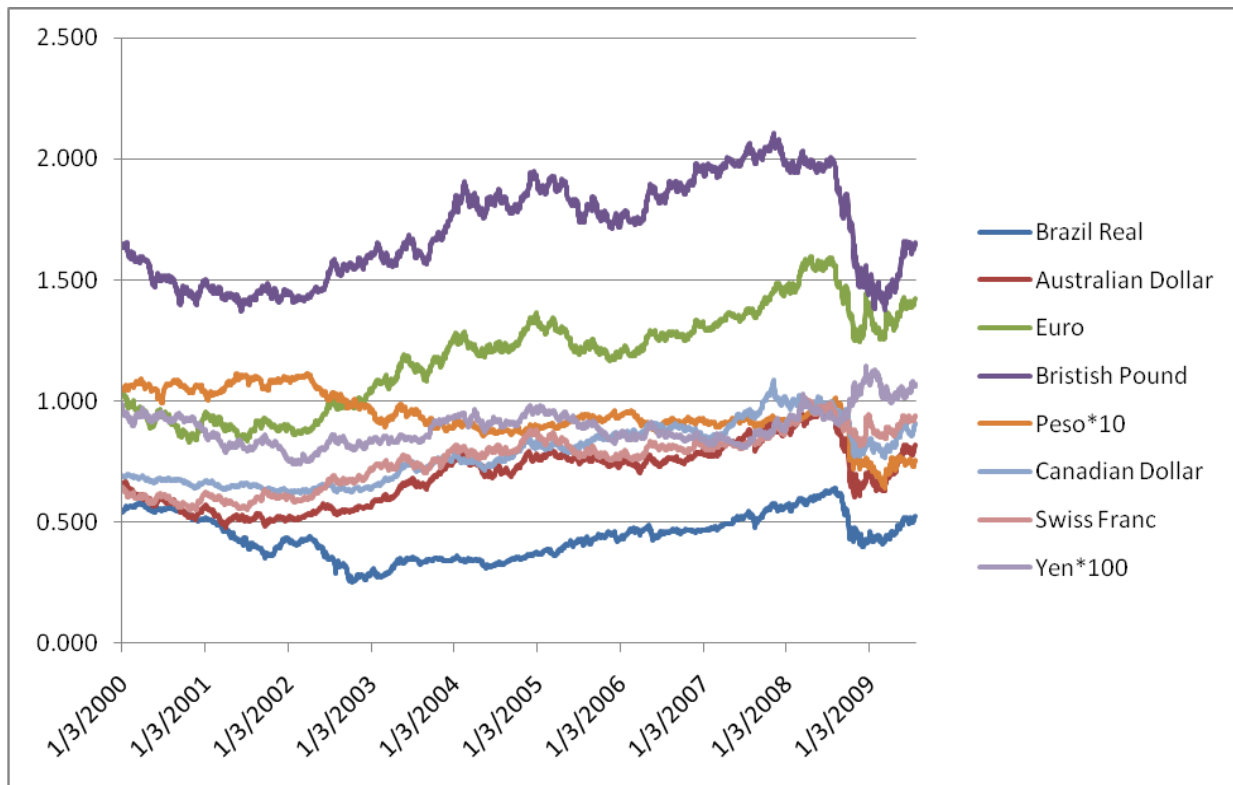
Association analysis is a popular data mining method that originated with the study of market baskets to see which items people purchased at the same time. It is often used as an exploratory method to help discover interesting relationships in the data that you may wish to analyze further. For an in-depth discussion of association analysis techniques, see Hand et al [10], or Berry and Linoff [4]. The variables in an association analysis model can be specified for use as inputs, outputs, or both. Association analysis then generates a set of rules of the form IF A THEN B where variables specified as inputs may occur after IF, variables specified as outputs may occur after THEN, and variables specified as both may occur in either position.

The set of rules that is generated also depend on the user-supplied minimum values of support and confidence. Support refers to the percent of times that some combination of inputs (also called antecedents) occurs in the data set. When the antecedent combination does occur, confidence reflects the percent of time that the output, or consequent, is also true. There are several major association analysis techniques, for example, Apriori, Generalized Rule Induction

(GRI), and Carma. These vary in the way they search for interesting rules within a large, and generally sparse, data set.

Generalized Rule Induction is a methodology that was introduced in 1992 by Smyth and Goodman [13] when they proposed a new method of automated rule acquisition based on large amounts of data. This methodology introduced a measure of information content called the J-measure. GRI uses the J-measure to quantify the amount of information in a rule, and then generates a set of optimal rules on the training data. The J-measure is defined as

$$J = p(x) [p(y|x) \ln (p(y|x)/p(y)) + (1 - p(y|x)) \ln (1 - p(y|x))/(1-p(y))] \tag{1}$$

where p(x) is the probability of the observed value of x and p(y) is the probability of y.

After running the GRI methodology, a model is generated that consists of a set of rules containing the most information. Along with the statement of each rule, GRI also specifies the specific support and confidence that occurred for each rule. Only rules that meet the user-specified minimum levels for support and confidence are listed. With the easy use of computers today to isolate conditions meeting the antecedents, the level of support (that is, the percent of occurrence within the database) is of less concern that the level of confidence. Thus, if a set of antecedent conditions occurs a very small number of times, but the consequent confidence is high, then this would be a rule worth exploring.

3. PROBLEM SETUP AND RESULTS

Association analysis problems most often use data in symbolic format. Thus, all data was converted from numeric values, reflecting their amount relative to 1 US dollar, into non-numeric data indicating only the direction of movement relative to the US dollar on any given day. The Generalized Rule Induction methodology was run using the SPSS product Clementine (now renamed IBM's SPSS Modeler since the purchase of SPSS by IBM).

There were two runs of the model for this study with the inputs and outputs divided into groups determined by time zone locations. In the first run of GRI, because of the time difference in markets across the globe, the Australian dollar and Japanese Yen were used as inputs with the Euro, the Swiss Franc and the British Pound as possible outputs. Following this run, the Australian dollar, Japanese Yen, the Euro, the Swiss Franc and the British Pound were used as inputs, with the Mexican Peso, the Brazilian Real and the Canadian dollar as outputs. Only the training set data was used in the generation of the rule sets. Then the validation data was run through the trained rule set to see whether the rules were stable over time.

The GRI models generated a total of 151 rules. Here we show a selection of 18 of those rules where the validation set results supported the training set results, sorted by the market of the consequent. Each rule displays the antecedent and consequent followed by the support and confidence in each of the data sets.

Rules 1 2, and 3 have, as a consequent, the Brazilian Real. Looking at the first rule, we see that, if both the Australian Dollar and the Euro were down at closing, then, on the same day, the

Brazilian Real was down at closing. This simultaneous downward movement in Australia and the Euro occurred in 38.06% of the training set cases. When that simultaneous movement occurred, then 61.09% of the time, the Real was also down in the training set. The support and confidence values are also given for the validation set. In the validation set, we see that the simultaneous down movement for Australia and the Euro occurred in about the same percentage of rows, but the Real's co-movement in the down direction occurred even more often, 73.08% of the time. Rule 2 has the same consequent but an antecedent with movement only from Australia specified. In this rule, we notice that the support is greater than in rule 1, but the confidence is lower. That is, fewer restrictions on the antecedent mean that more rows will display this antecedent, but the percentage of those rows displaying the consequent will be smaller. However, both rules have a high enough confidence to expect that the down movement in the Real is closely tied to that of the Australian Dollar.

Rule 3 focuses on conditions necessary for the Real to move up relative to the US dollar. In this case, movements in three markets must happen before a strong consequent appears. Thus, Australia must be up relative to the dollar, the Swiss France must move down, and the Yen must also move down.

Rules 4 and 5 have the British Pound as a consequent. These rules indicate that movement in the British Pound, either up or down, is tied to simultaneous co-movements in both Australia and Japan. The confidence levels in both the training and validation sets are in the mid-70s, though the occurrence of the antecedents drops from training to validation.

The Canadian dollar is the consequent for rules 6 and 7. It moves down when both Australia and the Euro are down, but up movement depends only on Australia. The percent of time that the consequent was true increased in the validation set for both of these rules. The consequent of rule 6 is true for over 82% of the cases in the validation set.

The direction of the Euro is the focus of rules 8 and 9. These show that the movement of the Euro is tied to having movement in both Australia and Japan go in the same direction. That is, all three go up and all three go down together about 80% of the time. These rules did not diminish in confidence from the training to the validation set.

Rules 10 through 16 all have the Mexican Peso as a consequent. For the Mexican Peso to move down, either Australia or the Euro must also be down, and some other market must move up, as we see in rules 10 through 13. Up movement in the Peso happens often when either the Euro or Australia are up and the Swiss Franc is down. These rules have much less confidence in the training set than in the validation set, an indication that there may be less robustness in these rules over the time period studied.

Finally, rules 17 and 18 indicate that simultaneous movement in Australia and Japan are an indication that the Swiss Franc will go the same direction that day. This relationship holds for both the up and down directions with high confidence. As we see in the table, over 80% of the time, this simultaneous movement occurs.

In addition to rule generation using GRI, Clementine allows us to generate a picture of pair-wise relationships among non-numeric data variables using a web graph. This graph, shown in Figure 2, illustrates pair-wise down movements on the left and pair-wise up movements on the right in the training data. Stronger relationships, that is, ones that occur more often, have darker lines connecting them. This graph does not combine up and down movement in the various markets as we can do with GRI, but it does give us some indications of places we might expect rules to be generated.

By looking at these graphs, we see that there are more relationships shown for up than for down. Also note that the relationships between the Euro and the Swiss Franc are the strongest in both directions of movement. This figure also enables us to see the power of the GRI methodology over a simple graphic view of a data set. For example, while the links of the Australian dollar and the Japanese Yen to the Swiss Franc are thin on the Up chart here, GRI has shown us that the combination of up values on these two are a powerful indicator of movement in the direction of the Franc as priced in dollars.

FIGURE 2.   Clementine Web-graph of pair-wise down and up movements
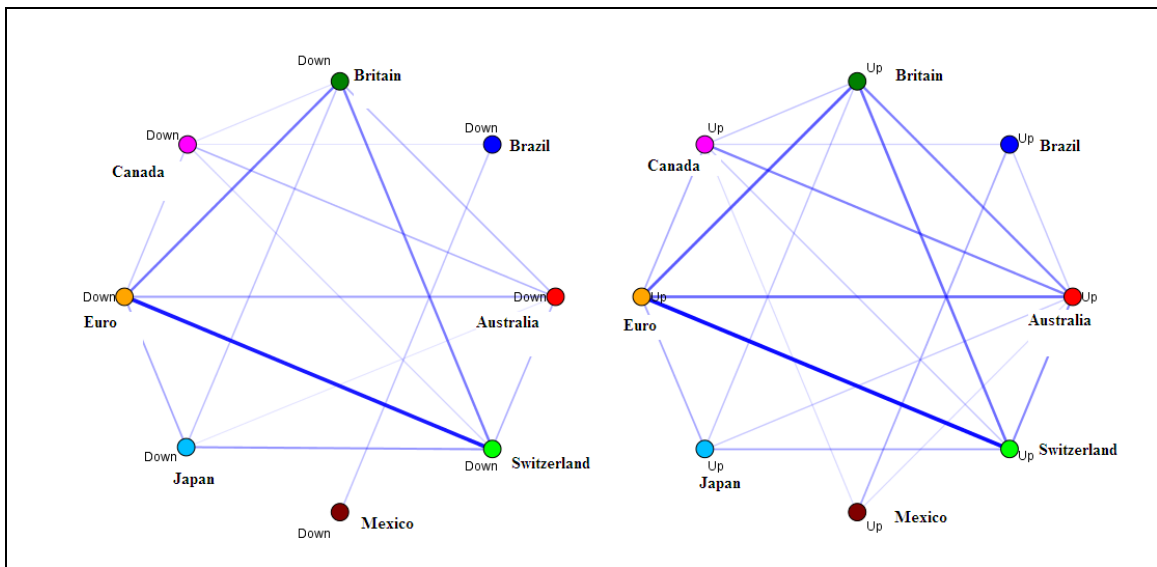
TABLE 1. Rules generated by the GRI algorithm

	RULE		TRAINING		VALIDATION	
	Antecedent	Consequent	Sup	Conf	Sup	Conf
1	Australia= Down and Euro= Down	Brazil = Down	38.06	61.09	37.68	73.08
2	Australia = Down	Brazil= Down	54.09	59.35	52.90	69.18
3	Australia = Up and Swiss = Down and Japan = Down	Brazil = Up	8.08	64.80	11.23	70.97
4	Australia = Down and Japan = Down	Britian = Down	32.37	76.71	22.83	76.19
5	Australia = Up and Japan = Up	Britian = Up	28.94	72.07	17.39	77.08
6	Australia = Down and Euro = Down	Canada = Down	38.06	72.24	37.68	82.69
7	Australia = Up	Canada = Up	45.91	65.68	47.10	78.46
8	Australia = Down and Japan = Down	Euro = Down	32.37	81.45	22.83	85.71
9	Australia = Up and Japan = Up	Euro = Up	28.94	78.00	17.39	85.42
10	Australia = Down and Euro = Down and Swiss = Up	Mexico = Down	4.24	64.89	6.16	76.47
11	Australia = Down and Britian = Down and Japan= Up	Mexico = Down	11.60	63.04	17.75	71.43
12	Euro = Down and Swiss = Up	Mexico = Down	7.00	63.23	8.33	69.57
13	Euro = Down and Britian = Down and Swiss= Up	Mexico = Down	3.97	63.64	6.16	64.71
14	Euro = Up and Britian = Up and Swiss = Down	Mexico = Up	3.02	64.18	5.80	87.50
15	Euro = Up and Swiss = Down	Mexico = Up	6.05	58.96	6.88	84.21
16	Australia = Up and Swiss= Down	Mexico = Up	14.90	57.88	13.41	78.38
17	Australia= Down and Japan = Down	Swiss= Down	32.37	80.06	22.83	82.54
18	Australia = Up and Japan = Up	Swiss = Up	28.94	76.44	17.39	87.50

4. CONCLUSIONS

For many centuries, national economies have been linked to one another financially, primarily because of trade. The importing nation received goods and paid in some pre-agreed currency. Currency trading predates both bond and stock trading as a financial innovation. However, there is very little doubt that during the past 50 years globalization grew at a remarkable pace and with it, currency trading.

Today, the daily volume of currency transactions in currency futures, forwards, swaps and options dominates all other types of trading volumes. This volume is driven by globalization that includes both trade and foreign direct investments, by portfolio diversification and by hedging and speculation, among other factors. Whether currencies move together or independently is a matter of importance for investors wishing to spread the impact of their portfolio decisions.

This study used a set of data spanning almost ten years with data reflecting the values of eight major currencies relative to the US dollar. The data was transformed into directional movement and then analyzed to see whether association analysis could uncover rules relating to simultaneous currency movement that remained stable over time.

Eighteen rules were discovered and discussed that appear to have some robustness over both the training and validation sets. These results suggest that there is reason to believe that co-movement among some specific markets exists over relatively long periods of time and does not exist among others. This information can be used to help diversify a portfolio by holding currencies for which related movements do not occur.

5. REFERENCES

[1] Albanis, G. and Batchelor, R. *Combining heterogeneous classifiers for stock selection*. Intelligent Systems in Accounting, Finance & Management, 2007, 15: 1–21. doi: 10.1002/isaf.282

[2] Batyrshin, I., Herrera-Avelar, R., Sheremetov, L, and Panova, A. *Association networks in time series data mining*, Proceedings of the Fuzzy Information Processing Society, 2005, pgs 754 – 759.

[3] Batyrshin, I., Herrera-Avelar, R., Sheremetov, L, and Panova, A. *Moving Approximation Transform and Local Trend Associations in Time Series Data Bases*. Perception-based Data Mining and Decision Making in Economics and Finance, Studies in Computational Intelligence, 2007, Springer, Berlin / Heidelberg.

[4] Berry, M. and Linoff, G. Data Mining Techniques, Second Edition, Wiley Publishing Inc., 2004.

[5] Bossomaier, T., Standish, R., Harre, M. *Simulation of trust in client-wealth management adviser relationships*, International Journal of Simulation and Process Modelling, 2010, 6(1), 40 – 49.

[6] Brock, W., and Malliaris, A.. Differential equations, stability and chaos in dynamic economics. North Holland, Amsterdam, The Netherlands, 1989.

[7] Brock, W., Scheinkman, J., Dechert, W. and LeBaron, B., *A test for independence based on the correlation dimension*, Econometric Reviews, 1996, Taylor and Francis Journals, 15(3), 197-235.

[8] Campbell, J.Y., Medeiros, K.S., and Viceira, L.M., *Global Currency Hedging*, NBER Working Paper, No. W13088, 2007.

[9] Devereux, M.B., and Sutherland, A., *Financial Globalization and Monetary Policy*, IMF Working Paper, 2007.

[10] Hand, D., Mannila, H., and Smyth, P. Principles of Data Mining, The MIT Press, 2001.

[11] Lane, P., and Milesi-Ferretti, G.M., *The External Wealth of Nations: Measures of Foreign Assets and Liabilities for Industrial and Developing Countries*, Journal of International Economics, 2001, 55, 263-94.

[12] Lane, P., and Milesi-Ferretti, G.M., *The External Wealth of Nations Mark II*, IMF Working Paper, No 06-69, 2006.

[13] Smyth P, Goodman RM. *An information theoretic approach to rule induction from databases*. IEEE Transactions on Knowledge and Data Engineering. 1992;4(4):301–316. doi: 10.1109/69.149926

[14] Tseng, C. *Data driven modeling of co-movement among international stock market*, Journal of Modelling in Management, 2007, 2(3), 195 – 207.

[15] Wang, X., Smith-Miles, K., Hyndman, R., *Rule induction for forecasting method selection: Meta-learning the characteristics of univariate time series*, Neurocomputing, June 2009, 72(10-12), 2581-2594.

Media Disclosure, Board Structure and CEO Compensation: Evidence from Taiwan

Bingsheng Yi, Chia-Wei Chen and Barry Lin*

Abstract

This paper reports empirical evidence in the relationship between media coverage, board structure and CEO compensation using data from Taiwan Stock Exchange. We find that CEO compensation is much higher for firms with more press coverage, firms with more analyst coverage, firms with larger board, and larger institutional holdings. CEO compensation is lower for firms with higher percentage of independent board members and for firms with CEO who also serves as the Chairman of the board. We controlled for potential endogeneity problem, and the results remain the same. Our findings confirm and extend results found by other papers in the complex relationship among board structure, CEO duality, CEO compensation, and press coverage.

*Bingsheng Yi is at Californian state University—Dominguez Hills, Chia-Wei Chen is at Tunghai University, Taiwan, and Barry Lin is at Simmons College.

1. Introduction

Mass media coverage is a major channel of financial information and often asserts an immense influence to public opinion ad emotion.   In recent time, press reporting often is associated with public anger and outcry in corporate fraudulent practices and untruthful financial disclosure.   Lily and Peress (2009) reports evidence of a relationship between media coverage and cross-sectional expected stock returns. In particular, they found stocks with no media coverage earn higher returns than stocks with high press coverage.

Dick, Volchkova and Zingales (2008) investigates media coverage and its effect on corporate governance on firms in Russia.   They find that Anglo-American press coverage increases the probability that a corporate governance violation is reversed.

In a separate strand of research, Chhaochharia and Grinstein (2009) find a significant decrease in CEO compensation for firms which were required by US stock exchanges to strengthen their board oversight.   Other papers documented evidence that board decisions often affect CEO compensation strongly (see, among others, Fama (1980), Fama and Jensen (1983), Jensen (1993), Hall and Murphy (2003), and Bebchuk and Fried (2003, 2004)).)

This paper uses a sample of firms listed in Taiwan Stock Exchange and investigates the effect of the degree of media coverage on CEO compensation, subject to other important variables such as board structure, analyst coverage, CEO duality, and institutional ownership.   We find strong evidence that media coverage is strongly associated with CEO compensation, with CEO of firms with high coverage receiving substantially higher compensation.   Analyst coverage and institutional ownership both have a similar positive effect on CEO compensation, so is board size, while percentage independent board member has an opposite effect.   CEO duality (CEO who also serves as the Chairman of the board) has a negative effect on CEO compensation.   Our results broadly mesh with other studies on CEO compensation, board structure, and other studies on corporate governance.

2. Data and empirical results

Our press coverage data is collected from TEJ (Taiwan Economic Journal Co., Ltd), one of the largest business publications.   Sample firms include all firms listed in TSE and OTC in Taiwan from 2005 to 2008.

Table I reports summary statistics for our sample firms.   CEOs in Taiwan receive, on average, NT\$12,062,000 of total compensation, with bonus accounting for about 75% of the package.   An average firm receives 206 times of press coverage, with 568 times for the 95 percentile, and 54 times for the 5% percentile, thus exhibiting a large range of press coverage among the firms.

Table II reports test results for our univariate comparisons.   Average CEO compensation of firms with higher than average press coverage is almost four times higher than that of firms with below average press coverage, which is a very substantial difference.   Almost the same degree of difference can be seen on the effect of analyst coverage.   Firms with above average board size (found to be less efficient in monitoring) pay their CEO almost twice as much as firms with below average board size.   Firms with above average % of independent board members (associated with better monitoring) pay their CEO less than firms with lower % of independent board members.   Finally, CEOs with dual titles (found to be associated with worse firm

performance) receive lower compensation.

Table III and IV report results of multivariate tests with OLS and fixed-effect regressions.   These results confirm our earlier findings in the univariate tests.   In particular, press coverage is found to have a highly significant and positive effect on CEO compensation.

Table I
Summary Statistics

	Mean	5% Percentile	Median	95% Percentile	Standard deviation
Panel A: Information Disclosure					
Press coverage	206	54	141	568	225
Number of analyst forecasts	3.43	0	3	7	2.28
Bonus (NT$000)	3,096	0	579	13,285	8,533
Salary (NT$000)	8,966	0	6,225	27,730	8,966
Total compensation (Bonus + Salary, NT$000)	12,062	0	7,579	38,298	16,982
CEO duality dummy	0.30	0	0	1	0.46
Board size	6.80	5	7	11	2.21
Independent director(s)	0.99	0	0	3	1.07
Proportion of independent directors in the board (%)	15.09	0	0	40	16.58
Institutional ownership (%)	35.69	5.28	31.74	77.91	22.52
Market to book ratio	1.34	0.62	1.05	2.85	1.33
Return on equity (%)	16.30	-25.09	11.85	68.90	33.74
Return on assets (%)	3.55	-15.58	4.73	19.10	13.26
TSE dummy	0.56	0	1	1	0.50
Firm age since listed in TSE or OTC	8.5	1	6	26	8.00
Sales (NT$000)	11,600	278	2,303	37,800	53,000
Total assets (NT$000)	13,000	577	3,111	45,500	44,700

Table II Univariate Tests

Comparison of average CEO compensation between different groups: The whole sample firms are classified into two groups according to whether a sample firm's value in press coverage (number of analyst forecasts, board size, the proportion of independent board members, and institutional ownership) exceeds its median value or not respectively.

Panel A: CEO compensation and press Coverage

	> median	<= median	
	(1)	(2)	(1) – (2)
Bonus (NT$000)	4,753	1,458	3,295^a
Salary (NT$000)	12,066	5,901	6,165^a
Total compensation	16,819	7,359	9,460^a
N	2,203	2,228	

Panel B: CEO compensation and number of analyst coverage

	> median	<= median	
	(1)	(2)	(1) – (2)
Bonus (NT$000)	4,554	1,728	2,826^a
Salary (NT$000)	11,479	6,608	4,871^a
Total compensation	16,033	8,337	7,696^a
N	2,145	2,286	

Panel C: CEO compensation and board size

	> median	<= median	
	(1)	(2)	(1) – (2)
Bonus (NT$000)	5,179	2,539	2,640^a
Salary (NT$000)	12,243	8,089	4,154^a
Total compensation	17,422	10,628	6,794^a
N	935	3,396	

Panel D: CEO compensation and proportion of independent board members in the board

	> median	<= median	
	(1)	(2)	(1) – (2)
Bonus (NT$000)	2,850	3,334	-484^c
Salary (NT$000)	8,927	9,003	-76
Total compensation	11,777	12,337	-560
N	2,176	2,255	

Panel E: CEO compensation and CEO duality

	CEO dummy =1	CEO dummy =0	
	(1)	(2)	(1) – (2)
Bonus (NT$000)	2,421	3,379	-958^a
Salary (NT$000)	8,320	9,237	-917^a
Total compensation	10,741	12,616	-1,875^a
N	1,310	3,121	

^{a} indicates significance at 1%, ^{b} indicates significance at 5%, ^{c} indicates significance at 10%.

Table III OLS Regression Results

Independent variables	Dependent variable		
	Bonus	Salary	Total compensation
Press coverage	11.08[a]	18.89[a]	29.97[a]
	(5.42)	(10.24)	(9.51)
Number of analyst coverage forecasts	311.63[a]	590.41[a]	902.03[a]
	(4.62)	(6.68)	(7.10)
CEO duality	-609.63[a]	-842.29[a]	-1451.92[a]
	(-2.75)	(-3.13)	(-3.63)
Board size	109.41	172.21[b]	281.63[b]
	(1.40)	(2.34)	(-2.19)
% of Independent directors in the board	2.24	-15.07	-12.83
	(0.24)	(-1.40)	(0.44)
Institutional ownership	20.26[a]	23.40[a]	43.67[a]
	(4.71)	(2.86)	(3.77)
Return on equity	-5.11	-41.71[a]	-46.82[a]
	(-0.96)	(-5.37)	(-4.42)
Market to book value	-213.43[a]	-450.47[a]	-663.89[a]
	(-2.83)	(-3.89)	(-3.67)
Sales	0.00001	8.70e-06	0.00002
	(1.33)	(0.99)	(1.23)
Firm age	30.53	73.64[a]	104.18[a]
	(1.47)	(2.77)	(2.70)
TSE dummy	1036.57[a]	2856.47[a]	3893.04[a]
	(5.30)	(10.23)	(9.74)
Year dummies	Yes	Yes	Yes
Industry dummies	Yes	Yes	Yes
R-square	0.207	0.325	0.343
Number of observations	4431	4431	4431

[a] indicates significance at 1%, [b] indicates significance at 5%, [c] indicates significance at 10%.

Table IV Fixed Effect Model Regression Results

Independent variables	Dependent variable		
	Bonus	Salary	Total compensation
Press coverage	1.12	5.26[a]	6.38[a]
	(1.06)	(5.36)	(3.93)
Number of analyst coverage forecasts	149.36[b]	97.14	246.50[b]
	(2.23)	(1.57)	(2.41)
CEO duality	-417.01	-206.66	-623.67
	(-1.17)	(-0.63)	(-1.14)
Board size	251.07[c]	26.81	277.89
	(1.88)	(0.22)	(1.36)
% of Independent directors in the board	-1.13	-4.86	-5.99
	(-0.08)	(-0.36)	(-0.27)
Institutional ownership	18.65	-5.86	12.79
	(1.55)	(-0.53)	(0.69)
Return on equity	12.60[b]	-6.69	5.91
	(2.38)	(-1.37)	(0.73)
Market to book value	-29.10	0.61	-28.50
	(-0.30)	(0.01)	(-0.19)
Sales	0.0003[a]	0.00003[a]	0.00005[a]
	(5.06)	(4.97)	(6.31)
Firm age	838.03[a]	707.39[a]	1545.42[a]
	(9.51)	(4.97)	(11.45)
TSE dummy	-74.13	-664.90	-739.04
	(-0.08)	(-0.77)	(-0.52)
Year dummies	Yes	Yes	Yes
Industry dummies	No	No	No
Model F-stat	13.82	16.39	22.37
R-square overall	0.053	0.068	0.079
Number of firms	1189	1189	1189
Number of observations	4431	4431	4431

[a] indicates significance at 1%, [b] indicates significance at 5%, [c] indicates significance at 10%.

Private Benefits in a contingent claim framework: Valuation effects and other implications

Franck MORAUX[1]
franck.moraux@univ-rennes1.fr
Université de Rennes 1
IAE de Rennes and CREM

Patrick NAVATTE
patrick.navatte@univ-rennes1.fr
Université de Rennes 1
IAE de Rennes and CREM

Abstract

This article extends the continuous time framework of the firm developed by Black, Scholes and Merton to analyze effects of private benefits. We highlight first how *straight* private benefits can lower stakeholders' wealth. We nevertheless point out that private benefits do not necessary mean expropriation of minority equity holders. Managers can indeed adjust the business risk of the firm so as to make private benefits innocuous for equity. This in turn has a couple of beneficial consequences for blockholders. First, this prevents minority equity holders to complain for expropriation to authorities. Second, the portion of equity they have does not suffer from any value decrease. Innocuous private benefits cannot have however unlimited value, because the associated volatility would be unreasonable and too detrimental for bondholders. We also consider cases where managers enjoy a retirement plan and account for this in their strategic adjustment. Our simulations suggest that minority shareholders may take advantage of it. Finally, we advocate that covenants in loan contracts may prevent or limit such strategic behaviour.

Résumé

Cet article analyse les implications des bénéfices privés à la lumière de la théorie des actifs contingents développée par Black, Scholes, et Merton. On montre que dans un tel cadre les bénéfices privés ne signifient pas forcément l'expropriation des actionnaires minoritaires. Les dirigeants peuvent adopter un comportement stratégique au profit des actionnaires majoritaires, mais adapter le niveau de risque des affaires de l'entreprise de telle sorte que les prélèvements opérés deviennent indolores pour les actionnaires minoritaires. De ce fait deux avantages en découlent. D'abord, les minoritaires ne peuvent se plaindre auprès des autorités compétentes, ensuite aucune perte de valeur de la participation des majoritaires en actions n'intervient. Bien sûr, les bénéfices privés ne peuvent devenir illimités, car la variation de volatilité impliquée serait trop importante, et de toute manière se réaliserait au détriment des prêteurs. Ces derniers peuvent dès l'origine prévoir dans les contrats de prêt des clauses empêchant ou limitant de telles pratiques. Si le dirigeant bénéficie en plus d'un plan de retraite sous la forme d'une part de dette subordonnée, les actionnaires minoritaires en bénéficieront.

Keywords: Agency theory, private benefits, tunneling, contingent claim analysis.

JEL classification: G12, G32, G33.

1

Private Benefits in a contingent claim framework: Valuation effects and other implications

Abstract

This article extends the continuous time framework of the firm developed by Black, Scholes and Merton to analyze effects of private benefits. We highlight first how *straight* private benefits can lower stakeholders' wealth. We nevertheless point out that private benefits do not necessary mean expropriation of minority equity holders. Managers can indeed adjust the business risk of the firm so as to make private benefits innocuous for equity. This in turn has a couple of beneficial consequences for blockholders. First, this prevents minority equity holders to complain for expropriation to authorities. Second, the portion of equity they have does not suffer from any value decrease. Innocuous private benefits cannot have however unlimited value, because the associated volatility would be unreasonable and too detrimental for bondholders. We also consider cases where managers enjoy a retirement plan and account for this in their strategic adjustment. Our simulations suggest that minority shareholders may take advantage of it. Finally, we advocate that covenants in loan contracts may prevent or limit such strategic behaviour.

Résumé

Cet article analyse les implications des bénéfices privés à la lumière de la théorie des actifs contingents développée par Black, Scholes, et Merton. On montre que dans un tel cadre les bénéfices privés ne signifient pas forcément l'expropriation des actionnaires minoritaires. Les dirigeants peuvent adopter un comportement stratégique au profit des actionnaires majoritaires, mais adapter le niveau de risque des affaires de l'entreprise de telle sorte que les prélèvements opérés deviennent indolores pour les actionnaires minoritaires. De ce fait deux avantages en découlent. D'abord, les minoritaires ne peuvent se plaindre auprès des autorités compétentes, ensuite aucune perte de valeur de la participation des majoritaires en actions n'intervient. Bien sûr, les bénéfices privés ne peuvent devenir illimités, car la variation de volatilité impliquée serait trop importante, et de toute manière se réaliserait au détriment des prêteurs. Ces derniers peuvent dès l'origine prévoir dans les contrats de prêt des clauses empêchant ou limitant de telles pratiques. Si le dirigeant bénéficie en plus d'un plan de retraite sous la forme d'une part de dette subordonnée, les actionnaires minoritaires en bénéficieront.

2

Introduction

Agency problems are important feature of the contractual view of the firm. Some of them derive directly from the separation of ownership and control (Jensen and Meckling, 1976). Managers, e.g., have discretion on how to allocate investors funds, so it is very feasible for them to expropriate all or some of shareholders (Shleifer and Vishny, 1997). Large investors, for their part, develop an active monitoring activity to represent their own interests (Cronquist and Fahlenbrach, 2009). They are sometimes suspected to use their control rights to maximise their own welfare and redistribute wealth. These interests are not, of course, necessary in line with those of other stakeholders of the firm. In most concrete cases, such profits may become private.

Private benefits occur each time controlling shareholders extract values from the firm without sharing with other equityholders. In family firms, they can take the form of insider trading (Villalonga and Amit, 2009). Private benefits are especially substantial when private information is available and good corporate governance as well as transparency not present. It is also well known that both the manager's compensation and the quality of the board of directors do matter.

Of course, the existence of a significant block holder is important to monitor managers and prevent associated private benefits. But, as the block holder gets more power and begins to control the firm, the effect may reverse. The manager and monitor may indeed both agree to expropriate minority share

3

holders and gain from the firm[2]. With the help of managers, controllers can extract values from the firm to increase their wealth[3].

This paper aims at exploring this last issue. Searching for quantitative insights and managerial implications, we take into consideration the most important satkeholders, i.e., the majority share holder, the minority equity holder, the CEO and the senior and junior debt holders. It is worth summing up most insightful results of the paper. We highlight first how *straight* private benefits can lower stakeholders' wealth. We then illustrate that private benefits do not necessary mean expropriation of minority equity holders. Managers can indeed adjust the business risk of the firm so as to make private benefits innocuous for equity. This in turn has a couple of beneficial consequences for block holders. First, this prevents minority equity holders to complain for expropriation to authorities. Second, the portion of equity they have does not suffer from any value decrease. Innocuous private benefits cannot have however unlimited value, because the associated volatility would be unreasonable and too detrimental for bondholders. We also find that managers' retirement plans do matter and that minority shareholders may take advantage of it.

The rest of this paper is organized as follows. Section 1 undertakes a short literature review. Section 2 presents our contingent claim framework. Section 3 then analyzes private benefits and studies their impact on the present value of shares, debts when they are *straight*, i.e. when everything else remains the same. Section 4 points out that a strategic behaviour for block holders is

[2] Even benefits of concentrated ownership and monitoring appear contrasted. Double voting rights, e.g., represent a simple way to expropriate shareholders (Masulis and al., (2009), Villalonga and Amit (2006). Active monitoring is costly and time-consuming, while long term monitoring requires an ownership stability that is difficult to reconcile with investors appetite for liquidity. Investors may actually fear not to be able to seize opportunities on the market. Quantifying all these costs is out of the scope of this paper.
[3] The current theoretical view of control transfers testifies of these two conflicting effects of block control. On the one hand, equity transactions are supposed to help control transfers to more efficient management (Burkart and al, 2000). On the other hand, one suspects that

4

feasible, which making private benefits innocuous for all share holders. The idea here is to increase the volatility of the firm's assets[4] so that the minority equity holder is compensated for the loss caused by private benefits. At the end of this section, we also discuss the most extreme case of tunneling. Section 5 explores situations where CEO owns, besides his/her share holding, a retirement plan. Because such a plan constitutes a fraction of junior debt, it may have a strong impact on the strategic volatility adjustment the CEO proceeds. After simulations, we compare the various scenarios and derive more qualitative managerial implications.

1. A short literature review

Private benefits have been well defined by Dyck and Zingales (2004) as more than the "psychic" value some shareholders attribute to controlling the decisions of the firm. This additional value has been empirically documented. Barclay and Holderness (1989) find, among others, that prices requested for buying a block of stocks are often higher than forecasted. This result suggests that associated benefits are not shared with others.

Private benefits may take various forms. Top executives may enjoy very high compensation schemes and perquisites in excess to what should be. Some significant corporate resources may be diverted by large shareholders, especially if the law does not prevent it (La Porta and alii, 2000). Tunneling may be viewed as the most extreme form of such private benefits. It is a significant transfer of resources out of a firm to its controlling shareholders[5]. More subtly, certain information on business may have some value outside the

---

acquirer's primary motive is to loot the firm at the expense of small shareholders (Baek and al., 2006).

[4] Modifying the business risk of the firm does not mean necessary changing its productive assets. The firm may, for instance, export a larger fraction of its output to emerging countries.

[5] Johnson and alii (2000) identify three main procedures: "cash-flow tunneling", "asset tunneling" (also discussed by Djankov and alii, 2008), and "equity or guarantee tunneling" (Atanasov and alii, 2010). Some also consider that excessive equity-based compensations (in the sense that they exceed by far the market price) and abnormal profits from "insider trading" are other kinds of tunneling.

5

firm within companies owned by dominant equity holders. In real transactions with such companies, there will be a never ending debate concerning what should be called the "fair" transfer price. Small deviations from the proper price are difficult to detect, but if applied to a large volume of transactions, it can represent some sizeable private benefits. In each case just considered, the value extracted by the controlling group is not shared among all shareholders in proportion of the stocks owned.

Practical approaches to estimate private benefits of control take into consideration proxies, like the block premium, or the voting right premium if different classes of shares are prevailing. Consequences are more or less distributional ones and not connected with efficiency. Dyck and Zingales (2004) use the block premium methodology trying to disentangle the effect of private benefit from the change in share value associated with the new block holder. Based on 393 transactions of control among 39 countries between 1990 and 2000, they find that on average corporate control value is 14% of the equity value of the firm, ranging from -4% to 65%. The premium paid is higher if the buyer belongs to a country that protects investors in a less intensive way. As La Porta and alii (2000) said, the existence of private benefits has a negative effect on the development of security markets. Kang and Kim (2008) found that block acquirers had strong preference for geographically proximate targets, and that they are more likely to engage in target governance activities that remote block acquirers. Albuquerque and Schroth (2010) estimate that, private benefits, as a mean, represent about 3% - 4% of the target firm's equity value or about 10% of the block's value. They found significant average profits. But their distribution seems highly positively skewed. This means that approximately 40% of trades correspond to private benefits representing less than 1%. Interestingly, they show that private benefits increase with the firm's cash holdings and decrease with short term debt. Strong competition prevents firm controlling coalition from higher private benefits.

6

2. A Contingent Claim Framework for Analyzing Private Benefits

Our analysis adopts the standard continuous time framework of the firm developed by Black, Scholes and Merton (1973, 1974). Our goal here is to extend the seminal setting to account for private benefits. The next section investigates implications. As usual, financial markets are perfect, complete and trading takes place continuously. There are neither taxes, nor transaction costs. There exists a riskless asset paying a known and constant interest rate denoted by r.

We consider hereafter a risky leveraged firm financed by equity and a couple of debts with different priority in case of default. Both debts are viewed as "zero-coupon" bonds maturing at time T. The face value of the senior debt is denoted by F_S, that of the junior F_J. Note that, in what follows, the junior debt may capture the manager's rights to retirement.

As well explained by the contingent claim analysis, the equity is essentially a standard call option written on the underlying firm's assets value whose expiration is the common maturity of debts. The strike price of such an option is the sum of face values: $F_S + F_J$. The firm's asset value[6], denoted by V, is supposed to be correctly described, under the risk neutral measure, by:

$$dV_t = rV_t dt + \sigma V_t dW_t \tag{1}$$

where $W = (W_t)_t$ is a Brownian motion and σ denotes the firm's assets volatility. Here, V_0 stands for the value of the firm's assets in absence of any private benefits.

Standard arguments provide the price of corporate liabilities. The equity price at time 0 is, e.g., given by:

$$Eq_0^{F_S+F_J} = V_0 N\left[d_1^{F_S+F_J}(V_0,\sigma)\right] - (F_S + F_J)e^{-rT} N\left[d_2^{F_S+F_J}(V_0,\sigma)\right] \tag{2}$$

---

[6] The firm's asset value is net of the manager's compensation.

The senior debt is worth $D_0^s = V_0 - Eq_0^{F_S}$ and the junior debt is worth $D_0^j = Eq_0^{F_S} - Eq_0^{F_S+F_J}$, so that the sum of $Eq_0^{F_S+F_J}$, D_0^j and D_0^s is V_0. Let's assume now that equity is shared by a majority block of equity holders and a single minority equity holder, whose respective proportions are θ and $(1-\theta)$. Of course, the fraction of the equity (θ) owned by the majority block is strictly higher than ½. If ever the manager holds a proportion $(1-\beta)$ of the majority block, the majority equity holder owns a $\theta\beta$ fraction of the total equity only. It is then straightforward to imply from this their wealth.

Private benefits essentially lower future incomes of the firm for the benefit of the block-holders' collusion. Denoting by I the present value of all future private benefits, the amount $x_0 = V_0 - I$ represents the current value of the firm's assets in presence of private benefits. We denote by η the fraction of private benefits the majority equity-holder receives. The manager then receives $(1-\eta)I$. Note that, for simulation purposes, one will use base case parameters given in Table 1.

Please Insert Table 1 somewhere here.

A couple of remarks on private benefits deserve to be done. In this paper, private benefits are expressed as a percentage of firm's assets because we assume that the firm size matters. Private benefits are indeed related to well documented corporate behaviour such as the hubris and the empire-building hypothesis. In most simulations, private benefits can reach at maximum ten percent of the firm's assets. When private become even larger, we enter into the world of "tunneling", but this special environment deserves a specific treatment. Let's first study *straight* private benefits that decrease the contemporaneous firm's assets value only (everything else remains the same in the business activities of the firm).

8

3. *Straight* private benefits and consequences on the wealth of stake holders

In case of *straight* private benefits, everything remains the same[7] except that the initial firm's asset value is essentially lowered by the amount of private benefits ($x_0 = V_0 - I$). Dynamics of this firm's assets value is therefore described by:

$$dx_t = rx_t dt + \sigma x_t dW_t \tag{3}$$

and the equity is worth:

$$Eq_0^{F_S+F_J}(I,\sigma) = (V_0 - I)N\left[d_1^{F_S+F_J}(V_0 - I,\sigma)\right] - (F_S + F_J)e^{-rT}N\left[d_2^{F_S+F_J}(V_0 - I,\sigma)\right]. \tag{4}$$

Wealth of the majority equity holders, of the manager and of the minority equity holder are respectively equal to $B_0(I,\sigma) = \eta I + \beta\theta \times Eq_0^{F_S+F_J}(I,\sigma)$, $Mg_0(I,\sigma) = (1-\eta)I + (1-\beta)\theta \times Eq_0^{F_S+F_J}(I,\sigma)$ and $m_0(I,\sigma) = (1-\theta) \times Eq_0^{F_S+F_J}(I,\sigma)$.

The total wealth of equity holders therefore decreases by an amount $\left(Eq_0^{F_S+F_J}(I,\sigma) - Eq_0^{F_S+F_J}\right)$. And, in particular, the minority equity holder incurs a loss of $\Delta m(I,\sigma) = (1-\theta)\left(Eq_0^{F_S+F_J}(I,\sigma) - Eq_0^{F_S+F_J}\right)$. The position of the majority equity holder is relatively more comfortable because they receive part of private benefits (ηI). Of course, he or she will set private benefits so that their net value $\Delta B(I,\sigma) = \eta I - \theta\beta\left(Eq_0^{F_S+F_J}(I,\sigma) - Eq_0^{F_S+F_J}\right)$ remains positive. The manager will receive $\Delta Mg = (1-\eta)I - \theta(1-\beta)\left(Eq_0^{F_S+F_J}(I,\sigma) - Eq_0^{F_S+F_J}\right)$. Creditors suffer from private benefits. The senior creditor faces a decrease of wealth equal to[8]:

$$\Delta D_0^s(I,\sigma) = D_0^s(I,\sigma) - D_0^s = \left[(V_0 - I) - Eq_0^{F_S}(I,\sigma)\right] - \left[V_0 - Eq_0^{F_S}\right] = -I - \underbrace{\left(Eq_0^{F_S}(I,\sigma) - Eq_0^{F_S}\right)}_{<0}$$

The junior creditor loses:

---

[7] Our assumption here implies that the business activities of the firm is continuously sizable and then can be downsized or upsized.

$$\Delta D_0^j(I,\sigma) = D_0^j(I,\sigma) - D_0^j = \left[Eq_0^{F_S}(I,\sigma) - Eq_0^{F_S+F_J}(I,\sigma) \right] - \left[Eq_0^{F_S} - Eq_0^{F_S+F_J} \right] \quad \text{or}$$

equivalently

$$\Delta D_0^j(I,\sigma) = \left[Eq_0^{F_S}(I,\sigma) - Eq_0^{F_S} \right] - \left[Eq_0^{F_S+F_J}(I,\sigma) - Eq_0^{F_S+F_J} \right]$$

Figure 1 displays four graphs showing effects of straight private benefits on the wealth of the different stakeholders. The upper graphs provide relative gains (or losses) enjoyed (or faced) by stakeholders. We compute, e.g. $\left[D_0^s(I,\sigma) - D_0^s \right]/D_0^s$, for senior creditors. The lower graphs provide absolute gains or losses and $\left[D_0^s(I,\sigma) - D_0^s \right]$, e.g., for senior creditors. The left graphs consider the situation where the manager receives a fraction of private benefits $((1-\eta))$ smaller than the portion of the equity they hold $((1-\beta))$. Labels "B", "Mg", "m", "J" and "S" stand for respectively block-holder, manager, minority, junior and senior.

Please Insert Figure 1 somewhere here.

Upper graphs of Figure 1 highlight the differences between majority ("B" and "Mg") and other stakeholders (creditors "J" and "S" and the minority equity holder "m"). As expected, private benefits appear to be a very bad news for both the minority equity holder and creditors in relative and absolute terms. The reason is that all of them suffer from the decrease of the firm's assets value without receiving any portion of private benefits. Note that the majority block holder and the manager also suffer from this deflation but they are largely compensated by receiving private benefits[9]. The lower graphs of Figure 1 highlight that, in absolute terms, creditors suffer less than the minority equity-

---

[8] Hereafter, the exponent on Eq_0^F indicates the face value F to account for.

[9] The upper left graph of Figure 6 clearly shows that the equity value decreases substantially. As a matter of fact, block-holders face this concern in a larger proportion than others. The reason for this is that they own a larger stake of the shares.

10

holder. To fix ideas, let's have a look at quantitative values for $1-\beta = 10\%$ and $1-\eta = 11\%$. This corresponds to the right graphs. If private benefits reach 3% of the firm value, the manager increases his or her wealth by 4.19% whereas the majority shareholder gets +3.25%. In relative terms, the senior creditor loses less than the junior one (-0.34% versus -1.57%). In absolute terms, the loss is almost identical. Finally, the minority shareholder bears the worst situation (-5.13%). As a matter of facts, the minority can feel the loss so substantial that he or she may complain to authorities for expropriation. Of course, these latter will defend minority rights and pursue both the majority holder and the manager. The next section shows how the coalition may behave strategically to render private benefits innocuous for equity holders. We call this the strategic behaviour.

4. A *strategic* way to make private benefits innocuous for every share holder

A *strategic* way to proceed is to act on the characteristics of the activities of the firm so that equity holders remain indifferent to private benefits in terms of "wealth". Minority equity holders will then not be able to claim against majority equity holders. In our (risk neutral) setting, changing the characteristics of the activities of the firm is equivalent to modify its business risk[10]. Dynamics of this firm's assets value can then be described by:

$$dx_t = rx_t dt + \sigma(I)x_t dW_t \qquad (5)$$

Where $\sigma(I)$ is the volatility the manager and block holders will strategically choose. The equity price is now worth: $Eq_0^{F_J+F_S}(I, \sigma(I))$ So that the total equity holders' wealth is $B_0(I, \sigma(I))$, $Mg_0(I, \sigma(I))$, $m_0(I, \sigma(I))$. Theoretically, the wealth of equity holders (viewed as a single group) decreases by an amount equal to

---

[10] The vega of an option is strictly positive so that the wealth of share holders increases as the firm's assets volatility gets larger.

11

$\left| Eq_0^{F_S+F_J}(I, \sigma(I)) - Eq_0^{F_S+F_J} \right|$. For some stakeholders, this loss is compensated by receiving part of the private benefits. But minority equity holders are more or less expropriated. A strategic solution for the manager (acting for the best of the interests of block holders) consists in changing the business risk to avoid significant (and visible) expropriation of minority. The result is to find $\sigma'(I)$ such that:

$$Eq_0^{F_S+F_J}(I, \sigma'(I)) - Eq_0^{F_S+F_J} = 0 \tag{6}$$

Expected consequences of the strategic behaviour have various effects. First, if the wealth of the minority equity holders remains the same, they will not be able to complain for expropriation anymore… Second, the majority share holders and the manager avoid dramatic consequences of the private benefits on their own financial assets.

Figure 2 displays four graphs highlighting effects of strategic private benefits on the wealth of stakeholders. The upper graphs provide relative gains or losses enjoyed or faced by stake holders. We compute, e.g. $\left[D_0^s(I, \sigma'(I)) - D_0^s \right] / D_0^s$, for senior creditors. The lower graphs provide absolute gains or loss. We compute, e.g. $\left[D_0^s(I, \sigma'(I)) - D_0^s \right]$, for senior creditors. Every thing else remains the same as Figure 1.

Please Insert Figure 2 somewhere here.

Graphs clearly show that the considered strategy makes private benefits innocuous for the minority equity holder. The wealth of block holders (majority equity holder and manager) increases by a substantial amount, at the expense of creditors. Senior creditors face larger loss than juniors in absolute terms. The reverse is true however in relative terms. When private benefits reach 3% of the firm value (with $1-\beta = 10\%$ and $1-\eta = 11\%$), the manager's

12

wealth increases by 9.32%. That of the majority shareholder gets +8.38%. Minority shareholder now loses nothing. Both creditors lose money, in a different way however. The senior creditor faces -3.79% (only) when the loss of the junior creditor is -12.95%. Adjusting volatility makes effects of private benefits more pronounced than before at the expense of debt holders.

The situation considered here may evoke the asset substitution problem of agency theory, as well as the overinvestment issue[11]. In its most conventional form overinvestment means that negative NPV projects may be favoured by managers because they have positive effects on a certain category of investors. For their part, creditors are not however without resources. First of all, shorten debt maturities should be worth. Some forms of covenants may attenuate or prevent temptations of private benefits too. The covenant suggested by Bhanot and Mello (2006) may, e.g., help. When the credit risk (captured by the rating) changes, this covenant allows creditors to precipitate the early partial or total repayment of the debt, to require a cash infusion from share holders, or to renegotiate the general conditions of the granted loans. Increasing the business risk of the firm clearly changes the debt value and the debt ratio. It may therefore cause the rating to change.

As final words of this section, it is worth mentioning tunneling. Tunneling may be viewed as very extreme situation where block holders capture privately a very large piece of the firm's value. Figure 3 draws the corresponding volatility as a function of very large private benefits. Unreasonable values of private benefits highlight a nonlinear impact on the volatility. Various qualitative results are found in Figure 3. First of all, it can be observed that effort to do is non linearly affected by the level of private

---

[11] It may also evoke dividend policy too (Brookman and Unlu, 2009), since a large (public) dividend is known to decrease the value of the firm and the debt value. Private benefits behave almost the same except that they are dedicated to only a certain coalition of share holders.

13

benefits. Second, given a firm, there exists a maximum level of private benefits strategically manageable. Third, shortening the maturity of the debt (in absence of retirement packages) does not imply the expected result as it lowers the level of the required volatility.

Please Insert Figure 3 somewhere here.

5. Retirement packages offered to managers matter

Let's now assume that the manager's incentives apparently diverge from the sole block holders wealth by considering that he or she plans to benefit from a retirement compensation (Sundaram and Yermack, 2007) designed as a fraction of the junior debt. In case of default senior creditors are paid first. Here, the manager's wealth is:

$$Mg_0 = (1 - \eta)I + (1 - \beta)\theta \times Eq_0 + (1 - \alpha) \times J_0$$

Figure 4 first illustrates effects of the retirement package on the straight private benefits, the manager can receive. As in Figure 1, these graphs provide relative gains (or losses) enjoyed (or faced) by stake holders when the manager can benefit from a retirement package. Graphs of figure 4 obviously show that only the manager's wealth is negatively impacted. The reason is that the junior debt value decreases with the amount of private benefits.

Please Insert Figure 4 somewhere here.

It is then clear that the manager has strong incentives to change the business risk not only to avoid significant expropriation of minority shareholders but also to prevent undesirable effects on his or her entire wealth including retirement packages. The result is to find $\sigma''(I)$ such that:

$$(1 - \beta)\theta[Eq_0(I, \sigma''(I)) - Eq_0] + (1 - \alpha)[J_0(I, \sigma''(I)) - J_0] = 0$$

14

We can then expect from his or her strategic behaviour several effects. First of all, if the wealth of the minority equity holders remains the same, they will not be able to complain for expropriation anymore… Second, the majority holders and the manager will avoid dramatic consequences of the private benefits on their own financial assets. Third, in order to compensate the decrease of value of the retirement package, the manager will increase the level of volatility more than $\sigma'(I)$. This is once again good news for equity holders. They (especially the minority equity holders) will benefit from this incentive to increase volatility. Figure 5 displays four different graphs that consider effects of strategic private benefits in presence of retirement packages.

Please Insert Figure 5 somewhere here.

Upper graphs of Figure 5 provide relative gains/losses, lower ones the absolute gains/losses. It appears that creditors face a more severe problem than before, the junior lenders loosing more that the senior one. The CEO owns part of the junior debt but he or she counterbalances this loss with the benefit extracted on the equity side. His or her wealth remains globally unchanged.

Just to go on our numerical illustration, when private benefits reach 3% of the firm value, the manager increases by 7.36% his or her wealth, the majority equity holder gets +14.91% on a relative basis, while the minority one obtains +6.53%. For their part, creditors face a significant loss: -8.84% for the senior and -24.58% (!) for the junior one. As anticipated, the relative wealth spreads are more pronounced that before at the expense of debt holders. The minority equity holder now takes benefits from the manager's behaviour because this latter decides to increase the volatility far above what compensate for private benefits only. This can also be viewed as a consequence of the subordinated status of the manager's retirement packages relatively to minority

15

rights. The presence of retirement package appears is here much favourable for the minority equity holder.

To sum up previous results, let's compare the different scenarios. Three graphs of Figure 6 compare losses or gains faced by stake holders and use the same scale (Billet and alii, 2010). They highlight consequences of private benefits in a much different ways than before. The lower right graph completes the analysis by comparing the corresponding firm's assets volatility.

Please Insert Figure 6 somewhere here.

The three graphs show that straight private benefits are not so damageable for creditors compared to what happens when private benefits are accompanied with strategic manager's behaviour. In case of straight benefits, equity is mostly and negatively impacted. When this effect is alleviated by changing the firm's assets volatility, both junior and senior debts are deeply impacted. This situation will even worsen if the manager behaves strategically to account for his retirement package. This lower graph suggests that granting a retirement package to the manager is worth for equity holders. For its part, the lower right graph illustrates the way the firm's asset volatility should be increased. It is worth noting that obtained values are fairly reasonable and credible. In addition, the case $\sigma''(I)$ with $(1-\alpha)$=5% highlights the fact that the suitable volatility is not linear with respect to $(1-\alpha)$.

Conclusion

Private benefits take place each time when the value extracted by the controlling shareholder group is not shared among all shareholders in proportion of the stocks owned. Our simulations showed that large benefits may be obtained by this group at the expense of others. We show that it is

16

possible to alleviate the negative effects of private benefits extraction on equity by increasing the business risk of the firm at a specified level. Then the conflict of interest between minority and control share holders vanishes. But for sure this manipulation is detrimental to bondholders or creditors (senior and junior) if they don't take into account this behaviour. Then we switch to another conflict of interest implying lenders and share holders Nevertheless, creditors can counterbalance these negative effects if firm's credit rating is affected by the risk increase with ad hoc debt covenants. When the manager or the CEO holds a fraction of junior debt as a retirement plan for example, it is interesting to note, if he develops a strategic behaviour in presence of retirement packages, that minority shareholders may benefit of it, but not the debt holders who bear heavier losses than before.

Our simulations suggest that some environments can be more favourable than other for private benefits. We can expect the size of the firm and that of the free cash flows (Jensen, 1986) to matter as well as the business profile of the firm. We can call to mind the empire-building inclination of managers and the propensity to changing the business profile to more intangible assets. LBO environments should be especially monitored because managers have generally strong position in the equity. To foster private benefits, firms can even decide to go darkness and to be delisted (Marosi and Massoud, 2007). To solve such agency problems, strip financing could obviously help. But this implies that creditors are willing to become equity holders and that equity holders accept to welcome *less easy to manage* investors.

Note, finally, that our paper sheds also a new light on the CEO compensation issue and connections it can have with private benefits. Clearly, the more difficult the CEO obtains a visible compensation from company, the more seriously he or she will take into consideration the opportunity of private benefits. It would even not be a surprise that this becomes a significant part of

17

the global compensation he or she can expect. Unfortunately, this global compensation remains difficult to observe.

18

References

Albuquerque R, E Schroth, « Quantifying private benefits of control from a structural model of block trades », 2010, *Journal of Financial Economics*, 96, p 33 - 55.

Atanasov V, B. Black, C. Ciccotello, S. Gyoshev, "How does law affect finance? An examination of equity tunneling in Bulgaria", 2010, *Journal of Financial Economics*, 96, p 155 - 173.

Baek J.S, J.S Kang, J.K Lee, "Business groups and tunneling: evidence from private securities offerings by Korean Chaebols", 2006, *Journal of Finance*, 61, p 2415 - 2449.

Barclay M, C Holderness, "Private benefits from control of public corporations", 1989, *Journal of Financial Economics*, 25, p 371-395.

Bhanot K, A.S Mello, "Should corporate debt include a rating trigger? ", 2006, *Journal of Financial Economics* 79, p 69 – 98.

Billett, M.T, D.C Mauer,Y Zhang, "Stockholder and bondholder wealth effects of CEO incentive grants", 2010, *Financial Management*, été, p 463 -487.

Brockman P, E. Unlu, "Dividend policy, creditor rights and the agency costs of debt", 2009, *Journal of Financial Economics*, 92, p 276 - 299.

Black F, M Scholes, "The pricing of options and corporate liabilities", 1973, *Journal of Political Economy*, p 637 – 654.

19

Burkart M, D Gromb, F Panunzi, " Agency conflicts in public and negociated transfers of corporate control", *Journal of Finance*, 2000, 55, p 647 - 677.

Cronqvist H, R. Fahlenbrach, "Large shareholders and corporate policies", *Review of Financial Studies*, 2009, 22, p 3641-3976.

Djankov S, R. La Porta, F.Lopez-de –Silanes, A. Schleifer, «The law and economics of self-dealing », 2008, *Journal of Financial Economics*, 88, p 430 - 465.

Dyck A, Zingales L, "Private benefits control: an international comparison", *Journal of Finance*, 2004, 59, p 537 – 599.

Kang, J.K, J-M Kim, "The geography of block acquisitions", 2008, *The Journal of Finance*, 53, 6, p 2817 – 2858.

Jensen M, "The Agency cost of free cash flow: Corporate finance and takeovers", 1986, *American Economic Review,* 76, p 323 - 330.

Jensen. M, W.H Meckling, "Theory of the firm: Managerial behaviour, agency costs and ownership structure", *Journal of Financial Economics*, 1976, 4, p 305 - 360.

Johnson S, La Porta R, F Lopez-de-Silanes, A Shleifer, "Tunneling", *American Economic Review*, 2000, Vol 90, p 22 - 27.

La Porta R, F Lopez-de-Silanes, A Shleifer, "Investor protection and corporate governance", *Journal of Financial Economics*, 2000, Vol 59, p 3 - 27.

Marosi A, Massoud N, "Why do firms go dark?" *Journal of Financial and Quantitative Analysis*, 2007, 42, p 421 - 442.

Masulis R, C, Wang, F Xie, "Agency problems at dual-class companies", *Journal of Finance*, 2009, 64, p 1697 - 1727.

Merton R. "On the pricing of corporate debt: The risk structure of interest rates", 1974, *Journal of Finance*, 29, p 449 - 470.

Villalonga, B., and Amit, R., 2009. How are U.S. family firms controlled? *Review of Financial Studies,* 22, p 3047 - 3091.

Villalonga, B., and Amit, R, "How do family ownership, control, and management affect firm value?", 2006, *Journal of Financial Economics* 80, p 385 - 417.

Shleifer, A., Vishny, R, "A survey of corporate governance", 1997, *Journal of Finance* 52, p 737 - 783.

Sundaram R.K, D. L Yermack. (2007) "Pay me later: Inside debt and its role in managerial compensation", *Journal of Finance*, 62, 1551-1588.

21

Table 1: Base case parameters

Initial Firm's assets value	V_0	100
Initial volatility	σ	30%
Risk-free interest rate	r	5%
Block holder's rights (in % of equity)	θ	2/3
Minority's rights (in % of equity)	$1-\theta$	1/3
Manager's fraction of majority block	$1-\beta$	10%
Debt maturity	T	5
Senior Face Value	F_s	50
Junior Face Value	F_J	15
Manager's fraction of junior debt (retirement package)	$1-\alpha$	10%
Manager's fraction of private benefits	$1-\eta$	9% or 11%
Majority equity holder's fraction of private benefits	η	89%

22

Figure 1: Effects of straight private benefits on stakeholders' wealth (with no business adjustment)
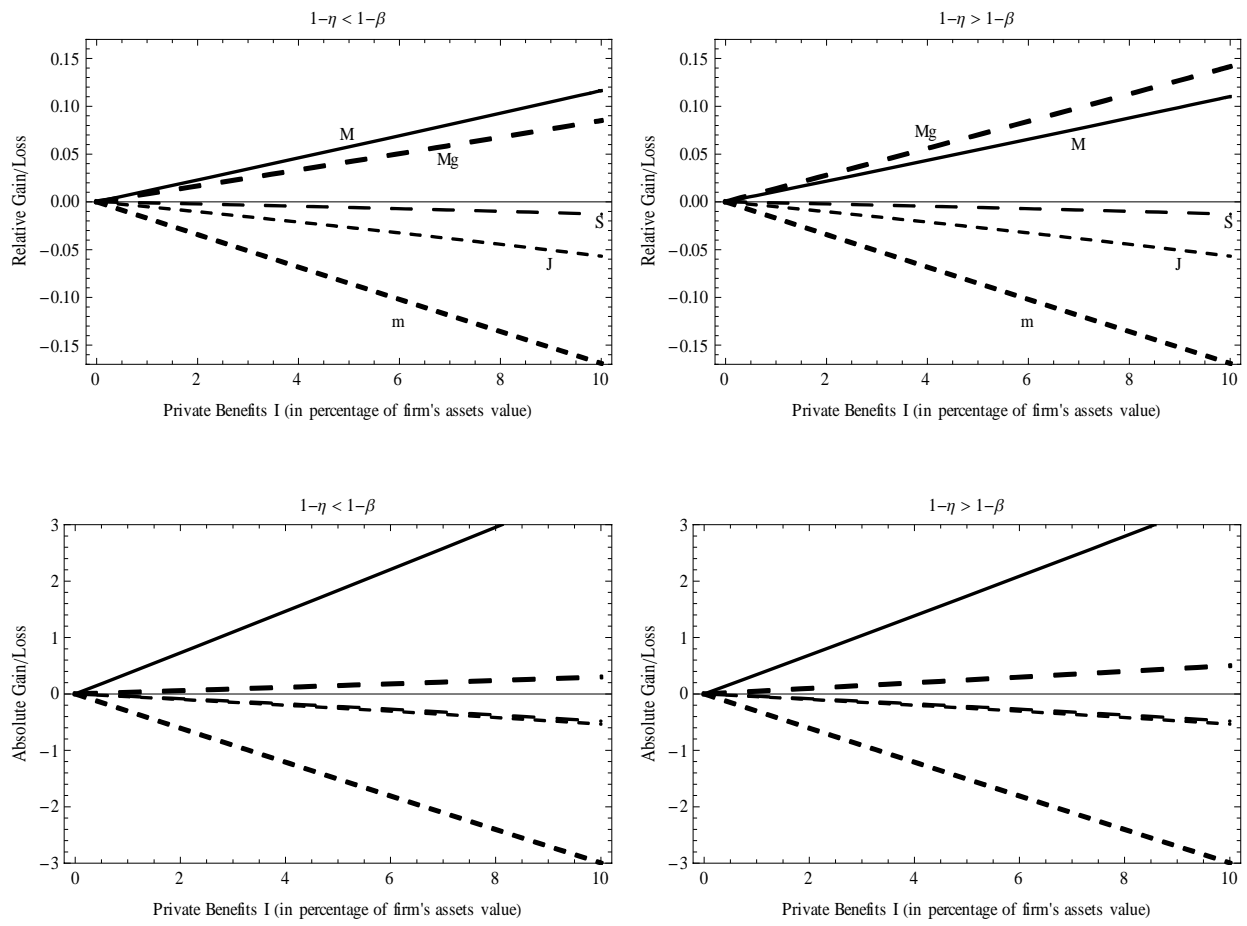
Base case parameters are given in Table 1. Share holders (minority, majority and the manager) are signalled by bold lines.

23

Figure 2: Effects of Private Benefits on stakeholders' wealth with a business adjustment (in absence of retirement packages)
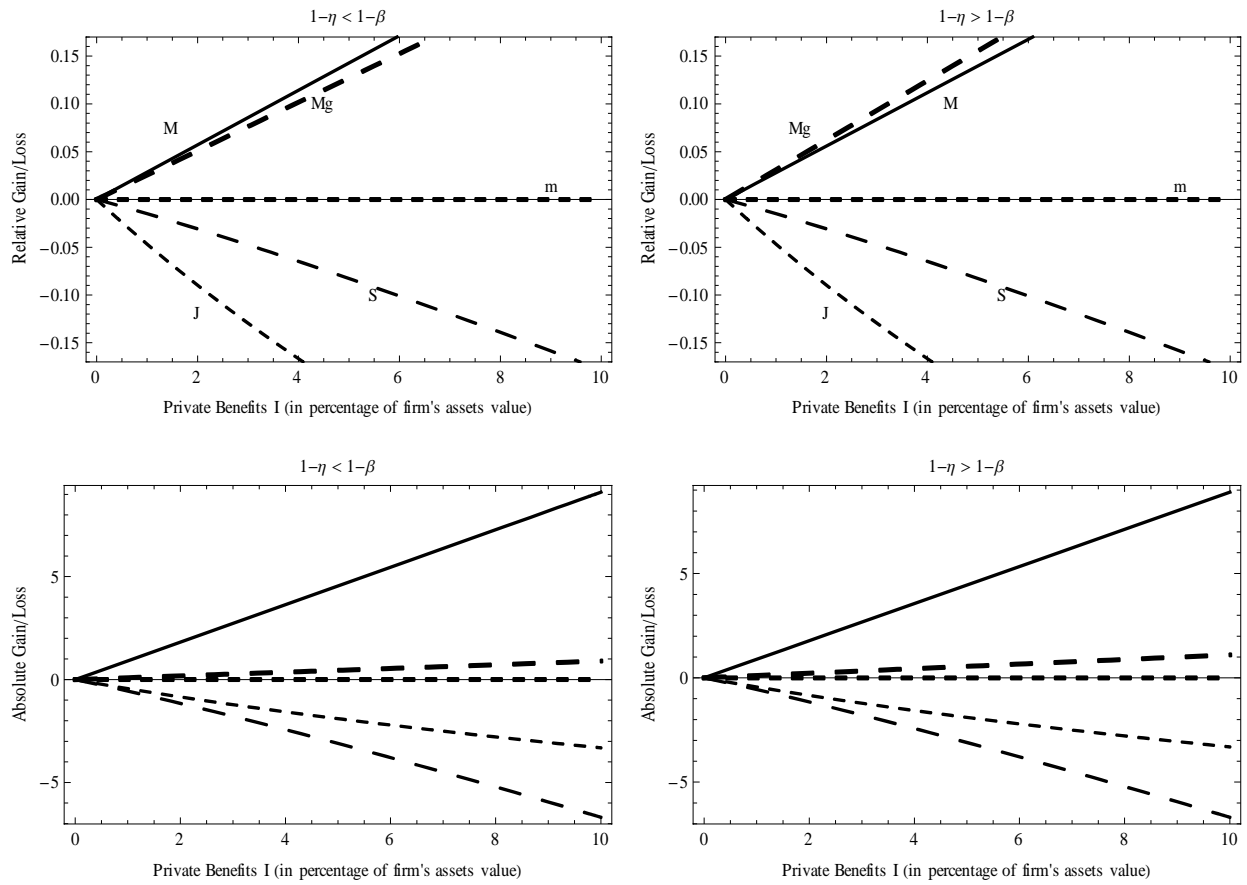
Base case parameters are given in Table 1. Share holders (minority, majority and the manager) are signalled by bold lines.

Figure 3: From Private Benefits to tunneling
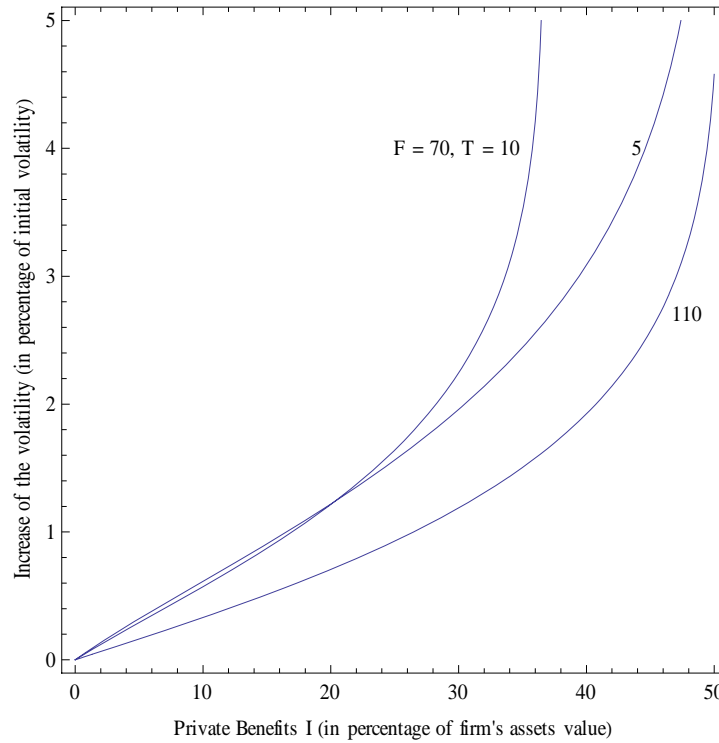
This graph shows the volatility to choose in order to make up the private benefits (in absence of any retirement package). Most parameters are inspired from Table 1, except the maturity and the face value of the total debt. "F=70, T=10" serves here as a benchmark. The label "5" mentions that the maturity is lowered everything thing else being equal. Similarly the label "110" mentions that the face value of the total debt is 110 everything thing else remaining the same.

25

Figure 4: Straight private benefits and retirement package
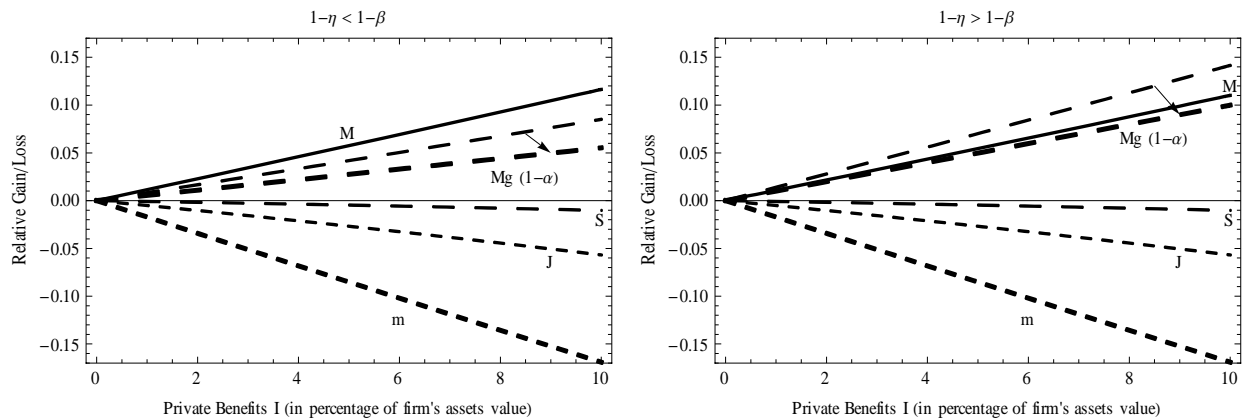
These graphs reconsider upper graphs of Figure 1 and highlight consequences of the retirement packages on the manager's wealth.

26

Figure 5: Effects of Private Benefits on stakeholders' wealth in presence of a retirement package when the business is adjusted strategically
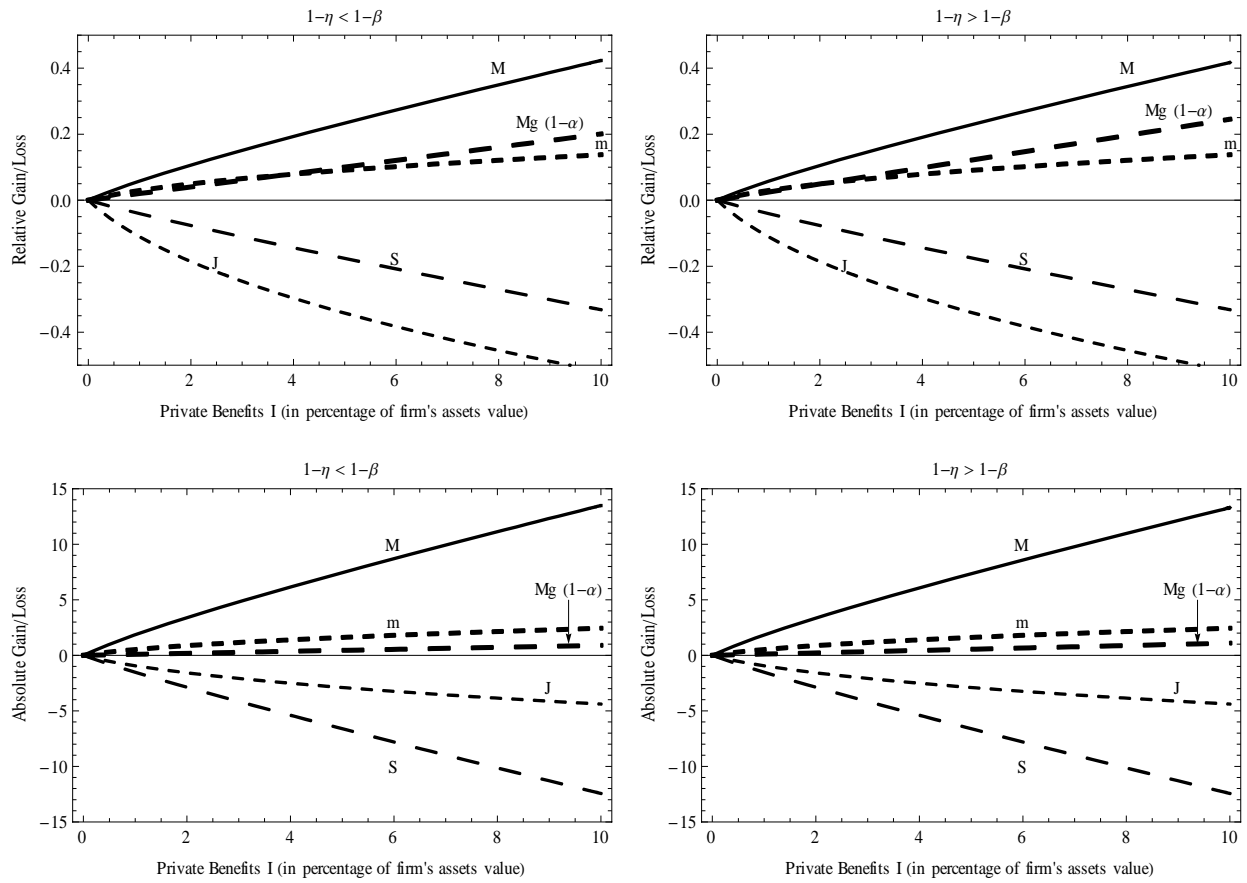
These graphs are equivalent to those of Figures 1 and 2. Share holders (minority, majority and the manager) are signalled by bold lines.

27

Figure 6: Comparing Consequences of Private benefits
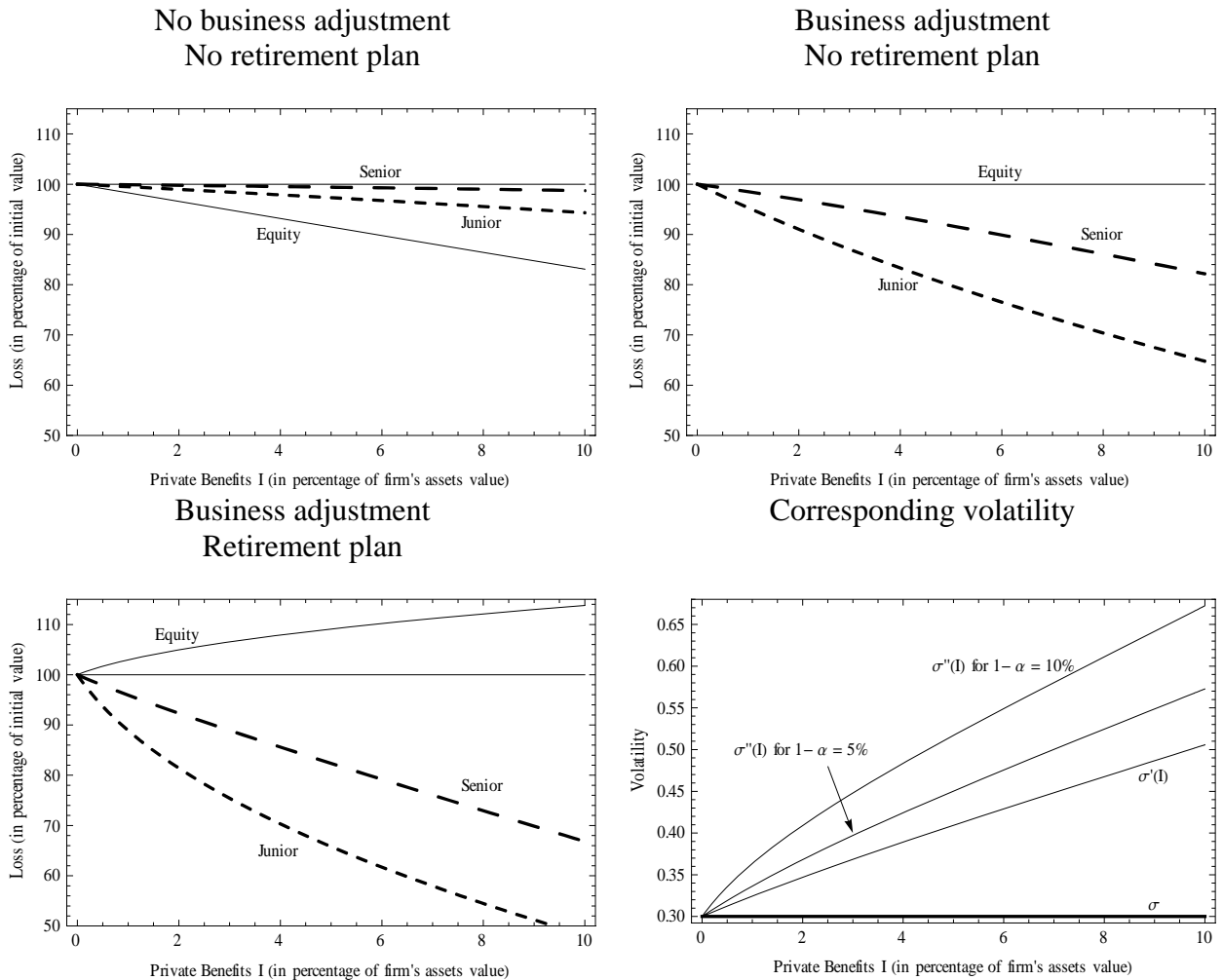
These graphs put together results obtained under the different scenarios. Three of them focus on the financial impacts of private benefits on values. The lower right graph plots the associated volatility. One observes there the initial volatility (σ =30%) that corresponds to the upper left graph, the adjusted volatility $\sigma'(I)$ and the adjusted volatility $\sigma''(I)$ obtained in absence or presence of a retirement plan respectively. $\sigma'(I)$ is associated to the upper right graph. $\sigma''(I)$ with $(1-\alpha)$=10% corresponds to the lower left graph. $\sigma''(I)$ with $(1-\alpha)$=5% illustrates that the suitable volatility is not linear with respect to $(1-\alpha)$.

28

A STUDY OF TIAA REIT

Malek Lashgari

University of Hartford
Department of Economics, Finance, and Insurance
Barney School of Business
200 Bloomfield Ave.,
West Hartford, CT 06117

Lashgari@hartford.ed

860-768-4655

The Teachers Insurance and Annuity Association – College Employees Retirement Fund (TIAA-CREF) provides various investment choices including real estate (REIT), inflation protection bonds (TIPS), and common stock (STOCK). A Real Estate Investment Trust (REIT) is a portfolio of investments in real estate properties known as Equity REIT. Equity REITs are traded on the financial markets in the same manner as common stock. In this way, a marketable security is designed by the fairly illiquid asset.

This study examines the performance of TIAA REIT in relation to changes in real estate prices, inflation and the real return in the economy during 1997-2010. The time horizon for the study is due to the availability of data since inflation protection bonds were issued by U. S. Treasury in January 1997. The study is further extended to MIT REIT database, publicly available from MIT.

Sources of Data

Changes in the real estate prices are based on the composite index of real estate transactions taken from Standard and Poor's / Case-Shiller data base. Inflation hedging properties of real

estate properties are tested with both the Consumer Price Index (CPI) taken from the Bureau of Labor Statistics (BLS) and TIPS from TIAA-CREF website. Meanwhile, the Treasury Inflation-protected bond provides a risk free asset reflecting the real return as well as changes in the market rate of interest, both of which are expected to influence real estate prices.

The Design of the Study

Statistical tests consisting of correlation and regression equations are performed as follows:

<u>Model 1</u> examines the relationship between successive changes in TIAA REITs and successive changes in market-based house prices.

$$\Delta \text{REIT} = a + b.\Delta \text{House} + e \qquad\qquad (1)$$

REIT and House values are based on market prices.
Δ denotes the amount of change. The sign of (b) is expected to be positive and statistically as well as economically significant. The results for Model 1 are shown in Table 2.

<u>Model 2</u> examines the relationship between successive changes in TIAA REITs and successive changes in common stock prices.

$$\Delta \text{REITs} = a + b.\Delta \text{STOCK} + e \qquad\qquad (2)$$

The sign of (b) is expected to be positive as REITs prices are market-based and are influenced by environment of the stock market. The results for Model 2 are shown in Table 2.

Model 3 examines the effects of changes in inflation as well as changes in the rate of interest in the market. The results for Model 3 are shown in Table 2.

$$\Delta \text{REITs} = a + b.\Delta \text{TIPS} + e \qquad\qquad (3)$$

TIPS denote prices for U. S. Treasury inflation protection bonds.

Model 4 examines the role of changes in inflation as measured by CPI on REIT prices. The results for Model 4 are shown in Table 2.

$$\Delta \text{REITs} = a + b.\Delta \text{CPI} + e \qquad\qquad (4)$$

Model 5 shows the relationship between TIAA REIT and MIT REIT as follows.

$$\Delta \text{TIAA REIT} = a + b.\Delta \text{MIT REIT} + e \qquad\qquad (5)$$

Empirical Results

Correlations between TIAA REIT and various financial variables are shown in Table 1.
In addition, the results of regressions of TIAA REIT with these variables are shown in Table 2.
The results for Model 1 indicate that changes in transaction based house prices explain about 9 percent of variance in changes in TIAA real estate retirement annuity unit values; its beta coefficient of 0.53 further shows that a one unit change in house prices is associated with 0.53 units change in TIAA REIT unit value. These results are statistically significant at the 5 percent confidence level.

Table 1: Correlation Between Changes in REIT's with Other Assets
1997-2010

Factors	TIAA REIT	MIT REIT Index
Changes in House Price	0.30	0.24
Changes in Stock Price	0.04	0.51
Changes in TIPS	-0.15	0.37
Changes in CPI	0.24	0.13
MIT REIT Index	0.10	1.0

TIAA REIT further appears to be a good hedge against inflation as shown by the results from model 4. Changes in inflation as shown by CPI explain 6 percent of variance in TIAA REIT. The beta coefficient of 0.97 shows an almost perfect matching in changes in inflation and TIAA REIT unit value. These statistics are statistically significant at the 5 percent level.

Changes in TIAA REIT unit values and changes in the Treasury Inflation Protection Bands (TIPS) reveal a negative beta. This result is surprising since TIPS show the effects of inflation and changes in interest rates. Equity REIT indexes examined here show different pattern of movements with economic and financial factors as shown by a 0.10 correlation with each other. For example, TIAA REIT moves inversely with TIPS with a correlation of - 0.15 whereas MIT REIT follows TIPS with a correlation of 0.37. Similarly, changes in TIAA REIT are independent of the stock market while MIT REIT has a correlation of 0.51.

Table 2: Results of Statistical Regressions for
TIAA Real Estate Retirement Annuity
1997-2010

Statistics	Model 1 House Price	Model 2 Stock Price	Model 3 TIPS	Model 4 CPI	Model 5 MIT REIT
Intercept→	0.24	0.51*	0.75*	0.16	0.51*
t-value→	0.96	2.01	2.68	0.60	2.03
p-value→	0.34	0.05	0.01	0.55	0.04
Beta Coefficient	0.53*	0.01	-0.75**	0.97*	0.01
t-value	4.0	0.44	1.85	3.09	1.21
p-value	0	0.66	0.07	0	0.23
F-Statistics	16.01	0.19	3.41	9.55	1.47
R^2	0.09	0	0.02	0.06	0.01

*denotes significance at the 5 percent level.
** denotes significance at the 10 percent level.
Data are for the time horizon covering January 1997-June 2010.Total observations are 161 months.

The varying behavior of TIAA REIT and MIT REIT with respect to the stock market and inflation may, in part, be due to their structure; TIAA REIT is an annuity accumulation unit whereas MIT REIT index reflects a regular marketable asset. In addition, TIAA REIT imposes severe restrictions on frequent trading. Table 3 shows the results of regressions of MIT REIT onto the same financial variable. As shown in Table 3, changes in MIT REIT prices show a strong relationship with market prices for real estate properties, common stock, and TIPS. In addition their reactions to changes in inflation appear to be strong but with a weak statistical significance (0.10 significance level).

Table3: Results of Statistical Regressions of MIT REIT;
1997-2010

Statistics	Model 1 House Price	Model 2 Stock Price	Model 3 TIPS	Model 4 CPI	Model 5 TIAA REIT
Intercept→ t-value→ p-value→	- 1.25 0.67 0.51	- 0.44 0.28 0.78	- 4.11* 2.11 0.037	- .99 0.49 0.63	0.03 0.02 0.99
Beta Coefficient t-value p-value	3.18* 3.18 0.002	1.50* 7.40 0.0	14.12* 5.01 0.0	3.77** 1.60 0.11	0.71 1.21 0.23
R^2	0.06	0.26	0.14	0.02	0.01

*denotes significance at the 5 percent level.
** denotes significance at the 10 percent level.
Data are for the time horizon covering January 1997-June 2010.Total observations are 161 months.

Conclusions

TIAA REIT appears to be a good choice for inclusion of real estate in an investment portfolio of common stock and TIPS. While mildly correlated with changes in real estate prices and inflation, it does not behave like common stock. In contrast, the MIT REIT shows a medium to strong positive relationship with prices of real estate properties, common stock price, and TIPS.

THE ETHICS OF YIELD SPREAD PREMIUM: A CASE STUDY OF ILLINOIS MORTGAGE BROKERS/BANKERS

James F. (Jim) Hoelscher, Lincoln Memorial University,:james.hoelscher@lmunet.edu
Luis Eduardo Rivera-Solis, Dowling College, riveral@dowling.edu

ABSTRACT

Illinois mortgage brokers/bankers suffered severe setbacks during the current economic downturn that began in 2007. Ethics among this group of mortgage lenders is a more significant topic of analysis during this time, as more and more instances of fraud surfaced amongst mortgage brokers/bankers even as they were going out of business at record numbers. This study analyzed the potential impact on ethical behavior of those firms licensed by the state of Illinois to function as mortgage brokers/bankers. Analysis of the demographic factors surrounding these firms included analyses of variance and regression analysis to determine if the demographic characteristics of this group affect ethical behavior. Analysis reveals that those mortgage employees who control charges assessed to borrowers, such as yield spread premium, exhibit the tendency to abuse ethical behavior. In addition, demographic factors such as length of time employed by their respective mortgage firm, overall compensation structure and intent to remain with their current firm impact evident ethical behaviors. Further research should be conducted on not only Illinois mortgage brokers/bankers, but also a broader segment of the mortgage industry overall to determine the extent of these impacts on ethics within the mortgage business. Further research should be conducted on the possibility of adjusting compensation structures to eliminate potential sources of ethical abuse.

Keywords: Mortgage Banking, Financial Ethics

INTRODUCTION

In the past several years, mortgage lending has come under severe scrutiny. Many practices, some of them questionable and occasionally unethical, within the mortgage lending industry have lead to these difficulties. Significant downturn in the overall economy in 2007 and 2008 can be attributed primarily to subprime lending problems and are driving proposed regulatory changes (Vartanian, 2008). One of the practices that has generated much discussion is the payment of yield spread premiums (YSP) to mortgage brokers/bankers.

Yield spread premium is an amount paid by a lender to mortgage brokers/bankers to deliver a completed loan package at an interest rate in excess of the rate that the lender normally wishes to charge. Interest rates, established by various lenders to whom brokers/bankers refer their loans, are established as "par"; that amount below which the lender will not lend. The payment of this one-time YSP to brokers/bankers produces a loan product that will yield higher interest returns for the lender for many years to come, sometimes for the life of the loan. For example, an increase of 100-200 basis points (1 to 2 %) in the interest rate paid by the borrower may produce a one-time payment to the broker/banker of 2% of the loan amount. In this example, a loan of $150,000 will pay the broker a

fee of $3,000 at closing. The one-time payment of $3,000 in this example then results in an increase of 1% in the interest rate the borrower pays. In the first five years of this loan, the lender will realize additional revenue of approximately $7,000. This $3,000 fee is disclosed to the borrower as a payment to the broker/banker on the Federal Housing Administration Department of Housing and Urban Development (HUD) HUD-1 disclosure form, but the method is so complex that less than one out of ten individuals has any real idea what is taking place in their transaction (Jackson & Burlingame, 2007). It is interesting to note that this process exists for all types of mortgage loan products including those commonly referred to as A, Alt-A, FHA, VA and subprime.

The imposition of this higher interest rate to the borrower, argued by mortgage brokers/bankers, reduces other fees and charges. However, a detailed study by Jackson and Burlingame (2007) shows that this is not the case. Consumers simply end up paying more because of the confusing yield spread premium disclosure requirements brought on by the Real Estate Settlement Procedures Act (RESPA) enacted in the early 1970s. Rather than mandate any fee or interest rate ceilings through the enactment of RESPA, regulators decided to use the disclosure process to attempt to control these costs (Jackson & Burlingame, 2007). However, the disclosure process for yield spread premiums is buried on the back page of the HUD-1 disclosure form and disclosed as "paid outside of closing" (POC) (Jackson & Burlingame, 2007). Additionally, some brokers/bankers use this knowledge to gain an upper hand in negotiations with borrowers. Knowing that disclosure is confusing they sometimes rely on shady sales pitches to entice the borrower to utilize their services. In one case, a broker advised a customer that he could "offer them lower points by getting the lender to pay a portion of the broker's fee for them" (Anonymous, 2008). He failed to advise the potential borrower that this payment by the lender to the broker would result in higher interest charges to the borrower for the life of the loan.

With the initiation of RESPA in the early 1970s, HUD attempted to adjust this disclosure process to make it easier for borrowers to understand the true costs of their loans. Various segments of the mortgage industry, most notably mortgage brokers/bankers, have successfully defeated most of these attempts. The current HUD proposal appears to do little to resolve the disclosure difficulties, as the HUD-1 form has simply been transformed from two pages to four, with new wording that will confound borrowers as much as the current disclosure process (Schulman & Spencer, 2008).

There are some instances where the utilization of the yield spread premium can be beneficial to homeowners (de la Torre & McClatchey, 2006). Those borrowers who prefer not to use out of pocket cash to pay for customary closing costs can rely on yield spread premium to do the job for them. Many times these transactions are advertised as "no closing cost" or "no points" loans. In these situations, the broker/banker simply adjusts the interest rate upward high enough to produce a payment of yield spread premium adequate to absorb all costs. The borrowers utilizing this approach are generally sophisticated enough to understand what is taking place. This complex and difficult disclosure process, however, confuses the majority of borrowers, regardless of attempts by various federal regulators to improve these disclosures (de la Torre & McClatchey, 2006).

Mortgage brokers/bankers thus have the upper hand in the vast majority of these negotiations. The National Association of Mortgage Brokers (NAMB, 2008) again expressed its disappointment in the

final HUD revision of the RESPA disclosures released in November 2008. Revising the disclosure form from two pages to four simply further complicates the disclosure process (NAMB, 2008). Borrowers then need to fend for themselves when dealing with mortgage brokers/bankers. They must rely on the ethical values of the individuals/companies in whom they place their trust. Individual loan officers are oftentimes at a significant advantage in those dealings due to the confusing state of affairs created by current disclosure requirements, as well as their knowledge of how to define and describe the various fees and charges the borrower must navigate (de la Torre & McClatchey, 2006). The imposition of these YSPs is generally at the discretion of the individual loan officer or may be an established policy of the mortgage broker/banker.

BACKGROUND OF THE STUDY

Since at least 60% of new mortgage originations in the U.S today are processed through mortgage brokers/bankers (NAMB, 2008), it is clear that a large portion of borrowers today rely on their mortgage broker/banker to obtain for them the best possible terms for their mortgage. Since the knowledge possessed by brokers/bankers surpasses individual borrowers due to the complexity of mortgage transactions, knowing that firms and/or individual loan officers subscribe to a clear and professional code of ethics is an important part of the transaction. As such, various trade associations have developed codes of ethics for their members' compliance. The Illinois Association of Mortgage Professionals (IAMP) has adopted an eight-page code that clearly defines what actions the association expects of their membership (IAMP Code of Ethics, 2008). This code, however, appears to lack teeth, as Illinois ranks third highest in mortgage fraud in the nation (Christie, 2008).

Attempts on a national level to establish ethical standards by the NAMB may be even weaker. Their code is a one-page document outlining these six areas for adherence:
-Honesty and Integrity
-Professional Conduct
-Honesty in Advertising
-Confidentiality
-Compliance with Law
-Disclosure of Financial Interests

Each of these six areas is detailed with a one-sentence description (NAMB Code of Ethics, 2008). Adherence to these various codes of ethics may be lacking, as more and more borrowers are paying costs, such as yield spread premium, of which they neither understand nor are completely aware of (de la Torre & McClatchey, 2006).

The ethical climate within various mortgage broker/banker firms may be a contributing factor. Victor and Cullen (1988) studied 872 employees across four firms and found that ethical climates are both multidimensional and multidetermined. This study showed differences within ethical climates and within various departments in the same firms. Differences within mortgage broker/banker firms may be based on different strategic approaches to the same line of work. Mortgage brokers/bankers generally compete for customers from within the same general pool, although refinancing and purchases tend to be handled differently. Differences may also arise due to different product

offerings, such as subprime, FHA or conforming loan products. These differences may create different ethical perceptions from firm to firm.

Various compensation practices within the mortgage broker/banker industry may also create differing ethical climates and behaviors (Helm, 2008). Most mortgage firms tend to be small businesses, employing from three to twenty individuals. Owners of mortgage firms may be more concerned with total overall revenue for the firm, whereas most loan officers are concerned with their individual compensation package. The drive and need for these sources of revenue, for both loan officers and the firms they work for, may create unethical behavior. Businesses often pursue courses of action due to the need for revenue, and behavior of employees can be altered accordingly. An example is the Pacific Lumber takeover in 1988. In this case, the purchasers of the company undertook such massive debt that to repay the debt they began a process of clear-cutting of the old stand Redwood forests. Without the need for those funds, their actions may have not endangered the environment as they did. The previous owners of the firm had employed a practice of selective cutting that still produced profits, but the new owners had amassed significant debt and needed much more revenue. This need drove their legal, but very controversial and likely unethical, behavior (Lindsey, 1988).

An analysis of the ethical climates existing within mortgage firms and among the employees within those firms may provide a glimpse into the reasoning for a variety of interactions with borrowers. In addition to the ethical climate in existence at these firms, individual ethical perceptions may guide the work of these employees, especially loan officers, who are compensated partially or solely by commission. Loan officers tend to justify charging the fees they do, through whatever means available to them, as practices that do not violate any law. Quite simply, that which is legal becomes ethically acceptable. They tend to approach this process as a reason for a reflection of what things are (Blanshard, 1984). Following this approach to maximizing revenue and following strictly legal approaches violates the ethics of reason. Blanshard (1984) posited that the use of reason in our ethical activities should be that "We must forecast the consequences of our proposed act in the way of intrinsic goods and bads in the experiences of those affected by it". Since many loan officers tend to function as sole proprietors within their respective firms, their individual ethical outlooks may affect their activities more than the ethical climates that exist within their respective firms.

As our respective financial institutions, including the entire mortgage industry, face the prospect of significant regulatory overall in 2010, analyzing and identifying the various ethical perspectives existing in mortgage brokers/bankers can assist regulators in establishing these new regulatory directions. Several activities recently, including vast numbers of mortgage fraud, have spurred this potential legislative effort. On the federal level, this effort has the potential to be the largest and most sweeping overhaul of the financial markets since passage of the Glass-Steagall act in 1935. Unethical and fraudulent activities are a significant part of the driver of this legislation (Vartanian, 2008).

STATEMENT OF THE PROBLEM

Housing is the single largest driver of the U.S. economy today. Virtually all U.S. economic activity depends on a vibrant housing sector. The list of associated businesses and trades that depend on this segment of our economy is almost endless. Practically all construction trades, appliance

manufacturers, landscapers, carpet manufacturers, city and village tax revenues, etc, etc depend on housing for their livelihood. In addition, that vibrant housing market is reliant on an active and compliant mortgage market. An absence of that strong mortgage market can create severe and significant overall economic problems.

Mortgage loans were originally made available to most U.S. homeowners through financial intermediaries known as savings and loan associations (S&Ls) (Woerheide, 1984). These financial institutions have been a part of the U.S. financial system since the mid 1800s, but they experienced rapid growth during the early part of the 1930s depression era because of stimulus from the U.S. Government (Woerheide, 1984). S&Ls accepted savings funds from individuals and then re-loaned those funds to others to purchase homes. The process of these transactions initially represented a social contract with community members as well as an investment for their savers. By providing individuals with a source of funds to purchase a home, these S&Ls invested in community expansion through funds from savers allowing those families to realize the American dream of homeownership. With the massive wave of financial deregulation that began in the late 1970s, these firms came under severe financial pressure and began to disappear. By the mid to late 1980s, virtually all S&Ls had disappeared from the landscape. This fragmentation created a mortgage loan process that today more closely resembles a generic product, where mortgage loans became strictly investment vehicles rather than any type of social contract with borrowers. These mortgage loans have evolved into very sophisticated investment vehicles that have opened new opportunities for many, but have also created some significant problems for our overall economy in the past two years (Mason, 2003).

Where the entire process of a mortgage loan application and its related activities was originally a complete vertical process within a given S&L, those activities have now become a completely horizontal process (Helm, 2008). All related functions completed in-house by S&Ls included completing the mortgage application, credit check and verification process, title search and title policy issue, appraisals and closing and disbursement actions. These functions today are performed by independent businesses that handle each step separate from the other. For example, the process of finding the customer (sales), completing the mortgage application and collecting related documents (processing) are now the purview of businesses we know as mortgage brokers or bankers. Independent title companies have replaced the property title search, title policy issue and closing and disbursement functions. Independent appraisal companies, available for hire by anyone, have replaced in-house appraisal departments. While the segmentation of these activities into separate businesses have created new entrepreneurial opportunities and enabled many more homeowners to obtain mortgage financing for their dream of homeownership, a multitude of problems attendant to the mortgage industry because of this segmentation have also arisen (de la Torre & McClatchey, 2006).

Most mortgage firms today are commonly called mortgage brokers or bankers. Their job is primarily sales in nature, as they obtain prospective borrowers, collect documentation from the prospective borrowers to verify information on the credit application, and then submit the loan package to a potential lender for underwriting and ultimately funding. These mortgage broker/bankers firms are referred to as mortgage companies, but they do not lend any of their own money. Laws governing mortgage lending, which vary from state to state, regulate mortgage brokers/bankers and in some

states, regulations are more onerous than in others. Mortgage brokers/bankers are also required to comply with a lengthy and often complex set of federal laws and regulations. In 2007, these mortgage brokers/bankers originated approximately 60% of all mortgage loans in the U.S. (NAMB, 2008).

These firms earn their compensation by charging fees to borrowers, more commonly known as points, which is generally a percentage of the loan amount. They may also charge for out-of-pocket expenses, such as credit, title and processing fees. These activities may then be further sub-contracted to processing companies and/or title insurance companies. Some brokers/bankers also charge a fee known as yield spread premium. Most employees of mortgage brokers/bankers are compensated on a commission only basis, similar to real estate agents. The sales agents within mortgage brokers/bankers are commonly known as loan officers or, in Illinois, loan originators.

Federal regulations require mortgage brokers/bankers to disclose their fees and charges to the borrower on a disclosure form known as the HUD-1. This form details how much the borrower is borrowing and to whom the various disbursements from the loan are paid. Some of those disbursements include the points and fees charged by the broker/banker. Yield spread premium, one significant potential source of broker/banker income, is also disclosed on the HUD-1, but in a very confusing and complex manner (Jackson & Burlingame, 2007).

Mortgage brokers/bankers charging yield spread premium may cause irreparable damage to the overall industry by increasing defaults and foreclosures due to higher interest rates charged to those borrowers. Ethical behavior within the mortgage industry, both individual and organizational, has suffered significant damage in the recent past (Harney, 2008). This ethical behavior is exacerbating the damage to mortgage brokers/bankers reputation in these turbulent times surrounding mortgage lending. This creates uncertainty in the mortgage industry further causing additional damage to the overall U.S. economy.

Public policy makers are in need of guidance as to what various regulatory changes are needed in light of the current damage caused to the economy by the subprime mortgage crisis. Mortgage fraud significantly affects all aspects of mortgage lending from origination through secondary market sales. This environment, where mortgage fraud pervaded many aspects of mortgage lending, may have also contributed to borrowers acquiring loans for which they ultimately where not qualified. This process has created significant financial impact on mortgage lending, and ultimately on the U.S. economy overall (U.S. Department of Justice, 2005). The overall financial impact of these white-collar type crimes has continued unabated since 2005, and there appears to be no end in sight (Christie, 2008). An analysis of mortgage fraud reveals that even though the overall mortgage market is shrinking, fraud cases have increased 42% in just the past year (Christie, 2008). While yield spread premium has not been directly linked to mortgage fraud, one recommendation by the FBI crimes report to the public indicates one method to help stem such fraud is for the borrower to understand the forms they are signing—that is, if they understand them at all. This revelation indicates serious problems surround the overall understanding of mortgage terms by the borrowing public at large. Inability to understand the complex mortgage terms through the required disclosure process has done little to solve these problems.

Public policy makers will work diligently to create new regulations as our financial markets have undergone the worst turmoil since the great depression. Knowledge of the difficulties and problems created by yield spread premium will be one of the issues foremost on their minds as new disclosures are created. Designing a newer, simpler disclosure form that will aid borrowers in truly understanding all costs in their mortgage transaction can significantly stabilize the mortgage market and assist in reestablishing mortgage lending to revitalize the U.S. housing market (Thordsen, 2008).

PURPOSE OF THE STUDY

Relationships between organizational climates and results of the behavior of employees have been shown to exist in previous research (Martin & Cullen, 2006). The five ethical climate dimensions defined by Victor and Cullen (1988) may promote behavior that causes harm to borrowers or to the success of their respective mortgage firms. The ethics of reason may also influence individual loan officer behavior (Blanshard, 1984). Motives that drive loan officer behavior, such as the desire to generate larger sums of income, may also affect their ethical behavior. This is an example of Kantian ethics in that use of other persons to fulfill loan officer needs abuses respect and dignity of others (Beauchamp & Bowie, 2004).

The purpose of this study was to determine what, if any, impact specific demographic behavior has on the ethical behavior of mortgage brokers/bankers in the state of Illinois. More specifically this study is an attempt to determine if behavior among mortgage brokers/bankers in Illinois is driven by the drive to earn higher compensation, and therefore encourage those who interact with the public in the extension of mortgage loans to act in an unethical manner.

SUMMARY AND CONCLUSION

Through three analysis methods undertaken in this study, ANOVA single question analysis, ANOVA combination question analysis, and regression analysis using ethical dimensions, it appears Illinois mortgage brokers/bankers exhibit ethical behavior that is affected by their desire to earn higher levels of compensation. Those who control the establishment of interest rates as well as closing costs exhibit different levels of ethical behavior than those with different compensation structures. Those who earn their compensation through commission exhibit strong levels of impact on their ethical behavior. It appears that these individuals are taking advantage of the confusion surrounding the current disclosure requirements such that they are using those laws to their advantage.

Specifically, four demographic questions appear to be the most significant drivers of ethical behavior. Those questions are, in the order of their importance—
1) 29 Compensation structure
2) 32 Borrower's overall interest rate established by
3) 44 Level of job satisfaction with current firm
4) 39 Intent to remain with current firm
From these indications, it is logical to presume those in a position to control borrower's interest rate and for those working on commission ethical behavior is significantly impacted. When the loan officer has the ability to charge the borrower more to increase their own compensation the impact to ethical behavior overall is three – five times more significant. In addition, interesting to note is that

overall level of job satisfaction and intent to remain with their current mortgage firm has a significant impact on overall ethical behavior.

Mortgage firms need to take actions to improve their standing with not only the borrowing public, but with various regulatory bodies as well. Illinois mortgage brokers/bankers have suffered significant setbacks due to high incidences of mortgage fraud in the past three years. Several cases of mortgage fraud have even risen to the issue of mortgage loans granted to deceased individuals. It is important that mortgage firms take the necessary steps to adjust their practices and procedures to repair the damage the mortgage business has suffered overall these past three years. One important way to do that is to either strictly enforce current ethics policies in the industry or reinforce ethics policies currently in place at those firms. Establishing accepted ethics codes of conduct for the industry as a whole, similar to those established by accountants or attorneys, would be a significant step in the right direction. Another important step is to adjust compensation processes such that there is not a direct link between the borrower's charges and the loan officer's compensation. Working to repair these two significant flaws in the mortgage broker/banker firms can do much to improve their standing with the general public and policymakers.

REFERENCES

[1] Anonymous. (2008). Interview with licensed mortgage professional regarding certain aspects of mortgage lending. Chicago, IL.

[2] Beauchamp, T., & Bowie, N. (2004). *Ethical Theory and Business (7th ed).* Upper Saddle River NJ: Pearson Prentice Hall.

[3] Blanshard, B. (1984). *Four reasonable men: Marcus Aurelius, John Stuart Mill, Ernest Renan, Henry Sidgwick.* Middletown, CT: Wesleyan University Press.

[4] Christie, L. (2008, 26 August 2008). Mortgage fraud still soaring. *CNNMMoney.com.* Retrieved 2 September 2008, from http://money.cnn.com/2008/08/25/real_estate/soaring_mortgage_fraud/index.htm.

[5] De la Torre, C., & McClatchey, C. (2006). PSST – hey buddy, want to buy a mortgage? A primer for financial planners. *Journal of Personal Finance, 5*(3), P21-42. Retrieved 10 Nov 2008, from Proquest Database.

[6] Helm, S. (2008). Interview with licensed mortgage professional regarding certain aspects of mortgage lending. Chicago, IL.

[7] IAMP. (2008). *Best practices handbook: A manual of ethical guidelines and professional standards for mortgage professionals and originators* (P1-8). Retrieved 20 November 2008, from http://www.iamp.biz/.

[8] Jackson, H., & Burlingame, L. (2007, Spring). Kickbacks or compensation: The case of yield spread premiums. *Stanford Journal of Law, Business & Finance, 12*(2), P289-362. Retrieved 10

November 2008, from Proquest Database.

[8] Lindsey, R. (1988, 2 March). Ancient redwoods fall to a wall street takeover. *New York Times.* Retrieved 3 May 2009, from http://www.nytimes.com/1988/03/02/us/ancient-redwoods-fall-to-a-wall-street-takeover.html?scp=1&sq=ancient%20redwood%20trees%20fall%20to%20a%20wall%20street%20takeover&st=cse.

[9] Martin, K., & Cullen, J. (2006, December). Continuities and extensions of ethical climate theory: A meta-analytic review. *Journal of Business Ethics, 69*(2), P175-194. Retrieved 25 November 2008, from Business Source Complete.

[10] Mason, D. (2003, 11 June). *Savings and loan industry (US)* (P1-11). Retrieved 22 August 2008, from http://eh.net/encyclopedia/article/mason.savings.loan.industry.us.

[11] NAMB. (2004). *National association of mortgage brokers code of ethics*. Retrieved 20 November 2008, from http://www.namb.org/namb/Default.asp.

[12] NAMB. (2008). *NAMB Bringing Homeownership to America's Communities*. Retrieved 14 Aug 2008, from National Association of Mortgage Brokers: http://www.namb.org/namb/Mission.asp?SnID=199072901.

[13] Schulman, P., & Spencer, H. (2008, June). RESPA reform: I all over again. *Real Estate Finance, 25*(1), P3-11. Retrieved 10 November 2008, from Business Source Complete.

[14] Thordsen, H. (2008, 2 May). The legal watch: Federal and state legislation proposed and/or effective in 2008. *Mortgage Originator*, P1-22. Retrieved 25 November 2008, from http://www.mortgageoriginator.com

[15] U.S. Department of Justice, Federal Bureau of Investigation. (2005, 20 May). *Financial crimes report to the public* (P1-58). Retrieved 24 November 2008, from http://www.fbi.gov/publications/financial/fcs_report052005/fcs_report052005.htm.

[16] Vartanian, T. (2008, August/September). Has the time for regulatory restructuring arrived? *Bank Accounting and Finance, 21*(5), P15-23. Retrieved 24 November, from Business Source Complete.

[17] Victor, B., & Cullen, J. (1988, March). The organizational bases of ethical work climates. *Administrative Science Quarterly, 33*(1), P101-125. Retrieved 8 November 2008, from Business Source Complete.

[18] Woerheide, W. J. (1984). *The savings and loan industry: Current problems and possible solutions.* Westport, CN: Quorum Books.

Investor Sentiment and IPO Performance:

The Case of Taiwan Electric and Non-electric

Chih-Cheng, Yeh,   04-23892088-3512,   Email：g9420802@yuntech.edu.tw
Chun-An, Li ,   05-5342601-2601,   Email：liica@yuntech.edu.tw
Wei-Zheng, Yan   0955- 912565,   Email：g9724711@yuntech.edu.tw
 Kun-Mei Pan   04-22196022,   Email：kmpan@ntit.edu.tw

ABSTRACT

Using the magnitude and duration of continuous rise to evaluate abnormal returns of IPOs unlike past literature, we find that the investor sentiment is strongly related to abnormal returns of IPOs in Taiwan. Further analysis indicates that, findings of the study about non-electronic IPOs are consistent with windows of opportunity and the bandwagon hypothesis. With these two hypotheses, both the magnitude and duration of long term abnormal returns of non-electronic IPOs slope downward. On the contrary, findings of the study about electronic IPOs are not consistent with these hypotheses. Both the magnitude and duration of short term and long term abnormal returns of electronic IPOs show a continuing upward trend.

Keyword: sentiment, IPO, abnormal returns,

INTRODUCTION

There have been many psychology scholars who generally argued that emotion has an important impact on decision making. Especially when people lack specific information necessary for making informed choices, their judgment tend to be influenced by emotions (Forgas,1995; Gendolla, 2000; Loewenstein et al., 2001). Market prices are driven by optimists especially when there are restrictions on short selling (Baker and Stein, 2004). In IPOs market always restrictions on short sales. Therefore, excess returns on IPOs may be related to investor sentiment. Underpricing at the beginning, hot IPO phenomenon and poor long-term performance are three circumstances being observed. Researchers are proposing hypotheses such as the winner's curse (Rock, 1986), the signal hypothesis (Firth, 1997), windows of opportunity hypothesis (Loughran and Ritter,1995) and the bandwagon hypothesis (Welch, 1992) as explanations of these phenomena. The study focuses on the relationship between the investor sentiment and IPOs performance in Taiwan to test and verify the windows of opportunity and the bandwagon hypotheses.

The windows of opportunity hypothesis posits that, when investors are optimistic about stocks during a period and will cause the large cycles of IPO volume. In other words, issuers will time their IPOs during these sentiment periods in order to take advantage of the opportune time to lower the cost of capital. IPOs issued during high sentiment period will have a higher probability to be overpriced than those issued during the other period. Hence, IPOs issued during high sentiment period will have a lower abnormal returns than those issued during the other period in the long-run. In Taiwan stock market, the flourishing electronics industry earns higher stock returns, causing its IPOs to attract greater attention from investors. Therefore, it was not necessarily for electronic firms in Taiwan to time their IPOs. On the contrary, non-electronic companies in Taiwan may have to time their IPOs in order to catch eyeballs. Such behavioral differences may cause long term abnormal returns of non-electronic IPOs be lower than those of electronic IPOs;

On the other hand, if potential investors pay attention not only to their own demand for a new IPO, but also to whether other investors are purchasing, bandwagon effects may develop. As IPOs of electronic companies catch more attention from investors than those of non-electronic firms, investor sentiment may start to play a role on abnormal returns of electronic IPOs from a point in time prior to listing. On the contrary, IPOs of non-electronic issuers get a lot less attention of investors. They may have to wait for those investors who are unable to participate in the electronic IPOs to redirect their investment demand, causing stock returns of non-electronic IPOs to be subject to bandwagon effects. For the reason, to non-electronic IPOs, the relevancy between investor sentiment and IPO abnormal returns before and after listing will be different from that of electronic IPOs in Taiwan.

Unlike past literature, the study evaluates abnormal returns of IPOs using the magnitude and duration of continuous rise. If Taiwan non-electronic IPOs are subject to windows of opportunity and bandwagon effects, then (1) the investor sentiment is strongly related to abnormal returns of IPOs ; (2) the long term abnormal returns of non-electronic IPOs will be lower than those of electronic IPOs; (3) the relevancy between non-electronic IPOs abnormal returns and investor sentiment before and after listing is different from that of electronic IPOs. Our empirical tests support all three propositions.

SAMPLING PROCEDURE AND METHODOLOGY

The study is limited to the data collected from January, 1997 to December, 2009. Long term IPO performance will not be available earlier than 36 months after the date of listing. In order to obtain sufficient data for evaluation, December of 2006 will be the deadline for sampling. The study has thus chosen companies that were first listed in Taiwan stock market during the period from January, 1997 to December, 2006. Finance companies and companies that are originally OTC listed or have been delisted or issued a temporary cease trade order are not included. Data are collected from Taiwan Economic Journal (TEJ).

In evaluating the sentiment variable, the principal component analysis filters the six proxies of Baker and Wurgler（2006）sentiment index and captures their common component - investor sentiment per month as follows:

$$SENTIMENT_t = 0.4142* CEFD_{t-1} + 0.6344* TURN_t - 0.2880* NIPO_{t-1} +$$

$$0.4368\ RIPO_{t-1} - 0.0104\ S_{t-1} - 0.3900* DP_t \qquad (1)$$

Abnormal returns of IPOs after listing are evaluated based on cumulative abnormal returns calculated in the manner proposed by Barber and Lyon (1997). The study also defines the magnitude of continuous rise and duration of continuous rise as follows:

1. Magnitude of continuous rise: As long as the *ith* IPO is still rising after listing and the extent of retreating is within 10%, then it is still regarded as a rising trend. The short run magnitude of continuous rise is the highest cumulative daily abnormal return over 30 days from the date of listing, and the long run magnitude of continuous rise is the highest cumulative monthly abnormal return over 36 months from the date of listing. It is a positive number. When the *ith* IPO is in a falling trend, the cumulative abnormal return of the day is the magnitude of continuous fall. It is a negative number.

2. Duration of continuous rise: As long as the *ith* IPO is still rising after listing and the extent of retreating is within 10%, then it is still regarded as a rising trend. The number of rising days prior to the highest cumulative daily abnormal return over 30 days from the date of listing is the short run duration of continuous rise. It is a positive number. When the *ith* IPO is in a falling trend, the number given is 0.

After the variables are defined, we first conduct statistical tests with standard error estimates in time series to assess short term and long term performance of IPO after listing and verify if there are cumulative abnormal returns respectively. Next, the paper calculates the magnitude and duration of continuous rise and observes whether electronic IPOs and non-electronic IPOs perform equally well to have the same long term and short term cumulative abnormal returns to verify if non-electronic IPOs are subject to windows of opportunity. Finally, the study examines the investor sentiment measured over the period of 1 month, 3 months, 6 months prior to listing and in the month of listing and conducts regression analysis to understand the link between investor sentiment and abnormal returns. Three control variables are used based on Fama and French (1998) three factor model. To verify if non-electronic IPOs in Taiwan are subject to bandwagon effects, we also examine if investor sentiment at these four different time points has the same impact on abnormal returns of both electronic and non-electronic IPOs.

$$EUPR_{i,t} = \alpha_o + \beta_o*SENT_{i,t} + \beta_1*SENT_{i,t-1} + \beta_2*SENT_{i,t-3} + \beta_3*SENT_{i,t-6} + \beta_4*RMRF_{i,t}$$
$$+\beta_5*SMB_{i,t} + \beta_6*HML_{i,t} + \varepsilon_{i,t} \qquad (2)$$

$$EUPT_{i,t} = \alpha_o + \beta_o*SENT_{i,t} + \beta_1*SENT_{i,t-1} + \beta_2*SENT_{i,t-3} + \beta_3*SENT_{i,t-6} + \beta_4*RMRF_{i,t}$$
$$+\beta_5*SMB_{i,t} + \beta_6*HML_{i,t} + \varepsilon_{i,t} \qquad (3)$$

Where $EUPR_{i,t}$ is the magnitude of continuous rise, $EUPT_{i,t}$ is duration of continuous rise, SENT is the investor sentiment that is divided into sentiment over the

period of 1 month, 3 months, 6 months prior to listing and in the month of listing, RMRF is the market risk premium, SMB is the size premium, and HML is the value premium.

EMPIRICAL RESULTS

Are there short term and long term abnormal returns to IPOs? There must be prior returns data for cross-section standard to avoid further biases with statistical tests of cumulative abnormal returns (Barber and Lyon, 1997). In the paper, standard error estimates in time series are used for statistically tests of cumulative abnormal returns. The results show statistically significant short term abnormal returns of 17.93% (t = 3.975) and positive long term abnormal returns of 47.25%(t = 4.626) for electronic IPOs, indicating that either short term cumulative abnormal returns or long term cumulative abnormal returns of electron IPOs are in a rising trend. For non-electronic IPOs, there are statistically significant short term abnormal returns of 2.51% (t = 4.1067), but long term cumulative abnormal returns takes a downward turn at the fifth month of listing. The cumulative abnormal returns become statistically insignificant from the 21st month of listing (t = 0.5088).

With regard to the magnitude and duration of short term continuous rise, the magnitude of continuous rise for electronic IPOs averages 29.85%, the length of time is 8 days. The magnitude of continuous rise for non-electronic IPOs averages 20.7%, the duration averages 5.5 days. The short term persistence of abnormal returns of electronic IPOs is longer than that of non-electronic IPOs either in the magnitude or duration and is statistically significant at the 1%.

With regard to the magnitude and duration of long term continuous rise, the magnitude of continuous rise for electronic IPOs averages 72.79%, the length of time is 12.6 months. The magnitude of continuous rise for non-electronic IPOs averages 54.45%, the duration averages 11.3 months. The magnitude of continuous fall for non-electronic IPOs averages 72.46%, the length of time is 18.1 months, indicating long term cumulative abnormal returns of electronic IPOs is in a rising trend, while that of non-electronic IPOs tend to take a downward turn (The magnitude of continuous fall is larger than that of continuous rise.)

Empirical evidence on the short-term abnormal returns of IPOs shows significant positive returns to both electronic and non-electronic firms. However, empirical results on the long term performance of IPOs reveal that electronic IPOs continue to enjoy a positive return, while non-electronic IPOs are having significant negative returns instead. The windows of opportunity hypothesis supposes that IPOs issued during good times will have a lower abnormal returns than those issued during the other times in long-run. Non-electronic issuers often time their IPOs, therefore, they have a higher probability of being overpriced than electronic IPOs to cause long term negative returns. The results are consistent with the windows of opportunity hypothesis.

The paper also examines if Taiwan stock market is subject to bandwagon effect. Empirical results reveal that:

(1) The magnitude and duration of continuous rise of short term and long term cumulative abnormal returns of electronic IPOs are positively related with investor sentiment over the period of 1 month, 3 months prior to listing and in the month of listing. They are significantly correlated to investor sentiment in the month of listing.

(2) The magnitude and duration of continuous rise of short term cumulative abnormal returns of non-electronic IPOs are positively related with investor sentiment in the month of listing and negatively related with investor sentiment over the period of 1 month and 3 months prior to listing with the sentiment over the period of 1 month having the greatest impact. The magnitude and duration of continuous fall of long term cumulative abnormal returns of non-electronic IPOs are positively related with investor sentiment over the period of 1 month and 3 months prior to listing and in the month of listing with the sentiment in the month of listing having the greatest influence.

Obviously, investor sentiment at different point of time has significantly different impact on short term abnormal returns of both electronic and non-electronic IPOs. Electronic IPOs start to attract investors' attention three months prior to listing, thus, the investor sentiment over the period of 3 months prior to listing is positively correlated with abnormal returns. Non-electronic IPOs do not start to attract investors' attention three months prior to listing, thus, the investor sentiment over the period of 3 months prior to listing is negatively correlated with abnormal returns. Not until the month of listing does the investor sentiment have a positive impact on abnormal returns.

Based on the findings of the study, electronic IPOs tend to attract more investor attention. The further analysis shows that investor sentiment over the period of 1 month and 3 months prior to listing and in the month of listing has an impact on abnormal returns of electronic IPOs with investor sentiment in the month of listing having the greatest influence. When seeing the boom and returns of electronic IPOs, investors who were unable to participate in electronic IPOs may redirect their investment demand to non-electronic IPOs, thus, abnormal returns of non-electronic IPOs are positively related to investor sentiment in the month of listing only. The redirection of investment demand evidences that non-electronic IPOs are subject to bandwagon effect. However, such short term demand has caused both the magnitude and duration of long term abnormal returns of non-electronic IPOs to slope downward.

CONCLUSIONS

In the IPO market, there is more severe information asymmetry due to lack of specific information necessary for making informed choices, the investor sentiment plays an important role in abnormal returns of IPOs. Using the magnitude and duration of continuous rise to evaluate abnormal returns of IPOs, we find the investor sentiment is strongly related to abnormal returns of IPOs. Besides, the purpose of the study is to examine if abnormal returns of non-electronic IPOs are significantly different from those of electronic IPOs under windows of opportunity and the bandwagon hypotheses. With these hypotheses, both the magnitude and duration of long term

abnormal returns of non-electronic IPOs slope downward. On the contrary, findings of the study about electronic IPOs are not consistent with these hypotheses. Both the magnitude and duration of short term and long term abnormal returns of electronic IPOs show a continuing upward trend.

REFERENCES

[1] Aggarwal, R.& Rivoli, P. *Fads in the Initial Public Offering Market?* Financial Management, 1990, 19(4), 5-57.

[2] Barber, B. & Lyon, J. *Detecting long-run abnormal stock returns: The empirical power and specification of test statistics*, Journal of Financial Economics, 1997, 43(3), 341–372.

[3] Baker, M. & Stein, J. C. *Market Liquidity as a Sentiment Indicator,* Journal of Financial Markets, 2004, 7 (3), 271-299.

[4] Baker, M. & Wurgler, J. *Investor Sentiment and the cross-section of stock returns*, Journal of Finance, 2006, 61(4), 1645-1680.

[5] Gendolla, G. *On the Impact of Mood on Behavior: An Integrative Theory and a Review,* Review of General Psychology, 2000, 4(1), 378-408.

[6] Firth, M. *An Analysis of the Stock Market Performance of New Issue in New Zealand*, Pacific-Basin Finance Journal, 1997, 5(1), 63-85.

[7] Fama, E. F. *Market efficiency, long-term returns, and behavioral finance*, Journal of Financial Economics , 1998, 49(3), 283-306.

[8] Forgas, J. P. *Mood and Judgment: The Affect Infusion Model (AIM)*, Psychological Bulletin, 1995, 117(1), 39-66.

[9] Loewenstein, G. F., Weber, E. U., Christopher, K.H. & Welch, N. *Risk as Feelings,* Psychological Bulletin, 2001, 127(2), 267-286.

[10] Loughran, T. & Ritter, J.R. *The New Issues Puzzle*, Journal of Finance, 1995, 50(1), 23-51.

[11] Rock, K. *Why New Issue are Underpriced*, *Journal of Financial Economice,* 1986, 15,187-212.

[12] Welch, I. *Sequential sales, Learning and Cascades*, Journal of Finance , 1992, 47, 695-723.

RETIRMENT PLAN SIMULATION

James Bishop, Bryant University, 1150 Douglas Pike, Smithfield RI 02917,
jabishop@bryant.edu, 401-232-6356
Phyllis Schumacher, Bryant University, 1150 Douglas Pike, Smithfield RI 02917,
pschumac@bryant.edu, 401-232-6328
Alan Olinsky, Bryant University, 1150 Douglas Pike, Smithfield RI 02917,
aolinsky@bryant.edu, 401-232-6266

ABSTRACT

The focus of this paper is the effect of changes to employer sponsored retirement plans on employee retirement benefits. Today's retirement benefits consist mainly of three types of plans: defined benefit (DB), defined contribution (DC), and "hybrid" plans. Many employers have changed the type of plan they offer in recent years. Specifically, there is a shift from DB plans to either DC or hybrid plans (hybrid plans have features of both DB and DC plans). In this paper we focus on a comparison of the retirement benefits between a DB plan and a DC plan under different scenarios which depend on years of work, market yield and predicted wage increases. We use a simulation model to compare DB and DC benefits over different career lengths for a worker with a starting salary of $50,000. Simulated fluctuations in annual market yield and average national wage increases are used to project DC balances. DB benefits are simulated using random fluctuations in wage increases. The resulting simulated benefits show that DC plans are generally inferior to DB plans.

Keywords: retirement, benefit projection, retirement simulation, defined contribution, defined benefit

INTRODUCTION\BACKGROUND

During the past thirty years, employers have transitioned from offering traditional defined benefit (DB) retirement plans to defined contribution (DC) plans. The drawback of contributory plans is that the money is typically invested in mutual funds and is subject to market volatility [2]. In addition, the money available to the individual at retirement under a contributory plan is a lump sum. This requires the retiree to then be responsible for turning that money into income during their retirement years, typically purchasing a life annuity with the funds. In contrast, defined benefit plans are not subject to market volatility and are paid out as life annuities.

Another common alternative to DB and DC plans in recent times is to combine the concept of the defined benefit and defined contribution plans into a hybrid plan – these plans have a variety of names such as: "cash balance" [7], "pension equity" [3], or "stable value". These hybrid plans accumulate an employee account balance in a similar fashion to the defined contribution plan,

but the income earns fixed interest and is not invested in the stock market. Typically, hybrid plans are less valuable than either DB or DC plans.

Employees across the country are facing changes to their retirement benefits. When such changes occur, employers typically send benefit statements to the employees, projecting how the change may affect future benefits. These projections are simplistic and assume constant yield, projected wage increases, and interest assumptions. Therefore, it is difficult for employees to understand the future impact of changes to their retirement plans.

Employers sponsoring a DB plan, whether changing the plan provisions or not, are required to send statements to their employees that typically include a benefit projected to age 65. In the case of a DC plan, the employer may send a periodic statement showing a projected benefit, otherwise, it is up to the employee to plan for their own retirement by doing a projection. For both DB and DC plans, these projections make constant growth assumptions for wage increases and market yield. In addition, the conversion rates between lump sums and annuities for future benefits are unknown, making a comparison between DB and DC benefits difficult.

For example, suppose the employee is 40 or 50 years old and their statement shows a projected lump sum benefit payable at age 65 of $553,462.00.   Presenting a benefit 15 or 25 years in the future is not realistic. In reality, the amount that the employee is likely to receive at age 65 is a range of values depending on future increases. Rather than using the typical constant increase method for projecting benefits for both DC and DB plans [1], simulation techniques allow us to project benefits in a more realistic manner. In addition, it is beneficial to see the variation that occurs over a long period of time.

The simulations presented within utilize actual national average wage increases over the past 59 years (as used by the social security administration), and market yield data over the same 59 year period. Simulated benefits are also computed over different length careers. This produces a more realistic projection, and includes the advantage of displaying a range of possible benefit values.

SIMULATION METHODOLOGY

Defined contribution account balances are projected based on simulated stock market performance using the S&P as the benchmark. The amount contributed to the account each year is a percentage of the employee's pay and so a wage projection is also necessary in order to determine the amount allocated to the account annually. In all cases, we have assumed a starting salary of $50,000.

An Excel-based simulation software package, @RISK, was utilized to test the various retirement scenarios. For each projected service period, i.e., 25 years, 35 years, and 45 years, 10,000 simulated projections were performed. A discrete uniform distribution was used to randomly generate wage increases and S&P increases based on equally likely probabilities for all annual increase percentages from years 1950 – 2009.

In order to minimize the variation in results for the comparisons between DB and DC benefits, the random numbers used in both formulas were the same.

Defined benefits only require a wage increase simulation. These benefits do not depend on the market and only depend on a benefit formula based on final salary and length of service. This benefit is expressed as a life annuity, therefore, conversion to a single lump sum benefit is required when the comparison to a DC balance is made. The conversion rate used within is the January, 2011 PPA projected spot rate required for conversion of annuities to lump sum benefits. The current conversion factor is 12.265 (i.e. the lump sum is worth 12.265 times an annual life annuity). However, the future conversion rate could be anywhere from approximately 10.0 to 14.0 for benefit conversions being performed in the future.

Both DC and DB retirement plans are simulated based on both a better than average plan formula, average plan formula, and worse than average plan formula. The following assumptions and simulation parameters are summarized below:

- Current salary of $50,000
- Discrete random selection of national average wage increases from 1950 – 2009
- Discrete random selection of S&P annual increases from 1950 - 2009
- Benefit accrual begins on the date of hire (also called service start age)
- Service start ages of 20, 30, and 40
- DC contributions are made mid-year
- Semi-annual interest compounding
- Actuarial conversion factors for various forms and timing of benefit payments are based on current Pension Protection Act assumptions (2010 mortality and yield rates for January 1, 2011) [4].

RESULTS

The simulation results presented within provide a range of likely retirement benefit amounts based on the level of benefit the employer offers. All benefit projections are simulated lump sum values payable at age 65 for both DB and DC retirement plans.

First, consider a typical projected defined contribution plan benefit assuming a 5% of pay employer contribution for an employee age 40, retiring at age 65. The 5% contribution assumption is based on an approximate average large company DC plan (range is roughly 2% - 8%). The resulting range of simulated lump sum retirement values is displayed in Figure 1a.
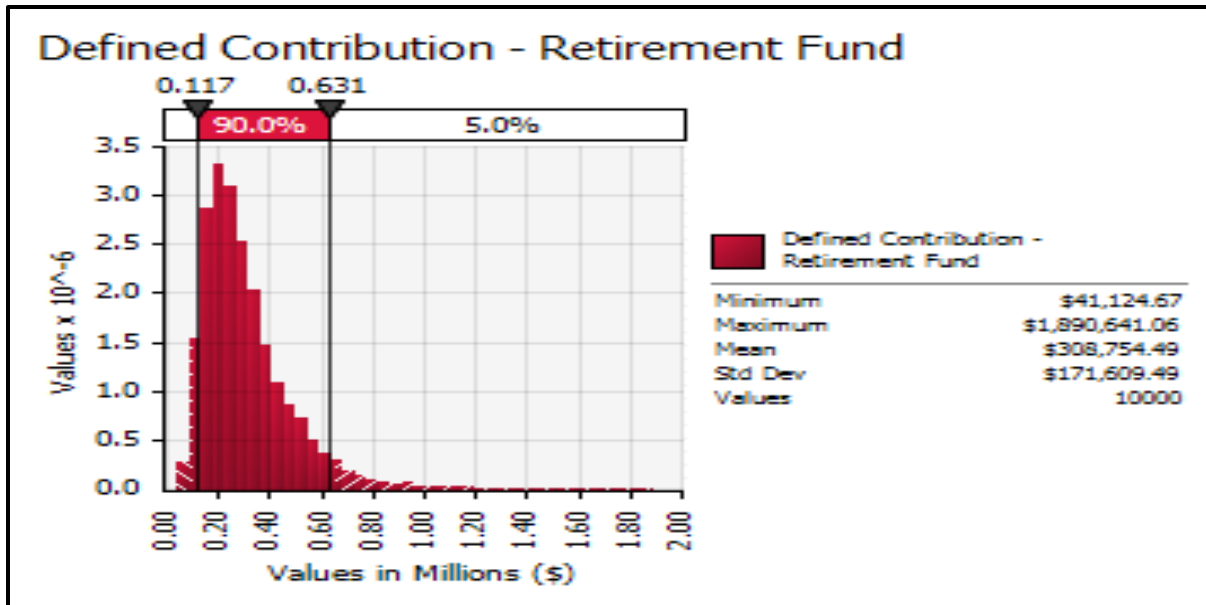
Figure 1a

In contrast, a typical DB benefit formula [4], also assuming a 40-year old employee projected to retire at age 65, computed using a typical 1.5% multiplier [1.5% times (years of service) times (final five average annual salaries) times (lump sum conversion factor)] is simulated in Figure 1b.
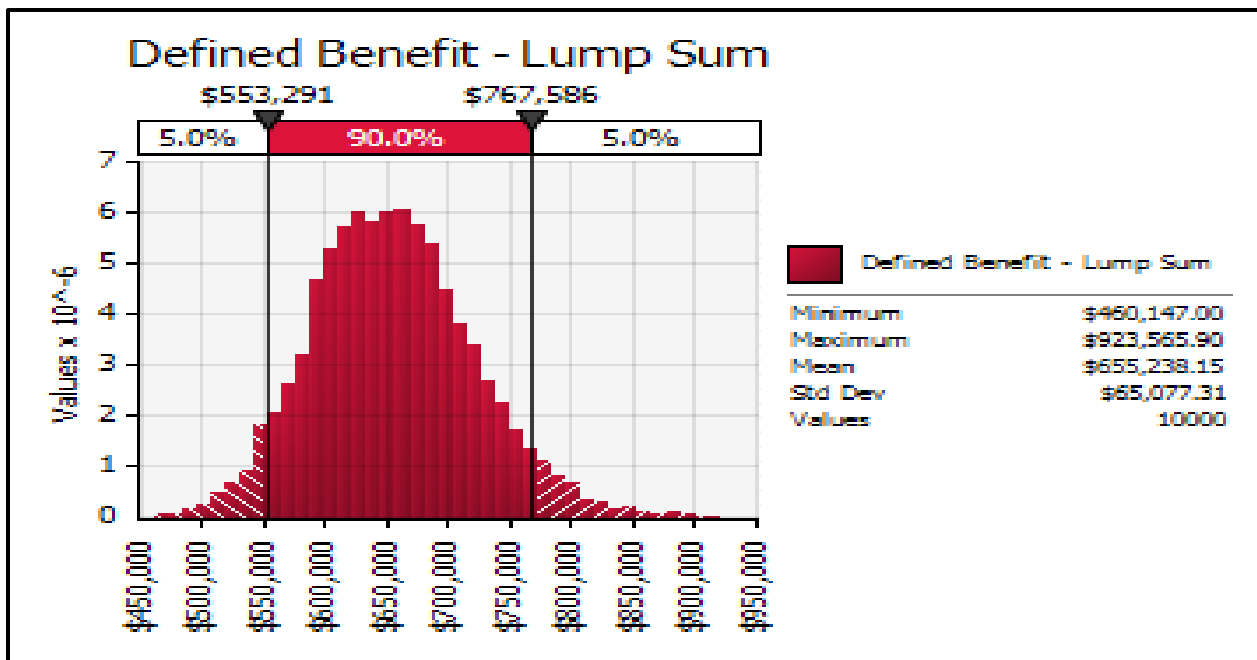
Figure 1b

For this example, the DB formula produces a significantly larger benefit value than the corresponding DC formula. However, there is a much larger deviation in the value of the DC benefit and a long upper-end tail to the simulation distribution.

The tables below summarize the simulations for employee's entering their retirement plans at ages 40, 30, and 20 respectively. All values assume retirement at age 65.

Contribution Rate/Fixed Benefit Rate	Defined Benefit			Defined Contribution		
	Mean	Median	Std. Dev.	Mean	Median	Std. Dev.
8%/2%	$ 873,651	$ 869,945	$ 86,770	$ 494,007	$ 429,638	$ 274,575
5%/1.5%	$655,238	$652,384	$ 65,078	$ 380,754	$ 268,524	$ 171,609
2%/1%	$ 436,825	$ 434,923	$ 43,385	$ 123,502	$ 107,409	$ 68,644

Table 1: 25 Years - Start Contributing Age 40 – Retire Age 65

Contribution Rate/Fixed Benefit Rate	Defined Benefit			Defined Contribution		
	Mean	Median	Std. Dev.	Mean	Median	Std. Dev.
8%/2%	$ 1,944,841	$ 1,931,843	$ 232,519	$1,341,659	$1,090,402	$ 948,552
5%/1.5%	$ 1,458,631	$ 1,448,882	$ 174,389	$ 838,537	$ 681,501	$ 592,845
2%/1%	$ 972,421	$ 965,921	$ 116,256	$ 333,563	$ 273,764	$ 231,596

Table 2: 35 Years - Start Contributing Age 30 – Retire Age 65

Contribution Rate/Fixed Benefit Rate	Defined Benefit			Defined Contribution		
	Mean	Median	Std. Dev.	Mean	Median	Std. Dev.
8%/2%	$3,976,011	$3,933,550	$ 545,029	$3,401,857	$2,561,304	$2,971,182
5%/1.5%	$ 2,982,008	$ 2,950,163	$ 408,772	$2,126,161	$1,600,815	$1,856,989
2%/1%	$ 1,988,005	$ 1,966,775	$ 272,514	$ 850,464	$ 640,326	$ 742,796

Table 3: 45 Years - Start Contributing Age 20 – Retire Age 65

Table 1 gives us the ability to compare benefit projections for a shorter career. The DC plan at 8% is comparable to a DB plan with a 1% formula, particularly if we look at the median of the projected values. An 8% contribution rate is considered better than average for a DC plan and

yet a 1% multiplier is (or was) considered less than average for a DB plan. Therefore, DC plans in general are not as generous as DB plans.

Over a longer career (Table 2 and Table 3), the DC plan fares better and the 8% DC benefit gets closer to being comparable to the 1.5% DB plan. The DB multiplier of 1.5% , considered an average DB multiplier, would require more than 8% contributions from a DC plan to provide the same retirement income.

CONCLUSION/DISCUSSION

The simulation of projected benefits clarifies the relationship between DC and DB lump sum retirement values. DB plans are generally more generous than their DC plan counterparts. In addition, DC plan benefits are much more volatile due to changes in the stock market.

The recent existence of hybrid plans is another interesting topic for discussion. These plans are usually not dependent on the stock market but also have a much lower average return rate on contributions. The overall lump sum value of a hybrid plan is generally comparable or less than that of a DC plan.

The one variable that creates uncertainty in the analysis presented within is the conversion factor (multiplier) that converts the DB annual annuity to a lump sum benefit. This factor can be significantly higher or lower based on interest rates at the time of payout (based on bond yield rates). Also, a lump sum payment is often not available under a DB plan.

REFERENCES

[1]   Bishop, J., Schumacher, P., Heeder, K. (2010). *Comparison of Retirement Benefit Plans in a Changing Economy.* Northeast Decision Sciences Institute Conference.

[2]   Coombes, A. (2009, October 6). *Steady savers gained over 5-year period*. Retrieved from http://www.marketwatch.com/story/401ks-took-big-but-not-devastating-hit-in-2008-2009-10-06.

[3]   Green, L. (2003, October 29). *What is a pension equity plan?*. Retrieved from http://www.bls.gov/opub/cwc/cm20031016ar01p1.htm

[4]   Hamilton, B, & Burns, B. (2001, December 31). *Reinventing retirement income in America.* Retrieved from http://www.ncpa.org/pub/st248?pg=5

[5]   IRS. (2009, September 5). *Choosing a retirement plan: defined benefit plan*. Retrieved from http://www.irs.gov/retirement/article/0,,id=108950,00.html

[6]   Kamenir, J. (2009). *How to Make Defined Benefit Pension Plans Attractive to 21st Century* Employers. Benefits Quarterly, 25(2), 51-56. http://search.ebscohost.com

[7]   Lowman, T. (2000, July 7). *Actuarial aspects of cash balance plans*. Retrieved from http://www.soa.org/files/pdf/actuarial_aspects.pdf

[8]   McGill, D, Brown, K, Haley, J, & Schieber, S. (2005). *Fundamentals of private pensions*. Oxford, NY: Oxford University Press.

[9]   Palisade Corporation. @Risk 5.7, [simulation software package]. New York, 2010

MUTUAL FUND GOVERNANCE, RETURNS, EXPENSES, AND CASH FLOWS

D.K. Malhotra, Philadelphia University, (215) 951-2813, MalhotraD@philau.edu
Maria Jaramillo, Philadelphia University, (215) 951-2950, jaramillo8365@philau.edu
Rand Martin, Bloomsburg University, rmartin@bloomu.edu

ABSTRACT

In this paper, we investigate the relation between mutual fund governance and fund's cash flows, expense ratios and returns. We use an overall fund governance grade and components of governance such as board quality, managerial incentives, corporate culture, and regulatory compliance to evaluate the impact of governance on fund performance. We find that better governed funds have higher returns and lower expense ratios. Impact of governance on mutual fund cash flow growth, however, is not very clear.

INTRODUCTION

Mutual funds have grown in popularity as the preferred investment vehicle. The number of mutual funds has increased to 23,148 at the end of 2006 with assets under management of $10.414 trillion.[1] In recent years, due to conflicts of interest in mutual funds between managers, fund sponsors and shareholders, the issue of mutual fund governance has attracted much popular, academic, political and even legal attention. Therefore, it is imperative to analyze the impact of governance on mutual fund's performance, cash flows and expense ratios.

In this paper, we investigate empirically if effective record of governance is noticed by the fund investors and results in increased cash flows for the fund. Secondly, we evaluate if effective governance results in improved rate of return for the investors. The third question that we investigate in this study is if effective governance results in reduced expense ratios for the funds.
 The study is important for several reasons. Firstly, if effective governance results in superior rate of return for fund investors, then the fund investors should notice it and cash flows to funds with effective governance records should go up. This will be a signal to funds with poor governance record to improve their fund governance record in order to attract cash flows from potential investors. If a fund with a record of effective governance does not attract cash flows from potential investors, it shows the presence of information asymmetry, because potential fund investors may not be aware of funds with a record of effective governance. Secondly, previous studies argue that expenses have a direct impact on fund's return to investors.[2] If effective governance results in reduced fund expenses, then fund shareholders should look towards funds with a record of good governance, with good governance, fund expenses will go down and return to investors will increase.

Thirdly, recent studies show that individual investors are paying attention to fund expenses and net fund flows are influenced by fund costs.[3] We reexamine this issue, because if funds with better governance enjoy lower expenses, they will also attract higher cash flows.

---

[1] Investment Company Institute; http://www.ici.org/stats/mf/trends_10_05.html#TopOfPage
[2] Chance and Ferris (1991) and Malhotra and McLeod (1997)
[3] See Ivković (2002), Sirri and Tufano (1998), Siggelkow (1998, 2003), and Nanda, Wang, and Zheng (2000), Barber, Odean, and Zheng (2004), and Khorana and Servaes (2004)

Our study will help the fund industry as well as regulatory agencies such as the Securities and Exchange Commission (SEC) get a better understanding of the impact of effective fund governance on fund's expenses and its benefits to fund shareholders in the form of higher returns and increased shareholder wealth.

This paper has six sections. In section II, we summarize previous studies on mutual fund governance. Section III briefly describes the data used in this study. Section IV describes the methodology and model used in this study. Section V presents empirical results. We summarize and conclude our study in Section VI of the paper.

LITERATURE REVIEW

A number of studies exist on corporate governance and most studies of corporate governance focus on industrial organizations. Studies on corporate governance focus on the impact of effective governance on shareholder wealth. Studies that find relationship between performance and better governance measure governance through the presence of more independent directors on the corporate board [Weisbach (1988), Byrd and Hickman (1992), Cotter, Shivdasani and Zenner (1997), and Brickley, Coles and Terry (1994)]. On the other hand, studies by Baysinger and Butler (1985), Hermalin and Weisbach (1991), and Klein (1998) find no evidence about the impact of board composition on firm performance and shareholder value.

In the context of mutual funds, earliest work can be traced back to Tufano and Sevick (1997) study that highlights relationship between board characteristics and fund fees. They report that the fees charged by open-end funds are lower when the funds' boards have characteristics that are consistent with effective governance. Relationship between fund governance fund expenses has also been explored by Del Guercio, Dann and Partch (2003) and they report that effective fund governance as evidenced by board structures consistent with effective board independence are associated with lower expense ratios. Cremers, Driessen, Maenhout, and Weinbaum (2005) report that governance does play a role in fund performance and funds with higher director ownership perform better.

Wellman and Zhou (2007) study reports that good governance results in superior performance of mutual funds. They also find evidence of cash outflows for funds with poor governance ratings. In a recent paper, Kong and Tang (2008) find that unitary boards are a better mechanism for mutual fund governance, because they have lower expenses and rank better on stewardship.

DATA

Mutual fund governance ratings are a logical response to the recent mutual fund scandals. In August 2004, just before mutual funds had to disclose information on the proxy voting records and regulations, the fund-rating firm *Morningstar* launched its fiduciary grade system of governance rating. Our data is from *Morningstar Principia*. The data set covers the years 2004, 2005 and 2006. Annual data for each fund includes *Morningstar* ratings ranging from excellent to very poor with accompanying commentary for each fund in five governance areas: board quality, manager incentives, regulatory issues, and corporate culture (which are rated by Morningstar on five scales: Excellent, Good, Fair, Poor, and Very Poor).

Board quality is the quality of a fund's board as evaluated by *Morningstar* on the basis of the following factors:

- Has the board taken action in cases where the fund clearly hasn't served investors well?
- Do the independent directors have meaningful investments in the fund? Morningstar allocates maximum score to a fund if at least 75% of a board's independent directors must have more money invested in the funds they oversee than they receive in aggregate annual compensation for serving on the board.
- Is the board overseeing so many funds that its ability to diligently protect the interests of the shareholders at this specific fund could be compromised?
- Does the fund meet the maximum SEC requirement for the proportion of independent directors, regardless of whether or not it is subject to the requirement?

Corporate culture looks at a wide range of factors in an attempt to assess how seriously a firm takes its fiduciary duty to fund holders. *Morningstar* consider the following factors:

- Has the firm launched "trendy" funds in attempt to gather assets?
- Has the firm closed funds at an appropriate size, or has it allowed fund assets to grow too large?
- Does the firm implement redemption fees or otherwise discourage rapid trading of its funds?
- Has the firm done a good job of retaining key personnel?
- How strong are the firm's shareholder communications?
- Does the firm direct fund brokerage in exchange for shelf space or use soft dollars?

To evaluate managerial incentives, *Morningstar* assesses two distinct components:
Fund ownership: Does the manager have significant investment in the funds he or she oversees, defined as 1/3 of his or her liquid net worth? If the funds run by the manager are inappropriate for such a large investment, does he or she have at least 1/3 of his or her liquid net worth in other funds at the same firm? *Morningstar* prefers compensation plans that reward long-term performance and does not emphasize asset growth. Incentive programs that encourage a focus on short-term performance or asset growth receive are viewed less favorably under *Morningstar's* rating structure.

Morningstar examines any regulatory issues at the fund company in the past three years. In the event of any breaches, *Morningstar* examines the remedies in place and the scope of and commitment to reform.

On the basis of a fund's regulatory compliance, board quality, manager incentives, fees, and corporate culture, *Morningstar* assigns an overall governance (stewardship) grade ranging from A (excellent) to F (worst). According to *Morningstar,* the governance grade allows investors and advisors to evaluate funds as to the manner in which funds are run, the extent to which the management company's and fund board's interests are aligned with fund shareholders, and the degree to which shareholders can expect their interests to be protected from potentially conflicting interests of the management company.

In addition to the above governance variables, we have information on investment objective, age of the fund (in years), assets (investment dollars) under management, number of brokerages through which a fund is being marketed and distributed, front-end and deferred load (percent), turnover ratio (percent), three-year annualized standard deviation and 12b-1 plans cost (percent).

Furthermore, we use the information given in the *Morningstar Principia* to compute number of funds in a fund family and the degree of concentration (fund focus) within the family. Fund focus or

concentration is defined as the extent to which a fund family is focused on an investment objective within the family.

Equation 1 shows the method that we use to compute fund focus within a fund family.

$$Fund\ Focus = \frac{Number\ of\ Funds\ in\ Family - Number\ of\ Fund\ Objectives\ in\ Family}{Number\ of\ Funds\ in\ Family} \quad (1)$$

The fund focus variable varies between zero for the lowest fund focus and less than 1.00 for the highest fund focus. If a large fund family concentrates on offering a limited number of fund objectives within the family, fund focus/concentration will be higher.

METHODOLOGY

In order to investigate the impact of effective governance in mutual funds, we use a two-part methodology. In the first part, we develop three models in which we study the impact of fund governance on a fund's total return, fund's cash flows, and fund's expense ratios. We model total return of a cross-section of mutual funds as a function of mutual fund governance in two ways. Our first model defines total return as a function of fund governance in which governance is indicated by board quality, manager incentives, regulatory issues, and corporate culture. In a variation of this model, instead of using board quality, manager incentives, regulatory issues, and corporate culture as measures of governance, we use one overall governance grade as an indicator of governance.

In addition, we introduce several control variables in the form of investment objective (stock or bond fund), age of the fund (in years), assets (investment dollars) under management, number of brokerages through which a fund is being marketed and distributed, front-end and deferred load (percent), turnover ratio (percent), three-year annualized standard deviation, and 12b-1 plans cost (percent). Equations 2a and 2b provide our models to study the impact of fund governance on fund's total return.

TR = a + b_1 (BOARD QUALITY) + b_2 (CORPORATE CULTURE) + b_3 (MANAGERIAL INCENTIVES) + b_4 (REGULATORY ISSUES) + b_5 (ASSETS) + b_6 (AGE) + b_7 (CASH FLOWS) + b_8 (DL) + b_9 (EXPENSE RATIO) + b_{10} (FL) + b_{11} (FUND FAMILY) + b_{12} (FUND FOCUS) + b_{13} (HISTORICAL EARNINGS GROWTH) + b_{14} (INSTITUTIONAL) + b_{15} (STOCK) + b_{16} (SD) + b_{17} (TURNOVER RATIO) + b_{18} (12B-1 PLAN) + e
$\qquad$ (2a)

TR = a + b_1 (EXCELLENT/GOOD GOVERNANCE) + b_2 (ASSETS) + b_3 (AGE) + b_4 (CASH FLOWS) + b_5 (DL) + b_6 (EXPENSE RATIO) + b_7 (FL) + b_8 (FUND FAMILY) + b_9 (FUND FOCUS) + b_{10} (HISTORICAL EARNINGS GROWTH) + b_{11} (INSTITUTIONAL) + b_{12} (STOCK) + b_{13} (SD) + b_{14} (TURNOVER RATIO) + b_{15} (12B-1 PLAN) + e
$\qquad$ (2b)

 Where:
$\qquad$ TR is defined as total return of a mutual fund for the past 12 months
$\qquad$ BOARD QUALITY is a dummy variable that equals 1 if the rating is excellent or good and 0 otherwise
$\qquad$ CORPORATE CULTURE is a dummy variable that equals 1 if the rating is excellent or good and 0 otherwise

MANAGERIAL INCENTIVES is a dummy variable that equals 1 if the rating is excellent or good and 0 otherwise

REGULATORY ISSUES is a dummy variable that equals 1 if the rating is excellent or good and 0 otherwise

EXCELLENT/GOOD GOVERNANCE GRADE is a dummy variable that equals 1 if the rating is A or B and 0 otherwise

Control Variables:

Assets: Net assets serve as a proxy for the size of the fund. As more and more investors invest in a fund, fund manager is presented with more cash than he/she can possibly invest. Fund manager may invest in instruments that are not optimal from the investor's viewpoint, thereby reducing fund returns as fund size rises. Studies by Grinblatt and Titman (1989), Gorman (1991), Ang, Chen, and Lin (1999) report that fund size is negatively related to fund return.

Age: Age may impact mutual fund returns, because there may be a greater pressure on new fund to perform in order to attract capita from fund investors in a highly competitive market. Redman and Gullet (2007) report that Fund Age is statistically significant in explaining fund returns. The direction of the impact of fund age changed in the two time periods where older funds had lower risk-adjusted returns in 1997–2000 and higher returns in 2001–03. Adkisson and Fraser (2003) argue that young *funds* tend to be smaller than older *funds*, which makes the young *funds' returns* and ratings more susceptible to manipulation. Also, Wisen (2002) shows that usually newer funds report higher returns due to bias in reporting of results by new funds.

Cash flows: Historical cash flow growth rate for a fund to analyze if a fund with higher total return has been attracting higher cash flows.

DL: DL refers to deferred load as a percentage.

Expense Ratio: Percentage of fund assets paid for operating expenses and management fees

FL refers to front-end as a percentage

The above three variables have been included as control variables, because previous research provides evidence on the relationship between fund performance and fund loads and fund expenses. According to Redman and Gullet (2007) that higher expense funds show lower returns. Dowen and Mann (2004) also report that expense ratios are negatively related to fund returns.

FUND FAMILY: Fund family refers to the number of funds being offered by a fund family. Funds that belong to large fund families may enjoy economies of scale in security research costs due to cost sharing with other members of the family. Also, larger fund families may be able to attract better fund managers who have better security selection skills, which translate into better returns for investors. Guedj and Papastaikodi (2004) show that funds that belong to larger families show more persistent performance than the entire universe of funds. Elton, Gruber, and Green (2007) report evidence that mutual fund returns are more closely correlated within than between fund families.

FUND FOCUS: We include fund focus as one of the variables that will impact fund returns. Siggelkow (2003) reports that focused funds perform better in contrast to funds that belong to more diversified fund families. Thus the degree of focus will be higher for a smaller number of fund objectives offered within a fund family. We hypothesize a positive relation between fund focus and fund's total return

HISTORICAL EARNINGS GROWTH: Growth of earnings per share over the past five years. We include this variable to evaluate if historical earnings growth impacts the returns during the past 12 months.

INSTITUTIONAL: It is a dummy variable that is assigned a value of 1 if the fund belongs to an institutional share class, 0 otherwise

STOCK: It is a dummy variable that is assigned a value of 1 if the fund belongs to an equity fund class, 0 otherwise

SD: Three-year annualized standard deviation

TURNOVER RATIO: Percentage turnover ratio for the fund. According to Israelsen (1998), in a study of 1,365 equity funds conducted by Morningstar, a low turnover ratio corresponded with higher returns and lower expenses.

12b-1 Plan: Percentage charges for 12b-1 plans. Philpot (1994) reports that funds with 12b-1 plan show lower return in comparison to funds without 12b-1 plan.

In the second model, we evaluate the impact of effective governance on funds' expenses. In equation 3a, we model fund's expense ratios as a function of board quality, corporate culture, manager incentives, and regulatory issues. In addition, we introduce several control variables in the form of age of the fund (in years), assets (investment dollars) under management, number of brokerages through which a fund is being marketed and distributed, deferred and front-end load (percent), fund family (number of funds in the family), fund focus (concentration), Institutional or retail (type of share class), stock or bond fund, three-year annualized standard deviation, percentage turnover ratio, and 12b-1 plans cost (percent). Equation 3b provides a variation of the 3a model by measuring governance through an overall governance grade (instead of measuring governance through board quality, corporate culture, managerial incentives, and regulatory issues separately).

$E = a + b_1$ (BOARD QUALITY) $+ b_2$ (CORPORATE CULTURE) $+ b_3$ (MANAGERIAL INCENTIVES) $+ b_4$ (REGULATORY ISSUES) $+ b_5$ (ASSETS) $+ b_6$ (AGE) $+ b_7$ (BROKERAGE AVAILABILITY) $+ b_8$ (CASH FLOWS) $+ b_9$ (DL) $+ b_{10}$ (FL) $+ b_{11}$ (FUND FAMILY) $+ b_{12}$ (FUND FOCUS) $+ b_{13}$ (INSTITUTIONAL) $+ b_{14}$ (STOCK) $+ b_{15}$ (SD) $+ b_{16}$ (TOTAL RETURN) $+ b_{17}$ (TURNOVER RATIO) $+ b_{18}$ (12B-1 PLAN) $+ e$

(3a)

$E = a + b_1$ (EXCELLENT/GOOD GOVERNANCE) $+ b_2$ (ASSETS) $+ b_3$ (AGE) $+ b_4$ (BROKERAGE AVAILABILITY) $+ b_5$ (CASH FLOWS) $+ b_6$ (DL) $+ b_7$ (FL) $+ b_8$ (FUND FAMILY) $+ b_9$ (FUND FOCUS) $+ b_{10}$ (INSTITUTIONAL) $+ b_{11}$ (STOCK) $+ b_{12}$ (SD) $+ b_{13}$ (TOTAL RETURN) $+ b_{14}$ (TURNOVER RATIO) $+ b_{15}$ (12B-1 PLAN) $+ e$

(3b)

Where *E:* the expense ratio of the fund

BROKERAGE AVAILABILITY: We include a variable that captures the number of brokerages through which a mutual fund is being marketed. Many mutual funds are available through several brokerage houses such as Charles Schwab or Fidelity. For instance, the Elite Income fund shows E3/A0/A2 listed for Brokerage Availability. Therefore, this fund may be purchased through Accutrade, Charles Schwab, and Charles Schwab-OneSource. Availability through big brokerage houses may allow individual investors the benefits of "one-stop" shopping as well as switching from one fund to another without charge. Some mutual funds are available through as many as 65 different brokerage outlets. A larger number of distribution channels are supposed to bring in new shareholders. This should increase the size of the fund and reduce expenses for the shareholders. However, a mutual fund has to pay to make it available through different brokerages. Also, a given brokerage will deal with a diverse group of funds, not limiting themselves to just a single fund family. We expect a positive coefficient for this variable.

All other variables are as defined above in equation 2.

In the third model, we investigate the impact of fund governance on the cash flows of the fund. We model fund's cash flows in equation 4a as a function of board quality, corporate culture, manager incentives, and regulatory issues. In addition, we introduce several control variables in the form of age of the fund (in years), assets (investment dollars) under management, number of brokerages through which a fund is being marketed and distributed, deferred and front-end load (percent), fund family (number of funds in the family), fund focus (concentration), Institutional or retail (type of share class), stock or bond fund, three-year annualized standard deviation, percentage turnover ratio, and 12b-1 plans cost (percent). Equation 4b summarizes the equation 4a model by dropping board quality, corporate culture, managerial incentives, and regulatory issues and including one overall GOVERNANCE grade as a measure of corporate governance.

$CF = a + b_1$ (BOARD QUALITY) $+ b_2$ (CORPORATE CULTURE) $+ b_3$ (MANAGERIAL INCENTIVES) $+ b_4$ (REGULATORY ISSUES) $+ b_5$ (ASSETS) $+ b_6$ (AGE) $+ b_7$ (BROKERAGE AVAILABILITY) $+ b_8$ (DL) $+ b_9$ (EXPENSE RATIO) $+ b_{10}$ (FL) $+ b_{11}$ (FUND FAMILY) $+ b_{12}$ (FUND FOCUS) $+ b_{13}$ (HISTORICAL EARNINGS GROWTH) $+ b_{14}$ (INSTITUTIONAL) $+ b_{15}$ (STOCK) $+ b_{16}$ (SD) $+ b_{17}$ (TOTAL RETURN) $+ b_{18}$ (TURNOVER RATIO) $+ b_{19}$ (12B-1 PLAN) $+ e$
(4a)

$CF = a + b_1$ (EXCELLENT/GOOD GOVERNANCE) $+ b_2$ (ASSETS) $+ b_3$ (AGE) $+ b_4$ (BROKERAGE AVAILABILITY) $+ b_5$ (DL) $+ b_6$ (EXPENSE RATIO) $+ b_7$ (FL) $+ b_8$ (FUND FAMILY) $+ b_9$ (FUND FOCUS) $+ b_{10}$ (HISTORICAL EARNINGS GROWTH) $+ b_{11}$ (INSTITUTIONAL) $+ b_{12}$ (STOCK) $+ b_{13}$ (SD) $+ b_{14}$ (TOTAL RETURN) $+ b_{15}$ (TURNOVER RATIO) $+ b_{16}$ (12B-1 PLAN) $+ e$
(4b)

Where *CF:* Cash flows of the fund and all other variables are as defined above in equation 2 and 3.

The second part is an estimation of coefficients for the above three models using panel data approach. The panel data approach allows for pooling of observations on a cross-section of open-end funds over two years. When observations possess the double dimension (cross section and time series), the crucial aspect of the problem is to have a clear understanding of how differences in behavior across individuals and/or through time could and should be modeled. A panel data set offers several econometric benefits over traditional pure cross section or pure time series data sets. The most obvious advantage is that the number of observations is typically much larger in panel data, which will produce more reliable parameter estimates and, thus, enable us to test the robustness of our linear regression results. Panel data also alleviates the problem of multicollinearity, because when the explanatory variables vary in two dimensions (cross-section and time series), they are less likely to be highly correlated. Panel data sets make it possible to identify and measure effects that cannot be detected in pure cross section or time series data. For instance, sometimes it is argued that cross section data reflect short-run behavior, while time series data emphasize long-run effects. By combining the cross-section and time series features of a data set, a more general and comprehensive dynamic structure can be formulated and estimated. The use of panel data suggests that individuals, firms, states, or countries are heterogeneous (Balestra 1995). Time series and cross-section studies not controlling for this heterogeneity run the risk of obtaining biased results (Baltagi 2000). Panel data controls for individual heterogeneity.

The most intuitive way to account for individual and/or time differences in the context of panel data regression is to use the fixed effects model. Fixed effect model assumes that difference across mutual funds can be captured in differences in the constant term. The regression coefficients (the slope

parameters) across groups in this model are unknown, but fixed parameters. It is also known as least square dummy variable (LSDV) model and we use LSDV fixed-effect model to estimate cost efficiencies in the mutual fund industry.

EMPIRICAL ANALYSIS

Table 1 shows cash flow growth on an average for each of the three years with highest cash flow growth for the year 2005. For each of the three years, cash flow growth is slightly higher for funds with poor governance record than for funds that received a grade of excellent or good. Funds with poor governance record also have higher brokerage availability on an average in each of the three years. This may explain higher cash flow growth. Fund focus or fund concentration has been rising each year and on an average, more focused funds show a below good governance record in comparison to funds with good or better governance grade. 12b-1 plan charges, on the average, have been between 0.43 and 0.44 during these three years. Funds with better governance record show lower 12b-1 charges in comparison to funds with a governance grade below good. Similarly, front-end and deferred load charges are higher on an average for funds with a governance grade below good. Three year annualized standard deviation is also slightly lower for funds with a governance rating of excellent or good. Furthermore, funds with a governance rating below good have a higher turnover ratio, which may indicate that management is more actively managing these funds through active buying and selling of securities to improve return. The average size of a fund with a good governance record is considerably higher than the average size of a fund with a governance rating below good. One of the criteria used in rating a fund's governance record includes expense ratio. Therefore, it is not surprising to find out that funds with better governance rating have a lower expense ratio in contrast to funds with a poor governance rating. On an average, funds with better governance record are older in age. Total 12 month return is also better for funds with a governance rating of excellent or good.

Table 2 shows that our model (equation 2a) explains 15 to 24 percent of the variation in total return. Board quality, which is our first indicator of governance, is highly positively related to the total return. Board quality as measured by independence of the board as well board supervision of the fund results in improved total return for fund investors.

Our second indicator of governance, managerial incentive is positively related to total return in one out of three years. For 2005 and 2006, we do not find statistically significant relation between managerial incentive and total return. Panel data analysis shows that better managerial incentives explain a higher total return and the relationship is statistically significant. Panel data analysis shows that regulatory issues and corporate culture do not play a statistically significant role in explaining total return of a fund on an average. Therefore, among the four governance variables, board quality is the only variable that, on an average, consistently explains the total return of a fund.

Equation 2b model shows that funds that have an overall rating of excellent or good show a higher total return for each of the three years. Panel data analysis supports the yearly results. With regard to control variables, Table 2 shows that on an average, fund size is positively related to a fund's total return and the relationship is statistically significant in both the panel data regressions. Age of the fund is negatively related to fund's total return in all the eight regressions and as per panel data regressions, there is a weak statistically significant negative relation between a fund's age and fund's total return. On an average, older funds exhibit a lower return in contrast with newer funds. It may be due to new fund bias as mentioned by Wisen (2002). Relationship between cash flow and fund's total return is mostly

negative except for the year 2005. For 2004 and 2006, on an average funds with a higher total return exhibited a negative cash flow growth. Panel data results also confirm the negative and statistically significant relationship between a fund's total return and cash flow growth. Deferred load is not statistically significant in explaining the total return of a fund. Front-end load is negatively related to total return of a fund and the coefficient is statistically significant as per panel data analysis. Expense ratio is positively related to the total return of a fund and the relationship is statistically significant for each of the three years. Therefore, the industry's argument that you get what you pay for holds ground as per our analysis.

Larger the number of funds in a family, larger will be the total net asset value return of a fund. In six out of eight regressions, the relationship between size of the family and total return is statistically significant. Fund focus or fund concentration is positively related to the total return except for the year 2004. Panel data models also show better performance for focused funds. Therefore, consistent with previous studies, we find that focused funds perform better than non-focused funds. Historical earnings growth is negatively related to current twelve month total return of a fund. High historical earnings growth does not translate into higher total return on an average for a fund. As expected, volatility as measured by three-year annualized standard deviation is positively related to the total return. Aggressive fund management through large scale buying and selling of securities does not translate into a higher total return, because turnover ratio is negatively related to the total return of a fund. Funds with lower turnover ratio get a better total return. For year 2004, the coefficient on stock fund is negative, which means stock funds earn a lower rate of return. In 2005 and 2006, there is no statistically significant relation between a fund being an equity fund and total return. Panel data analysis shows a statistically weak, but negative relation between total return and a stock fund. Funds with 12b-1 charges do not perform better. 12b-1 charges have a negative coefficient and are highly statistically significant in each of the three years. Results are consistent with panel data analysis.

Table 3 shows that our model explains 77 to 79 percent of the variation in expense ratio of funds during the year 2004 through 2006. Each of the governance variables, board of directors, managerial incentives, compliance with regulatory issues, and corporate culture are negatively related to the expense ratio of the funds. Panel data analysis shows that the negative coefficient on board quality, managerial incentive, regulatory issues, and corporate culture is statistically significant in explaining the expense ratio. Therefore, better governed funds have lower expense ratios. Empirical analysis of the equation 3b model also shows that funds that received an overall governance of excellent or fair have lower expense ratios.

Among the control variables, stock funds, fund family, fund focus, institutional funds, age, and size of the fund as measured by assets are negatively related to the expense ratio of the fund. These findings are consistent with previous studies except for the stock fund showing a lower expense ratio in contrast to bond funds. In two out of three years, stock funds do not show any statistically significant relation to the expense ratios. Only in 2004, stock funds show a statistically significant relation to the expense ratio of the fund.

Furthermore, 12b-1 plan, front-end load, deferred load, standard deviation, turnover ratio, brokerage availability, and total returns are positively related to the expense ratio. These findings are also consistent with previously published studies on mutual fund expense ratios. Cash flows of a fund do not show any statistically significant relation to the expense ratio except for the year 2006. In 2006, the cash flow is negatively related to the expense ratio and is weakly statistically significant, which mean

investors have started paying attention to the expense ratio of a fund. Higher expense ratio funds are attracting lower cash flows.

Table 4 shows that our regression model explains 15 to 42 percent of the variations in the cash flow of a fund from 2004 to 2006. Governance variables like board quality, managerial incentives, regulatory issues, and corporate culture do not show a consistent relation to the cash flow of a fund. Board quality is weakly negatively related to the cash flow of a fund in 2006 (model 4a). Panel data analysis, on the other hand, shows a weak positive relation between board quality and cash flows. Therefore, better board quality is attracting higher cash flows as per panel data. Managerial incentives show a positive relation to cash flows in 2004. Similarly, regulatory issues are negatively related to cash flows in the year 2006.

Empirical results from equation 4b also show funds that received an overall governance grade of excellent or good attract lower cash flows in 2004 and 2005. In the year 2006, relationship between cash flows and fund governance rating shows a statistically significant and positive relationship. It seems that average investor is only beginning to understand the impact of governance quality on the return and expenses of a fund. Therefore, governance rating shows a positive relation to the cash flows of a fund in 2006 only.

Among the control variables, fund size as measured by fund assets, fund age, deferred load, expense ratio, institutional, and 12b-1 plan variables do not show any statistically significant relation to the fund's cash flows. Availability of a fund through several brokerage houses, does on an average help improve a fund's cash flows. Panel data models show that there is a positive and statistically significant relationship between a fund cash flows and brokerage availability.

Panel data models also show that size of the fund family does not have statistically significant relation to the fund's cash flows, but in the year 2004, model 4b shows that fund family affects cash flows in a positive way. In 2005, both models show a statistically weak and negative relation to the cash flows. Beginning with 2005, focused funds, on an average, have been able to attract higher cash flows as reflected by the positive coefficient on fund focus for the years 2005 and 2006. Panel data models also confirm that investors have put more money in focused funds. Historical earnings growth is highly positively related to the cash flows of a fund. Investors pay attention to historical earnings growth and invest in funds that have shown high earnings growth in the past. High standard deviation is also positively associated with cash flows of a fund and is statistically significant. Funds that are more actively managed attract a higher amount of cash flow. This is reflected in the positive and statistically significant coefficient associated with fund turnover ratio. Fund cash flows are negatively related to the current 12-month total return of a fund. This may be due to investors cashing out on high fund returns.

SUMMARY AND CONCLUSIONS

Corporate governance issue has been a hot topic in finance literature in recent years especially after the corporate scandals of the late 1990s. Within the framework of corporate governance, conflict of interests between mutual fund managers, fund sponsors and shareholders have also attracted attention. Using the annual data from *Morningstar* for the year end 2004 through 2006, we investigate if funds with good governance record are able to earn a higher rate of return and have low expense ratios. We also investigate if track record of good governance is noticed by potential investors and translates into larger investment dollars for a mutual fund. *Morningstar's* overall fund governance grade that ranges from A

(best) to F (worst) is based on regulatory compliance, board quality, managerial incentives, fees, and corporate culture. We analyze the impact of overall fund governance grade on fund performance, fund expenses and fund cash flows. Furthermore, we also evaluate the impact of each of the components of the governance grade on fund returns, fund expenses, and fund cash flows.

An analysis of the annual data for each of the three years shows that on average better governed funds have a higher cash flow growth, lower 12b-1 charges, front-end and deferred loads, and expense ratios. Also, better governed funds on average are larger in size and older in age. Poorly governed funds market themselves by making them available through more brokerage houses. We find that governance does impact overall fund performance, because superior governance (governance grade of A or B) is associated with superior fund total return. Also, better governed funds have lower expense ratios, which ultimately translate into higher returns for investors. Fund investors have started paying attention to fund governance recently only. In the year 2006, better governed funds attracted higher cash flows.

This study also finds that governance variables as measured by board quality, managerial incentives, regulatory issues, and corporate culture do play a role in influencing fund returns, fund expense ratios, and fund cash flows. Board quality and managerial incentives are associated with better fund returns.

On an average, all the four measures of good fund governance are negatively related to a fund's expense ratio. Funds with good board quality, better managerial incentives, good regulatory compliance, and good corporate culture have a lower expense ratio, which means higher returns for the fund shareholders.

Since fund investors have started paying attention to the governance issues only recently, fund cash flows do not show a statistically strong relation to governance variables.

Morningstar mentions that *Morningstar's* governance (stewardship) grade for funds is entirely different from the *Morningstar Rating* for funds, commonly known as the star rating. There is no relationship between the two. Therefore, an interesting extension of this study would be to analyze the relationship between fund rating assigned by *Morningstar* and fund governance. Do better governed funds enjoy a higher rating? Another extension would be to examine if there is a statistically significant difference in the risk-adjusted performance of better governed and poorly governed funds.

TABLES, FIGURES, & REFERENCES

Tables, figures, references, and full paper available upon request from the authors.

A Semantic Approach to Secure Electronic Patient Information Exchange in Distributed Environments

Atif Khan, Helen Chen, and Ian McKillop

David R. Cheriton School of Computer Science,
University of Waterloo
200 University Avenue West Waterloo, ON, Canada N2L 3G1
`{a78khan,helen.chen,ian}@uwaterloo.ca`

Abstract. Modern medical information systems collect large amounts of diverse patient data in order to facilitate a higher level patient care. Although desirable, this functionality has a tremendous potential for abuse, where patient information can be shared, disclosed and used for other (secondary) purposes. In most cases, patient consent is solicited and institutional policies are put in place to limit the privacy and security risks. However, in practice these measures have proven to be inadequate, resulting in violation of patient consent even for non-life threatening scenarios. We propose a framework to capture privacy & security policies and to protect exchange of sensitive medical patient information. Our framework is comprised of distributed multiagent environments reflecting healthcare institutions and personnel. We utilize semantic techniques for data representation and reasoning. Furthermore, we do not require pre-established trust relationships to be present for exchanging private sensitive information between multiple parties. In our proposed framework, all decisions to share information, are backed up by semantic proof of authorization that can be verified by an independent third party.

1 Introduction

Modern medical information systems, along with Information Communication Technologies (ICTs), are key enablers for providing patients with a high level of medical care [15]. These systems capture and process large amounts of heterogeneous data from many diverse and distributed sources. Furthermore, information can be exchanged between various parties in many different formats, ranging from summaries of medical records to detailed diagnoses and test results.

Although the use of above mentioned technologies drastically improves the effectiveness of information exchange, coordination, and use, it also raises the critical issue of patient information privacy and security. Usually, patient information is collected for the primary purpose of providing healthcare for a specific episode. Any secondary usage must be in accordance with and governed by patient consent. It has been argued that a patient should be aware of all the systems collecting their information, and should be able to specify how this information can be used [14]. Given the complexity and diversity of medical information exchange scenarios, patient consent and privacy policies are often ignored for the greater good.

The use and protection of sensitive patient medical records is further complicated by the fact that these records are fragmented over many distributed heterogeneous storage systems. Although the information contained in these distributed segments is to be used in accordance with a global patient consent policy, each individual medical information system may be governed by its own privacy and security policies. Furthermore, these privacy and security policies might be augmented by jurisdictional policies (such as a provincial healthcare privacy and security act).

In this paper, we present a framework for describing, capturing, processing and managing sensitive patient medical information, based on patient consent and other applicable privacy and security policies. Our solution is centralized around the observations that (a) information contained in a patient's medical record should be governed by the patient's consent policy and (b) the security and privacy of the physical/electronic medical records must be guaranteed by the institutional privacy and security policies. Our proposed solution strives to honour both of the above mentioned conditions.

In our proposed system, all knowledge representation, acquisition, and exchange is based on the use of ontologies best suited to describe the various information sets. For example, patient consent is obtained and represented based on any ontology capable of defining the concepts required to capture such information. The main motivation to use an ontological approach is (a) to facilitate the exchange of information with other parties and (b) to represent information in an automated machine processable format.

Given the diversity of healthcare professionals and the heterogeneity of medical information systems, along with the complexity of information exchange, we have chosen to represent the healthcare domain as a distributed multiagent system. In our multiagent environments, each intelligent agent either represents an entity or acts as a proxy to an entity. All entities exist in a cooperative mode working towards maximizing the overall system utility. The system utility is measured in terms of the total benefit provided to a patient as a result of agent interactions.

In order to ensure the privacy and security aspects of all information exchanged between agents, we define an *information exchange handshake protocol*. We make the assumption that all agents can be monitored within their respective environments and therefore are not capable of being malicious. The protocol starts with the *request for information* where an agent sends a request for accessing sensitive medical data from another agent. The second phase of the protocol deals with validation of the various policies guarding the sensitive information. The requesting agent is asked to fulfil the security and privacy criteria identified by the agent holding the information. The requesting agent generates a proof in fulfilment of the requested criteria. The proof has the unique property of being verifiable by any party capable of computing it. Upon the validation of this proof, the requested information is delivered in response to the original request.

It is worth noting that at no point does our framework assume a (pre-established) trust relationship to have existed between the two agents exchanging information prior to the exchange. Each request for information is independently verifiable. Therefore, we can establish dynamic trust on a per request basis. Furthermore, considering that the validation of the policies protecting the patient information are part of each request response cycle, changes to these policies take effect almost immediately. This dynamic enforcement of patient consent and institutional privacy & security policies represent a major improvement over the trust-based [2] and role-based-access [22] medical information exchange frameworks.

In order to provide motivation for our work, consider the following example. *Let us assume that a patient John is primarily treated by Dr. Smith at the Toronto General Hospital (TGH). During his vacation in Calgary, John was found unconscious and was admitted to Calgary General Hospital (CGH). John is now being treated by Dr. Jane who requires access to John's past medical history, in order to properly diagnose and treat John. Dr. Jane requests access to John's medical records from TGH. For the purpose of illustration, lets us further assume*
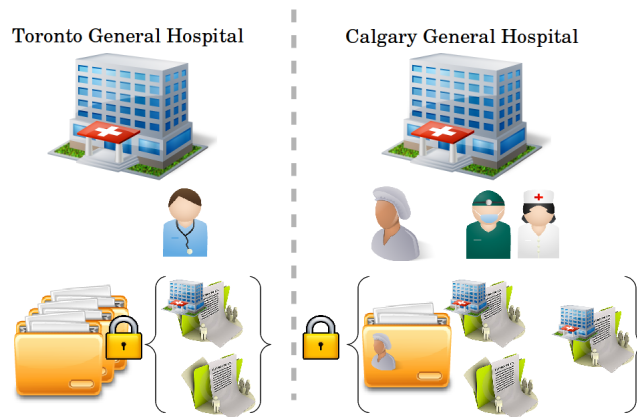
Fig. 1. A motivational example: John's EMR is protected by his consent and TGH's privacy & security policies. When Dr. Jane at CGH requests access to these documents, they are shared under the same protection that was offered by TGH. Furthermore, CGH may apply their own privacy and security policies in addition to the existing ones.

that John's medical records are protected by his consent policy and the TGH institutional privacy and security policies. Please refer to Fig:1.

The goal of our system is to allow Dr. Jane to be able to retrieve the required patient medical information (for John) from TGH, while still honouring all of the established privacy and security policies put in place by the institution and John's consent policy.

2 Related Work

Our proposed framework utilizes semantic knowledge representation and multiagent systems. Although both of these areas of artificial intelligence have been well researched and well understood, we have not found any existing work that utilizes semantic technologies with multiagent systems to address our goals specifically. In this section, we present a general literature review of these fundamental AI concepts. Our framework is motivated by the core concepts introduced and discussed in the papers selected in the literature review. The literature review focuses on semantic knowledge representation, various aspects of multiagent systems (such as use of semantic technologies in multiagent systems and trust establishment between agents) and electronic patient consent.

2.1 Traditional Authorization & Access Control

There have been several authorization and access control systems suggested in the literature [4,7,14,22]. Most of these systems are based on role-based access control (RBAC) and do not address patient consent. In RBAC, roles are associated with (access) privileges, and system users are then assigned roles based on the nature of their job or functionality.

Although the RBAC approach is quite effective within the confines of a single organization, it does not scale well when dealing with dynamic healthcare environments, where entities (such as doctors) can take on many different roles across many different organizations. Furthermore, RBAC based security frameworks are not real-time in nature. There is a considerable lag when it comes to enforcement of updated/new roles.

Our proposed framework is fundamentally different from the RBAC type approaches, as in our framework all access decisions are made based on the available information. Therefore, we can easily accommodate for the varying user roles and apply any changes to these roles in real-time.

[30] defines a dynamic role-based access control system with similar properties to our framework (such as a multiagent system with trust negotiation). However, their approach does not offer the semantic compatibility that our system provides. The trust establishment requirements are also not as flexible, requiring trust to be pre-established. Our approach is far more flexible and secure, since it requires all agents to negotiate trust on a per request basis.

2.2 Ontology Based Knowledge Representation & Healthcare

Ontologies have been heavily utilized in the area of medical informatics. However, the main goal of these ontologies has been to define and represent medical knowledge, and not privacy and security related concepts (as is the case in our solution).

Binfeng et al. in [3] explore building a medical knowledge base using a medical ontology for coronary heart disease. Their knowledge base has the interesting property of being able to map concepts back and forth between traditional Chinese and modern Western medical ontologies.

Cassimatis et al. in [5] argue that "Systems with human-level intelligence must both be flexible and be able to reason in an appropriate time scale. These two goals are in tension, as manifest by the contrasting properties of structured knowledge-based systems". They propose an interesting approach (reasoned unification) for representing and reasoning over linguistic and non-linguistic knowledge, within the scope of an inference context. Considering that medical information is a complex combination of various different types of data-sets, the ideas present in [5] have a significant application potential for healthcare information systems.

Another salient property of healthcare information systems is the use of many specialized domain specific ontologies by the respective specialized faculties. Therefore, in order to exchange and reason with information across all systems, we need the ability to translate the ontological concepts back and forth. [17] suggest an ontology learning framework for similar purposes. Their proposed framework provides an ontology learning environment with semi-automatic ontology-construction tools.

2.3 Multiagent Systems & Structured Knowledge

Semantic technologies for knowledge representation and processing seem to be very well suited for multiagent systems [6,9,11,16,28,31]. This comes as no surprise, considering that the premise of semantic knowledge representation is to facilitate machine processing of information.

[31] enhances the single coordination server limitation of the Trading Agent Competition (TAC) [25] scenario to work under the Agentcities [10] distributed agent model. Semantic web languages and tools are used to define (i)FIPA compliant ontology based agent communication language (ACL) and (ii)the knowledge-base for the agents to work with. [31]

found the use of semantic web technologies enhanced the interoperability between agents in multiagent environments.

[16] presents a very similar multiagent framework that has the ability to support multiple ontologies. Furthermore, their solution makes use of FIPA-compliant JADE agent framework to define semantic web ontology service and an inference service. These services act as middleware to support agent management, agent communication, and agent interaction protocols.

[6] proposes a ubiquitous computing system facilitating context-aware intelligent agents for the purpose of providing meaningful relevant services to individual participants in a meeting room setting. Their system is context-aware in order to personalize the user experience. The multiagent environment utilizes a semantic representation to describe the context, making it feasible for the agents to exchange and reason with the information present within the context.

[11] considers web enabled multiagent environments, where agents are distributed and provide unique services. The agents utilize customized ontologies to define and process their services. The semantic representation makes it possible for multiple agents to work together. [11] also explore the use of semantic technologies for agent communication language (ACL), where the terms used during agent-to-agent communication may originate from various ontologies.

[28] investigates the impact of agents using multiple domain specific ontologies in multiagent environments. They propose an inter-agent semantic concept learning approach to deal with the proliferation of (domain specific) ontologies. However, their solution is quite restrictive and requires closed world representation of information, where all agents are known to each other and have complete knowledge of all the information contained in the world.

Intelligent multiagent systems have been successfully utilized in healthcare settings for various applications. [9] proposes an architecture for an intelligent multiagent clinical decision support system . Although the proposed architecture is not specifically targeted for privacy and security, there are some fundamental similarities between the propose architecture in [9] and our framework (such as the use of intelligent agents to augment physician productivity in a multiagent environment).

2.4 Trust Management in Multiagent Systems

There are many definitions of trust when it comes to distributed systems. [21] defines trust in the context of multiagent systems as - "a belief an agent has that the other party will do what it says it will (being honest and reliable) or reciprocate (being reciprocative for the common good of both), given an opportunity to defect to get higher payoffs".

A trust relationship between two agents can be reasoned about and calculated using trust models capturing reliability and honesty of agents involved. Furthermore, an agent can define the various levels of trust it might place with another agent. [21] defines two broad categories of conceptualizing trust:

- "**Individual-level trust**, whereby an agent has some beliefs about the honesty or reciprocative nature of its interaction partners."
- "**System-level trust**, whereby the actors in the system are forced to be trustworthy by the rules of encounter (i.e. protocols and mechanisms) that regulate the system."

[12] recognizes trust as a major issue in the area of multiagent systems. Their investigation is motivated by the following three questions: (i)Why does an agent trust another? (ii)How do agents judge or evaluate the trustworthiness of others? (iii)What does an agent do after obtaining the trustworthiness of others?

[13] argues that existing trust and reputation models cannot be used for dynamic multiagent environments, where agents continuously join and leave the system. This dynamic behaviour adversely impacts the overall performance when existing trust models are used. They propose a new trust and reputation model for multiagent systems, which utilizes various forms of trust (such as role-based trust, witness reputation etc.) to produce a comprehensive score corresponding to an agent's trustworthiness.

[26] explores how non-cooperative distributed agents, when forced into working together, can utilize a trust based model to facilitate their interactions. Their mechanism focuses on two basic parameters (i)agent attributes and (ii)reliability values.

2.5 Electronic Consent

There are numerous studies dealing with electronic consent. However, these studies ignore the semantic aspect of information and focus mainly on security aspects [4,7,22]. O'Keefe et al. [18] undertake a feasibility study of electronic consent management systems in the medical arena. They expose various challenges faced by different consumer groups of electronic consent management systems. The study provides a sound set of recommendations for a generic implementation of a patient consent management system.

Song et al. [24] introduces the notion of an e-consent object, encompassing all relevant information concerning patient consent in the e-consent object. Lack of semantics is the biggest drawback of this model. The rules of consent are not expressed in any formal language and therefore are ambiguous at interpretation time.

Win et al. [29] describe an interface based approach through which patient consent can be expressed. The solution lacks organic growth as it hard codes the information and lacks the required flexibility for the user.

Pruski et al. [20] propose e-CRL language designed with the following two goals in mind (a)facilitate capturing of patient consent information (b)formalize the expression of patient consent information. "The language has a well defined BNF (Backus Naur Form) based syntax and semantics defined based on first-order logic and set theory which allow eHealth systems to fully control the access to critical health data" [20].

Although the e-CRL language provides support for semantics, it lacks some important features, such as proof generation. Furthermore, the defined language is not compatible with the RDF [1] based solutions and approaches, making integration difficult.

3 Proposed Solution

We propose a framework for exchanging sensitive patient information between multiple parties, while enforcing all required security and privacy policies along with patient consent. Our framework represents the various entities (such as hospitals, doctors, patients, staff etc.) using distributed multiagent environments, where each environment consists of intelligent agents (IAs) either representing or augmenting the functionality of an entity. All agents exist in a

cooperative mode working towards maximizing the overall system utility. We measure this system utility as the gain the patient receives from the use of our framework.

Information exchange between multiple parities is facilitated via their respective agents, where each agent takes part in a three-step information exchange handshake protocol. We currently do not address malicious agents based on the assumption that an agent's malicious activity can be detected by its corresponding environment. We utilize a semantic ontology based approach to define (a)the patient consent and (b)the privacy and security policies. The main motivation behind this approach is as follows:

– Ease of information exchange and distribution between multiple parties, while preserving the meaning of the exchanged data.
– All exchanged information is machine processable. This allows agents from different environments to not only share data with each other, but also process data in an intelligent format. Since the information is presented in a semantic format, agents can reason with the available data. They can also infer additional knowledge that might not be obvious from just the raw data.

All access decisions are processed by a semantic reasoner such as [23]. The semantic reasoner consumes the knowledge-base and the rules for knowledge inference to produce an answer for the information-access query. For each decision, the reasoner also outputs a semantic proof. This proof has a verifiability property by which any third party can compute and verify the proof independent of the proof creator. This allows us to establish trust for each access request, thus eliminating the need for pre-established trust requirements between the parties involved.

Our framework treats each access request as a single transaction. All transactions are mutually exclusive. That is to say, any previously computed access decision is not reused for the fulfilment of future requests. This allows for enforcement of all related policies on a per request basis. Therefore, a change in any policy will be applicable almost instantaneously, allowing dynamic enforcement across all environments and access requests.

Let us now discuss the details of the various components of our proposed framework.

3.1 Patient Consent Representation

There are many different forms of consent that a patient may choose from. Coiera and Clarke [8] define four general forms of patient consent. In our framework, we define the various types of patient consent policies (based on [8]) as follows:

– *opt-in*: A patient who has an opt-in policy allows any treating doctor to access their information.
– *opt-in-with-sensitive-documents-override*: In this case, the patient allows access to all their information except for documents that are classified as sensitive. For example, these documents may include human immunodeficiency virus (HIV) test results or sexually transmitted disease (STD) records.
– *opt-in-with-entity-override*: In this case, the patient allows access to all their information, but may specifically deny certain individuals, healthcare providers and organizations from accessing their data.
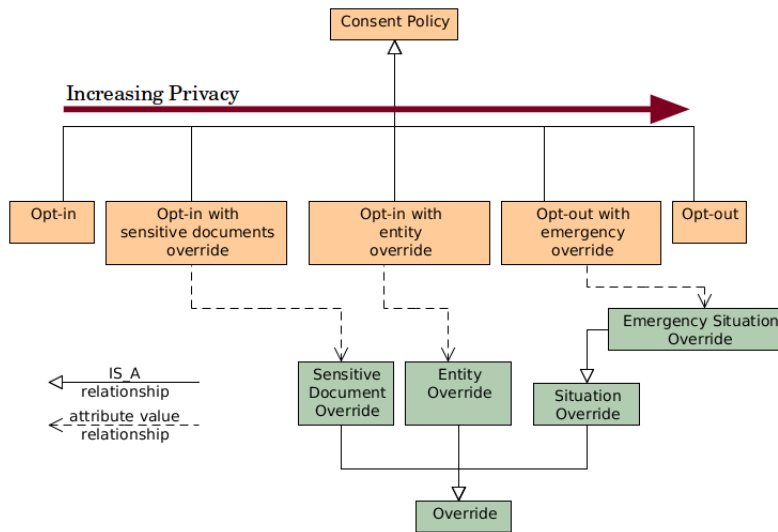
Fig. 2. A simple patient consent semantic model

- *opt-out*: A patient with an opt-out policy explicitly denies access to all their information regardless of who is trying to access the data or why they are trying to access it.
- *opt-out-with-emergency-override*: In this case, the patient agrees to grant access to their information only if it is an emergency situation, such as a life threatening scenario.

In order to establish a semantic representation of the above mentioned patient consent options, we need to build an ontology. It is important to note that we do not limit the type of ontology used to define patient consent. Any ontology capable of representing the core concepts (as stated above) will be acceptable by our framework. For the benefit of simplification and understanding, we define a very simple ontology model to represent patient consent concepts (please see Fig:2).

Based on our motivational example, let us assume that patient John had chosen *opt-out-with-emergency-override* as his consent policy C_{John}. Therefore, when Dr. Jane requests access to John's medical records from TGH, she should only be allowed access to this information if John's condition at CGH has been identified as life threatening. Otherwise, Dr. Jane's request for accessing John's medical records should be denied.

Note that for the purpose of John's consent policy, we do not really need to know the details of his life *threatening condition*. We simply need an agent at CGH to assert this fact based on the available information. Also note that this assertion will be part of the proof when executing the information exchange handshake protocol, and is also verifiable by an independent third party.

3.2 Privacy & Security Policy Representation

Traditionally, institutional privacy and security policies have proven to be complex in nature. This complexity is generally a function of the information and the relationships/rules defined in the policy. Therefore, an ontology based semantic representation of these policies seems like a natural choice. The overall process of translating a specific institution's policies into

its semantic representation might be time consuming, but it is not a technically challenging activity.

Similar to the consent policy, our framework supports any ontology for representing institutional privacy and security policies. The only requirement being imposed is that the selected ontology be capable of describing the relevant privacy and security concepts and their relationships.

For the purpose of clarification, we now turn back to our example under consideration. The two institutions involved are TGH and CGH. Let us assume that the privacy and security policies for these institutions are defined as follows:

– Pol_{TGH} : TGH only allows employees to access patient medical records. The employee accessing patient information must also be treating the patient (or part of the team taking care of the patient). Furthermore, the employee must be on shift and must be a physician, when accessing patient information.
– Pol_{CGH}: CGH defines a very different and relaxed privacy and security policy. It allows access to patient medical records by all hospital employees. There are no constraints similar to the ones defined in Pol_{TGH}.

Security & Privacy Axioms: Based on John's consent (C_{John}) and the the institutional policies (Pol_{TGH} & Pol_{CGH}), we can now define the following privacy and security axioms:

– All information for patient John must abide by $\{\text{C}_{\text{John}}\}$, independent of the location where the information resides.
– All information for patient John at TGH is protected by the protection set PS = $\{\text{C}_{\text{John}}, \text{Pol}_{\text{TGH}}\}$. Any entity or agent requesting information for patient John must provide proof of fulfilment to satisfy the set PS.
– When information is exchanged, the receiving party can augment the received protection set PS with its own privacy and security policy. Therefore, when CGH receives John's medical records, the exchanged information is now protected by a new protection set PS' = $\{\text{Pol}_{\text{CGH}}, \text{PS}\}$ at CGH.

3.3 Distributed Multiagent Environments

Our framework utilizes the multiagent systems [27] paradigm. Intelligent agents are used to either augment or replace the functionality of real world entities. For example, a physician agent can enhance the efficiency of a physician by interacting with the various information systems to collect patient data on the physician's behalf.

We group multiple intelligent agents together in an environment. All agents within a single environment are structured in hierarchical subgroups. An institution agent is the top most agent in an environment, and governs all agent interactions. We assume that all agent interactions can be monitored within the scope of an environment. By using this construct, we can define an environment for a hospital made up of supporting agents.

Agents from different environments are allowed to communicate with each other. However, all sensitive inter-environment communication should be in accordance to the information exchange handshake protocol. We have chosen to include intelligent agents in our framework for the following reasons:
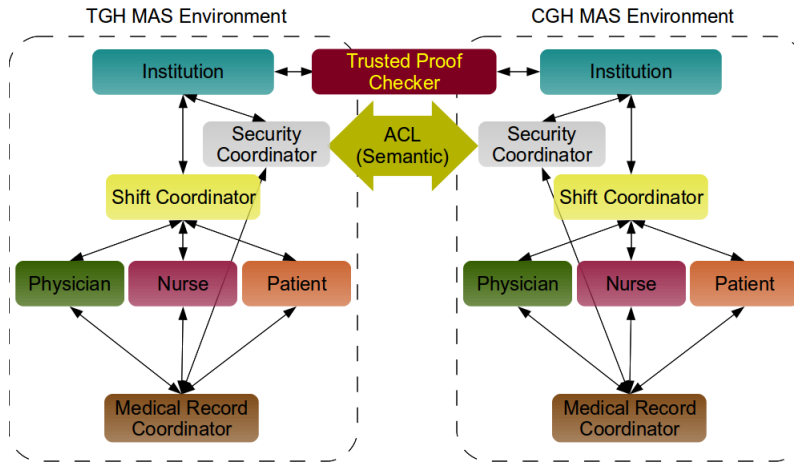
Fig. 3. Multiagent system healthcare environment model

– Distributed multiagent environments provide a reasonable approach to model the distributed heterogeneous healthcare institutions and healthcare providers.
– Using semantic representation of knowledge, intelligent agents can consume and process diverse data-sets from many different sources.
– A semantic reasoner can easily be integrated into an intelligent agent as opposed to an entity from the real world. This ensures complete adherence to evidence-based decision making that human agents might not be able to achieve.
– It is far easier to rely on an intelligent agent to follow all the required privacy and security protocols. All malicious activities of an intelligent agent would require breech of protocol that can easily be detected by the other agents in the environment.
– Given the same knowledge-base and reasoning rules, an agent will make the same predictable decision, hence reducing inefficiencies caused by human agent errors.
– Considering that we are proposing dynamic trust on a per request basis, it is far more efficient to utilize intelligent agents rather than human agents.

Agents utilize an agent communication language (ACL) when communicating with each other. Our framework imposes no restrictions on the choice of a particular ACL as long as the information required for successfully executing the information exchange handshake protocol can be exchanged. However, having a semantic based ACL would ensure that heterogeneous multiagent environments will be capable of communicating with each other.

We assume that all agents operate in a cooperative mode within an environment. All agents share the same utility function and work towards maximizing the overall system utility. We propose to measure this overall system utility as a function of an increase in patient privacy and security, as a result of using the multiagent system.

In support of our motivational example, we define two simple environments representing TGH and CGH (see Fig:3). The agents in both environments are similar in nature. An institutional agent governs its local environment and all encompassing agents. A security coordinator provides all privacy and security guarantees and is responsible for identifying institutional privacy and security policies. A medical record coordinator works in conjunction with the security coordinator to manage the sensitive patient information. Patients, physicians and other healthcare entities are represented by their corresponding agents. We also

define a trusted third-party proof checker agent that can be utilized to validate the semantic proofs required for information exchange, as discussed in the next section.

In order to exchange sensitive information between agents, it is desirable to have a trust relationship exist between the agents. Our framework assumes that no pre-established trust relationship exists between agents. Rather, it establishes trust on each individual request for information. The request based trust is a function of an agent's ability to satisfy the required privacy and security guarantees protecting the sensitive information that is to be exchanged. Therefore, any changes in patient consent or the institutional policies are reflected almost instantaneously.

3.4   Information Exchange Handshake Protocol

In our proposed framework, all information exchange between agents takes place after a successful completion of the handshake protocol. The handshake protocol is interactive in nature and requires a communication channel between all participating agents. Although synchronous real-time communication would be ideal, the protocol can be completed successfully even if the agents communicate in an asynchronous manner over any period of time. A time delay in the completion of the handshake protocol will only impact the delivery time of the requested information, and has no negative impact on the privacy and security guarantees.

The protocol has three phases that must be completed between the agent requesting the information and the agent responding to the request. We will refer to this as the request-response cycle. Given the fact that patient information may need to be aggregated from distributed heterogeneous systems, a requesting agent might engage in multiple request-response cycles with the distributed agents. Note that each request-response cycle is atomic in nature. That is to say, that an outcome of the request-response cycle does not influence the result of any other request-response cycle.

The phases of the protocol are (i)request for information (ii)proof generation and (iii)validation of the proof. These phases are described in detail as below:

Phase 1 – Request for Information: This is the first phase of the protocol in which a requesting agent identifies the patient and the various sources from where the patient information is to be accessed from. Note that our framework does not constrain how this identification process is to be completed. Once identified, the requesting agent initiates one or more request-response cycles based on the number of information sources identified.

Following our example, when Dr. Jane asks for patient John's medical records from TGH, the physician agent initiates the request. Since in our framework all communication between distributed agents must be facilitated by the corresponding institutional agents, the CGH institutional agent forwards the request from Dr. Jane's physician agent to TGH. The request is processed by TGH's institutional agent by consulting with the local (security coordinator and medical record) agents to identify the protection set PS = $\{C_{John}, Pol_{TGH}\}$ (John's consent and the privacy and security policies of TGH).

The protection set PS is returned to CGH asking them to provide a proof of fulfilment for all elements in PS. This marks the completion of the first phase.

Phase 2 – Proof Generation: This phase is initiated in response to receiving a protection set PS. The main goal being to mine the local knowledge-base and find evidence in
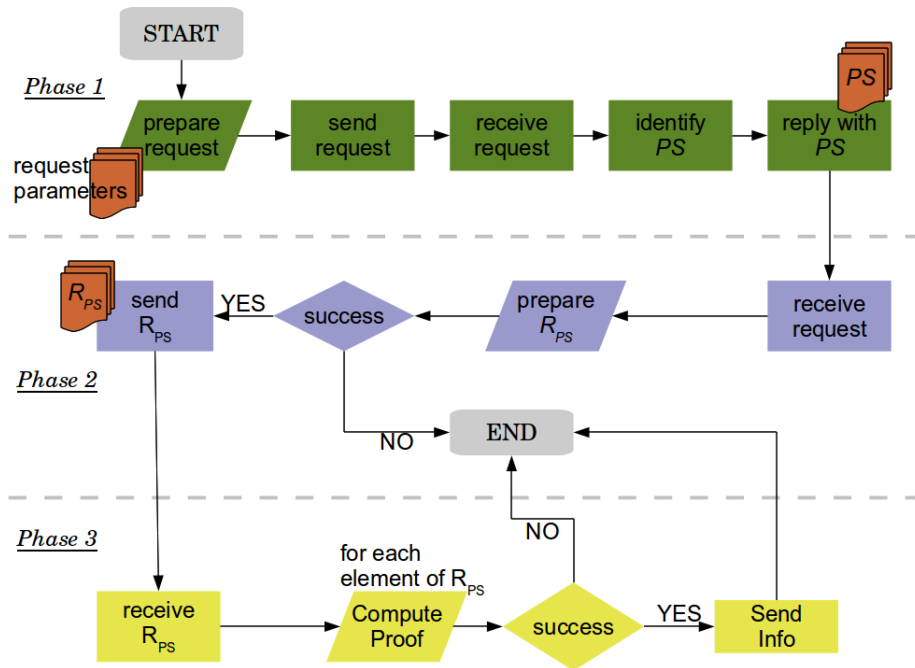
Fig. 4. Information exchange handshake protocol flow control.

support for fulfilment of the elements of PS. Since all our knowledge-base is represented in a machine processable semantic format, a semantic reasoner is utilized to discover facts and generate evidence from the knowledge-base. The reasoner then provides a semantic proof for the discovered knowledge.

Following our example, the CGH institutional agent will receive the protection set PS from TGH. The first element is patient John's consent that states opt-out-with-emergency-override consent policy. In conjunction with the local agents, the CGH institutional agent finds evidence that John's condition has indeed been diagnosed as life threatening. The second element in the PS represents TGH's privacy and security policy. The CGH institutional agent would then look for evidence to fulfil the requirements of Pol_{TGH}. It will establish that Dr. Jane (receiver of John's information) is indeed an employee of CGH, and is currently on shift and is treating patient John. Again a semantic reasoner is utilized to find the results and generate the proofs.

At this point, the CGH institutional agent has fulfilled all the requirements for PS. All the proofs are aggregated in a response set $R_{PS} = \{P1, P2, \ldots Pn\}$ which is sent back to TGH. This marks a successful completion of the second phase.

It is equally possible that the criteria in the protection set PS may not be successfully proved. For example, if Dr. Jane was not on shift, or another healthcare personnel (such as a nurse) was the initiator of the request for information. In the case of unsuccessful completion, the request-response cycle will terminate here.

Phase 3 – Validation of Proof: This phase begins when the the response set R_{PS} is received. Recall that the elements of R_{PS} are the various proofs generated by a semantic reasoner, with the characteristic of being verifiable by any independent third party. A party

validating the proof will generally compute the proof based on the evidence found to ensure the validity of the proof.

Based on our motivational example, the TGH institutional agent will receive a response set R_{PS} from CGH. It will then iterate over all the individual elements of R_{PS} and validate them. In an ideal case, this validation can be performed by an independent trusted third party proof checker TTPC agent. However, the existence of a TTPC is not necessary, since the proof can be locally verified at TGH.

Upon successful completion of the request-response cycle, the requesting agent will receive the information identified in the initial request. Please refer to Fig:4 for the flow control of the information exchange handshake protocol.

4   Future Extensions

4.1   Local Ontologies

Our assumption that patient consent can be expressed using the same ontology across all multiagent environments is not practical. Each institution may use a different ontology to define its privacy and security policies. These more realistic scenarios represent a major hurdle as it nullifies the global semantic understanding of data that we require.

It is also reasonable to assume that an institution may use a hybrid ontology, where the concepts are aggregated from various other ontologies. [11] actually advocates this to be a realistic scenario given an increase in semantic technologies. Furthermore, with an increase in domain specific ontologies, agents in multiagent environments will be forced to accommodate multiple domain ontologies [28].

In order to address this weakness, we will need to augment our proposed framework with the ability to identify and map similar ontological concepts across different ontologies. There are various approaches that can be taken here. A static mapping can be pre-defined between the different ontologies, mapping the concepts from one onto the other. A better approach would be to dynamically determine if two ontological concepts are the same. This could be achieved via a detailed inspection of the attributes and relationships defined for a given concept.

4.2   Context-based Information

When exchanging patient information, we identify the patient, the institution holding the information and the patient medical data. We do not consider the context of the information exchange. By introducing contextual information, we can further infer as to what parts of a patient's electronic medical record might be relevant for the exchange. The notion of a context by itself is nothing new. Many other system designs are context-aware such as [6,19,31]. Therefore, this should be a simple extension to our framework with high rewards.

4.3   Caching and Expiration of Proof Tokens

Currently we assume that all privacy and security related information, that is compiled and/or computed to generate the required proofs, is not reused. However, if a physician agent had computed that the physician is on shift as a result of some previous proof, then

the agent could potentially cache and reuse this proof for the duration of the physician's shift.

Although caching proofs, as stated above, would greatly increase the system performance, it has potential privacy and security risks associated. For example, caching a proof that Dr. Jane is a licensed physician in Ontario, over a long period of time might not be practical since Dr. Jane could have her license revoked at any time in the future.

4.4 Implementation

The main concepts of our framework (such as knowledge representation, inference rules definition, automated reasoning etc.) have been individually validated using simple proof-of-concept techniques and application. For example, we utilize N3 triple (knowledge) store to define patient consent according to the simple ontology created for the purpose of demonstration. Institutional privacy and security policies are also represented as N3 statements in a triple-store. An open source semantic reasoner, Euler [23], is used to query the knowledge-base. In order to fully realize the impact of our solution, we will need to integrate it into a real world system.

5 Conclusion

We presented a semantic multiagent framework to enhance the privacy and security guarantees protecting sensitive medical patient data. Our proposed framework utilizes an ontological representation of patient consent and institutional privacy and security policies. These protection guarantees are communicated and satisfied before sensitive information is exchanged. We allow distributed agents to share sensitive information without having pre-established trust relationships between two parties. Our framework utilizes the trust-establishment-per-request approach to ensure real-time enforcement of any changes to either patient consent or institutional privacy and security policies. All information exchange is governed by an information exchange handshake protocol, that takes place at the beginning. The protocol ensures all protection guarantees are fully met by requiring a semantic proof for each access requirement protecting the information. This proof has a unique property that it can be validated by an independent third-party.

References

1. Resource description framework (rdf). Website. http://www.w3.org/RDF/.
2. Oluwafemi Ajayi, Richard Sinnott, and Anthony Stell. Dynamic trust negotiation for flexible e-health collaborations. In *Proceedings of the 15th ACM Mardi Gras conference: From lightweight mash-ups to lambda grids: Understanding the spectrum of distributed computing requirements, applications, tools, infrastructures, interoperability, and the incremental adoption of key capabilities*, MG '08, pages 8:1–8:7, New York, NY, USA, 2008. ACM.
3. Xu Binfeng, Luo Xiaogang, Peng Chenglin, and Huang Qian. Based on ontology : Construction and application of medical knowledge base. In *Complex Medical Engineering, 2007. CME 2007. IEEE/ICME International Conference on*, pages 1586 –1589, May 2007.
4. B. Blobel. Authorisation and access control for electronic health record systems. *International Journal of Medical Informatics*, 73(3):251–257, 2004.
5. N Cassimatis. Flexible inference with structured knowledge through reasoned unification. *Intelligent Systems, IEEE*, PP(99):1 –1, 2009.

6. Harry Chen, Filip Perich, Dipanjan Chakraborty, Tim Finin, and Anupam Joshi. Intelligent agents meet semantic web in a smart meeting room. *Autonomous Agents and Multiagent Systems, International Joint Conference on*, 2:854–861, 2004.

7. X. Chen, D. Berry, and W. Grimson. Identity management to support access control in e-health systems. In *4th European Conference of the International Federation for Medical and Biological Engineering*, pages 880–886. Springer, 2009.

8. E. Coiera and R. Clarke. e-Consent: The design and implementation of consumer consent mechanisms in an electronic environment. *Journal of the American Medical Informatics Association*, 11(2):129, 2004.

9. G. Czibula, I.G. Czibula, G.S. Cojocar, and A.M. Guran. Imasc - an intelligent multiagent system for clinical decision support. In *Complexity and Intelligence of the Artificial and Natural Complex Systems, Medical Applications of the Complex Systems, Biomedical Computing, 2008. CANS '08. First International Conference on*, pages 185 –190, November 2008.

10. Jonathan Dale, Steven Willmott, and Bernard Burg. Agentcities: Challenges and deployment of next-generation service environments. In *Proc. Pacific Rim Intelligent Multi-Agent Systems*, 2002.

11. James Hendler. Agents and the semantic web. *IEEE Intelligent Systems*, 16:30–37, 2001.

12. Hongbing Huang, Guiming Zhu, and Shiyao Jin. Revisiting trust and reputation in multi-agent systems. In *Computing, Communication, Control, and Management, 2008. CCCM '08. ISECS International Colloquium on*, volume 1, pages 424 –429, August 2008.

13. T.D. Huynh, N. R. Jennings, and N.R. Shadbolt. An integrated trust and reputation model for open multi-agent systems. *Journal of Autonomous Agents and Multi-Agent Systems*, 13(2):119–154, 2006.

14. E.H.W. Kluge. Informed consent and the security of the electronic health record (EHR): some policy considerations. *International Journal of Medical Informatics*, 73(3):229–234, 2004.

15. S. Yunkap Kwankam. What e-Health can offer. *Bulletin of the World Health Organization*, 82:800 – 802, 10 2004.

16. Jaeho Lee. Toward a practical multi-agent framework utilizing the semantic web. In *Intelligent Agents and Multi-Agent Systems*, volume 2891 of *Lecture Notes in Computer Science*, pages 122–132. Springer Berlin / Heidelberg, 2003.

17. Alexander Maedche and Steffen Staab. Ontology learning for the semantic web. *IEEE Intelligent Systems*, 16:72–79, 2001.

18. C. OKeefe, A. Goodchild, P. Greenfield, A. Waugh, E. Cheung, and D. Austin. Implementation of electronic consent mechanisms. *Final Analysis Paper*, 2002.

19. Federica Paganelli and Dino Giuli. A context-aware service platform to support continuous care networks for home-based assistance. In *Proceedings of the 4th international conference on Universal access in human-computer interaction: ambient interaction*, UAHCI'07, pages 168–177, Berlin, Heidelberg, 2007. Springer-Verlag.

20. Cedric Pruski. e-crl: A rule-based language for expressing patient electronic consent. *eHealth, Telemedicine, and Social Medicine, International Conference on*, 0:141–146, 2010.

21. Sarvapali D. Ramchurn, Dong Huynh, and Nicholas R. Jennings. Trust in multi-agent systems. *Knowl. Eng. Rev.*, 19:1–25, March 2004.

22. Jason Reid, Ian Cheong, Matthew Henricksen, and Jason Smith. A novel use of rbac to protect privacy in distributed health care information systems. In *ACISP'03: Proceedings of the 8th Australasian conference on Information security and privacy*, pages 403–415, Berlin, Heidelberg, 2003. Springer-Verlag.

23. Jos De Roo. Euler proof mechanism. Website. `http://eulersharp.sourceforge.net/`.

24. H. Song, TK Win, and P. Croll. Patient e-consent mechanism: Models and technologies. *Proceedings of CollECTeR, Melbourne, Australia*, 2002.

25. TAC Team, Michael P. Wellman, Peter R. Wurman, Kevin O'Malley, Roshan Bangera, Shou-de Lin, Daniel Reeves, and William E. Walsh. Designing the market game for a trading agent competition. *IEEE Internet Computing*, 5:43–51, March 2001.

26. A. Terauchi and O. Akashi. Trust-based cooperative action control in multi-agent systems for network management. In *Advanced Information Networking and Applications Workshops, 2009. WAINA '09. International Conference on*, pages 278 –283, May 2009.

27. Gerhard Weiss, editor. *Multiagent systems: a modern approach to distributed artificial intelligence*. MIT Press, Cambridge, MA, USA, 1999.

28. Andrew B. Williams. Learning to share meaning in a multi-agent system. *Autonomous Agents and Multi-Agent Systems*, 8:165–193, March 2004.

29. K. Win, H. Song, P. Croll, and J. Cooper. Implementing patient's consent in electronic health record systems. *Proceedings of CollECTeR, Melbourne, Australia*, 2002.

30. Jae Wook Woo, Myung Jin Hwang, Chun Gyeong Lee, and Hee Yong Youn. Dynamic role-based access control with trust-satisfaction and reputation for multi-agent system. In *Advanced Information Networking and Applications Workshops (WAINA), 2010 IEEE 24th International Conference on*, pages 1121 –1126, 4 2010.

31. Youyong Zou, Tim Finin, Li Ding, Harry Chen, and Rong Pan. Using Semantic web technology in Multi-Agent systems: a case study in the TAGA Trading agent environment. In *Proceeding of the 5th International Conference on Electronic Commerce*, pages 95–101, September 2003.

SIMULATION FOR PATIENT DISCHARGE PROCESS REENGINEERING IN HOSPITALS

Sung J. Shim
Seton Hall University, USA
sung.shim@shu.edu

Arun Kumar
RMIT University, Australia
a.kumar@rmit.edu.au

Jianxin (Roger) Jiao
Georgia Tech Savannah, USA
roger.jiao@me.gatech.edu

ABSTRACT

Radio frequency identification (RFID), which uses radio waves to identify physical items, is rapidly supplanting bar code technology as the principal means of identifying objects in a wide variety of applications. While RFID has drawn increasing interest in healthcare, it seems that few have examined specific RFID implementations for patient services. This paper reports on a field study of RFID implementation for patient discharge process reengineering in a hospital. The results will prove helpful to those who consider implementing RFID to improve the patient discharge or other similar processes in hospitals.

INTRODUCTION

We conducted a field study at a hospital (referred to as "the Hospital" for anonymity). The Hospital was in the process of deploying RFID technology to reengineer and improve its patient discharge process. The Hospital seemed to be an ideal site for this field study, since it provided data for us to compare the current process without RFID and the reengineered process with RFID. By comparing the 'before' and 'after' processes, we could evaluate the impact of RFID implementation on the process.

THE 'BEFORE' PROCESS

The Hospital management was concerned about the lengthy and inefficient process for discharging patients. This not only caused frustration and dissatisfaction on the part of patients

but also led to delays in processing incoming patients and low ratios of bed turnovers. The Hospital management identified two major bottlenecks in the current patient discharge process. The first bottleneck was at the stage of arranging follow-up appointments for patients. The second bottleneck was at the stage of completing medical certificates and prescription forms by doctors. Data from the Hospital showed that the whole discharge process in each ward could take at least two hours.

THE 'AFTER' PROCESS

In order to alleviate the patient discharge process, the Hospital management considered setting up a new discharge lounge, where patients certified for discharge by their doctors are transferred to and wait until the discharge procedures for them are completed. The Hospital management expected to congregate patients into the central discharge lounge and streamline the discharge process. Further, the RFID system was expected to enhance the usability of the discharge lounge by allowing patients to be tracked with less staff and not allowing any patient to leave the Hospital before completing the discharge process.

SIMULATION RESULTS

Using Arena® software, we constructed a simulation model for the patient discharge process at the Hospital and ran the simulation model for both the 'before' and 'after' processes. The simulation results showed that the 'after' process could decrease patient wait times as well as staff service times in the process, and that it could increase ratios of bed turnovers as well as utilization of staff in the process.

Automated Reporting and Data Mining: In Support of Health-Care Quality Management

Patrick Samedy
NYC Health and Hospitals

Farrokh Guiahi
Zarb School of Business
Hofstra University

ABSTRACT

As a result of new government policy supporting the acceleration of electronic medical record (EMR) adoption, there will be an ever-increasing amount of clinical information captured along the health-care continuum. Accordingly, there is a need to create methods of integrated information delivery that are well suited to the timely and accurate delivery of clinical value through automated reporting and data mining. The immediate objective of this paper is to describe the methods, experiences and outcomes related to development of a reporting solution aimed at supporting quality management in the health-care setting. It includes a data structure that describes patient encounters in terms of quality performance measures and reporting and data mining tools used to deliver clinical intelligence. It was found that the automation of quality reporting streamlined the medical record review process and facilitated quality compliance improvements of up to 19%.

SIX SIGMA IN HEALTHCARE DELIVERY

Matthew J. Liberatore, Villanova University, 610-519-4390, matthew.liberatore@villanova.edu

ABSTRACT

This paper reviews and assesses the extant literature on the application of six sigma in health care delivery, focusing on the areas of application, process changes initiated and outcomes, including improvements in process metrics, cost, and revenue. The reported six sigma applications were classified using a two-dimensional framework: area of application within health care delivery and key process metrics improved. The findings suggest that although six sigma has been effective in improving health care delivery, more emphasis needs to be placed on improving the process of identifying and evaluating alternatives, and verifying that the changes implemented offer significant and sustainable improvements.

INTRODUCTION

The United States continues to devote substantial amounts of its resources to health care. U.S. health care spending growth decelerated in 2008, increasing 4.4 percent compared to 6.0 percent in 2007. Hospital spending growth increased 4.5 percent to $718.4 billion compared to 5.9 percent growth in 2007. Health spending growth for state and local and private sources of funds also slowed while federal health spending growth accelerated in 2008. Total health expenditures reached $2.3 trillion in 2008, which translates to $7,681 per person and an increase to 16.2 percent of the nation's GDP or Gross Domestic Product [132].

While health care spending continues to rise, so do concerns about health care quality. A major impetus toward recognizing the need to improve health care quality and patient safety occurred when the Institute of Medicine (IOM) released a report in November 1999 estimating that as many as 98,000 patients die as the result of medical errors in hospitals each year [65]. Through process and quality improvement efforts, the quality of health care for millions of Americans improved in 2007 but significant variations in performance continue to leave many people receiving substandard care [86].

The IOM report and the ongoing interest in improving operational cost and quality led a number of authors to recommend the application of six sigma to health care in the US and elsewhere [6] [21] [48] [60] [64] [68] [83] [92] [127]. Carrigan and Kujawa [16] states that six sigma is an effective strategy tool that can be used to establish and sustain competitive advantage and facilitate achievement of long-term strategic goals in health care. Physician participation is critical, and strategies to support their engagement are discussed in [42].

Some authors suggest that Six Sigma grew out of the total quality management (TQM) movement. TQM had a number of shortcomings, including not providing evidence of better patient outcomes, increased satisfaction, or improved financials. These factors, along with its inability to remove root causes of problems and demonstrate a strategic importance, led to TQM's eventual decline. In contrast, Six Sigma offers time and money deliverables; the sigma

metric to indicate the current state of process, outcome, or service quality; and a focus on improving the "critical to quality" (CTQ) characteristics vital to internal and external customers [11]. Woodard [143] describes evolution of quality control methods and compares them to six sigma, and ultimately advocates for six sigma. However, Landek [76] argues that six sigma is a useful tool that may not be effective in hospitals because of the cash and resources required. Neff (2003) states that six sigma can be overwhelming in scope if not broken down into manageable pieces, and requires significant investment in time.

The purpose of this paper is to review and assess the extant literature on the application of six sigma in health care delivery, focusing on the areas of application, process changes initiated, and outcomes, including improvements in process metrics, cost, and revenue. Those areas that have seen the most successful application are identified, and suggestions for other application areas and improved usage of six sigma are discussed.

OVERVIEW OF SIX SIGMA

Six sigma is credited with helping Motorola win the Malcolm Baldridge Award in 1988. Six sigma is a process improvement goal that was developed by Motorola in the early 1980s and subsequently has been adopted by many organizations. Traditionally a capable process was one in which its natural variation of plus or minus three standard deviations, or sigma, from the mean was less than the target specifications. Under the assumption of normality, this translates to a process yield of 99.73 percent. Motorola's Six Sigma asks that processes operate such that the nearest target specification is at least plus or minus six sigma from the process mean. This translates into an error rate of 2 parts per billion. Often, an error rate of 3.4 parts per million is associated with Six Sigma quality, under the assumption that the process mean can shift 1.5 standard deviations on either side of the mean [80].

Six Sigma projects are undertaken to improve the process of interest, focusing on the CTQ. A structured approach is used to uncover the root cause of problems using the DMAIC (Define-Measure-Analyze-Improve-Control) methodology:
Define the problem within a process
Measure the defects
Analyze the cause of defects
Improve the process performance to remove causes of defects
Control the process to make sure defects do not recur.

 DMAIC is a data-driven process that uses various quality and process improvement tools that have been developed over time, including: statistical analysis, cause and effect diagrams (fishbone, Ishikawa), control charts, design of experiments, Pareto Analysis, process mapping, Failure Modes and Effects Analysis (FMEA), Quality Function Deployment (QFD)/House of Quality, and Suppliers, Inputs, Process, Outputs, and Customers (SIPOC diagrams), among others.

Organizations that use Six Sigma emphasize employee participation and training through three levels:
Green belts: individuals that have completed basic training and participate in Six Sigma projects

Black belts: individuals competent to serve as on-site consultants and lead project teams
Master black belts: individuals who have mastered the Six Sigma process and are capable of teaching it to others and acting as resources for project teams [35] [93].

Six sigma is sometimes combined with lean management, which is based on the principles of the Toyota Production System [141] [142], and the resulting method is called lean six sigma. A kaizen event may be used, which is a focused, intensive, short-term project targeted to improve a process. Jacobson and Johnson [67] argue that the combined implementation of lean and six sigma drives effective results in healthcare, when DMAIC and the lean principles of speed, efficiency, and immediate action are applied. Daley [28] addresses the most common misconceptions regarding lean six sigma using several mini cases to provide anecdotal evidence.

LITERATURE REVIEW

There are some studies that have attempted to assess the implementation of six sigma in health care. Martin and McLennan [82] surveyed health care organizations and found that six sigma was the most common approach utilized by nearly one in five (18.5%) of the respondents followed closely by lean processes (13.3 %). Antony et al. [6] include a summary of outcomes and financial savings from ten health care organizations having six sigma programs. Of the ten firms, six were able to estimate cost savings and/or revenue increases. Revere et al. [105] provide a summary of some six sigma applications, while Gras and Philippe [54] review the application of six sigma in some clinical laboratories. All of the specific applications mentioned in these reviews are included in our application review if sufficient information was found in either these or other articles.

RESEARCH METHODOLOGY

To identify those journal articles that describe the application of six sigma in healthcare, an extensive literature search was conducted. The research process used the keyword "six sigma," in combination with the keywords "healthcare," "health care," and "hospital." We searched Pub Med, ABI/Informs Proquest (business), Compendex (engineering), CINAL (The Cumulative Index to Nursing and Allied Health Literature), and PsycINFO using these keywords.

The topics of the articles that were uncovered in the database searches were screened to determine if the six sigma methodology had either been recommended for application or actually applied in a health care context. Our search excluded conference proceedings and doctoral dissertations since we assume that important research will eventually appear in academic or professional journals. We also exclude non-English language publications from our search. Applications were included in this review if they provide sufficient details concerning the study approach, process changes initiated, and the results achieved, often including improvements in metrics.

RESULTS

Ninety-seven six sigma applications were identified, and address a variety of application within health care delivery, as shown in Table 1. At the highest level, area of application was defined to

include inpatient care, therapeutic support, emergency care, ambulatory care, and administrative. Inpatient care includes the processes associated with the major flow of patients through a hospital or medical center, including admission, medical/surgical, critical/intensive care, hotel services, and discharge. For example, hospital admission and pre-registration processes and the time required to obtain subsequent treatment are areas that have been addressed using six sigma. These studies indicate that six sigma has been effective in improving admission-related processes, and is an area of application that might be considered by more hospitals.

Table 1: Six Sigma Applications in Health Care Delivery

Inpatient
   Admissions
- Registration process
- Pre-registration system
- Bed assignments for patients needing to be isolated
- Development of kidney transplant list

   Medical/Surgical
- Catheters: infection, UTIs, groin injuries, insertion
- Cardiac surgery infections
- Insulin, heparin, breast milk administration
- Correct blood transfusions
- Double stapling technique (surgical procedure)
- Incidence of falls pressure
- Medication errors, timing, compliance, usage
- CHF, COPD, Ventilator, Delivery Room LOS
- Surgery, OR TAT & productivity
- Pain management, nursing efficiency

   Intensive/critical care
- Ventilator days spent, incidence of pneumonia
- Transfer time to regular patient unit
- Compliance with hand hygiene regulations

   Hotel services/discharge/other
- TAT for bed assignment
- Time to complete patient discharge: routine, SNF, Medicare
- Centralized equipment delivery
- Removal of old inventory
- Unresolved work orders

Therapeutic Support
   Laboratory
- TAT for ED laboratory orders
- TAT for MCI test

- Pneumatic tube system wait/travel time
- Laboratory requisition errors
- Phlebotomists' efficiency

Diagnostics/Radiology
- Defect rate of X-ray films
- Communicating findings without defects
- Test wait time
- Report creation time
- Stress test TAT
- Mammography screening cycle time
- Radiology scheduling process
- CT capacity

Pharmacy
- Medication safety, errors
- In-hospital drug sales
- Pharmacy call backs to physicians

Emergency Care
- Physician turnover
- Reduce patient "walk out" rate: improve flow; add capacity, reduce staff turnover
- Patient flow: throughput rate, LOS
- Door-to-doctor time
- Patient wait time for treatment
- Patient wait time for a bed

Ambulatory Care
Ambulance
- Best hospital destination, response time

Home health
- Prospective payment system – process automation
- reduce low utilization payment adjustments
- use of telehealth device

Outpatient/Surgical Clinics
- LOS for post-anesthesia and ambulatory surgery
- Medicine dispensing errors
- Patient access to OB/GYN clinic
- Delays in starting treatments for oncology patients
- Wait time from lab order placement to specimen collection
- Reduce rate of follow-up to new patient in Genitourinary Medicine clinics

Administrative

    Billing/Employee Management

- Billing accuracy
- Number of invoice mistakes from temporary employment agencies
- Employee recruitment process time
- Employee vaccination rate
- Develop tool for staff effectiveness

Figure 1 shows the number of reported six sigma applications over time. Note that the initial applications coincided with the publication of the IOM report. After a period a rapid growth, the number of reported six sigma applications has remained steady since 2004, the year when the first lean six sigma application appeared. Interestingly, the number of reported lean six sigma applications has remained at a steady level since 2005, perhaps indicating that those health care organizations that are adopting lean principles are not necessarily combining their efforts with six sigma.

Figure 1: Reported Six Sigma Applications in Health Care Delivery Over Time
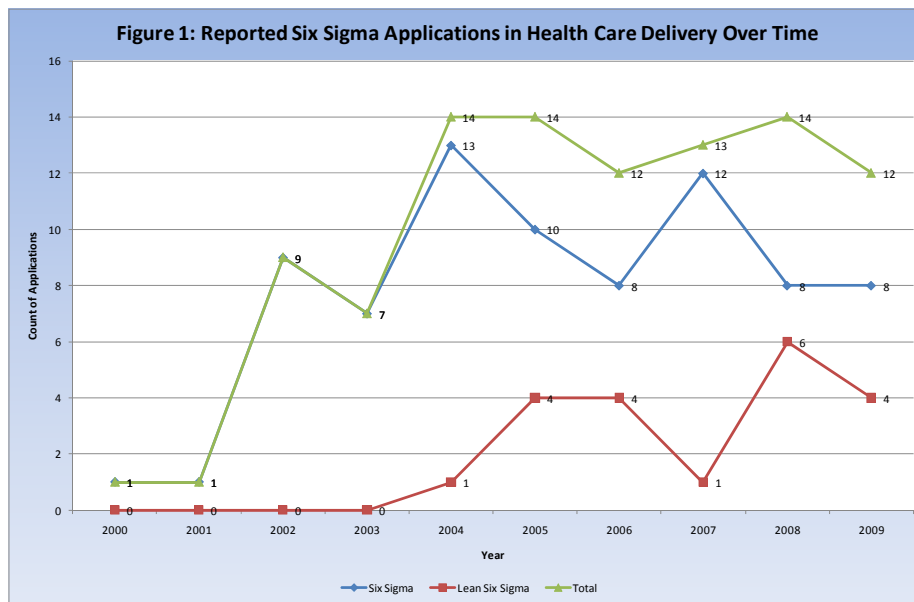
ANALYSIS

To obtain a better understanding of the extent of successful Six Sigma implementation, the reported applications were classified using a two-dimensional framework: area of application within health care delivery (as described above) and key process metrics improved (Table 2). The second dimension, process metric, includes defect rate, medication error, process time, compliance rate, and productivity. We note whether data on the level of improvement of the metric(s) are provided, whether cost and revenue improvements were obtained, and whether the application is six sigma or lean six sigma.

Focusing first on the rows in Table 2, we see that of the 97 applications reviewed, the health care delivery areas receiving the most attention are medical/surgical (33), diagnostics (14), and emergency care (11). These applications offer useful information and guidance to other health

care organizations facing similar problems. A limited number of applications were found in inpatient areas such as admission, discharge, and hotel services, which should be amenable to six sigma process and quality improvement efforts. Also, limited applications were found in ambulatory care areas such as surgery and clinics, which face issues similar to those in the medical/surgical area. In addition, there were no reported applications in other support areas within the hospital such as respiratory therapy and rehabilitation, as well as in physician practices, an important area of primary care. These findings suggest that six sigma can see much more widespread application in health care delivery.

Second, nearly all of the reported applications indicate the level of improvement of the key process metric after implementation. Only about one-third of the applications translate the level of process improvements into cost savings or revenue enhancement to demonstrate value and significance. The value of six sigma applications need to be clearly demonstrated to help maintain commitment to the process changes implemented. Along the same lines, only three of the applications discussed the sustained improvement in the key metrics. The control process is critical, so that the process and quality improvements are maintained.

Six sigma focuses on reducing process variation and errors, so it is clear why over one-third of the reported applications have error rate (defect rate, medication errors, or compliance rate in Table 2) as their driving metric. Of the remaining applications, about one-third focus on process time (e.g., cycle time, TAT, LOS, wait time) and slightly less than one third focused on productivity (e.g., resource utilization, throughput, capacity) metrics. Selection of the key metric is a critical task that directs six sigma process improvement efforts.

CONCLUSIONS

In this paper we have reviewed the reported applications of six sigma in health care delivery and presented a two-dimensional framework that has categorized the applications by area and the metrics improved. Our research demonstrates that many health care organizations have reported the application of six sigma or lean six sigma. Across the reported health care delivery applications, DMAIC is widely used to implement six sigma, and a number of hospitals have utilized consultants, including GE Healthcare Systems. Over time, the number of reported six sigma applications has remained at a steady rate. Hospitals that have not implemented six sigma or lean six sigma can learn from the successful applications discussed in this paper.

Table 2: The Six Sigma Applications in Health Care Delivery Literature Classified Using a Two-Dimensional Framework						
Functional Area	Defect Rate	Medication Error	Process Time	Compliance Rate	Productivity	Total
Inpatient Care						
Admission			[111**],[43*^C], [46**^C]		[115**^R, 58**^C]	4
Medical/ Surgical	[47**, 120], [33*], [84*], [27*, 55*, 90**], [57**], **[59***]**, [116**], [88**], [29], **[137**]**	[109**], [9**], [40*^C], [77, 99, 144], [37]	[1**^R], [109***], [122**^C, 41**, 52**], [98**], [121**], [**10**^C, **135**^C, **134**^C], [95**]	[125^C], [25*], [102***]	[135**^CR, 30^C,134], [32], **[44**]**, [50**^C], [111**], **[73**]**, **[140**]**, [34**]	33
Intensive/ Critical	[118**]		[4**], [69**^C]	[38**]		4
Hotel Services			[100**]			1
Discharge			[123**], [5*]		[12**^C]	3
Other	[110**]				**[30**^C], [78^C, 117**]	3
Therapeutic Support						
Laboratory			[75**], [91**^C], [117**]		**[106**]**, **[128**^C]	5
Diagnostics	[23**^C], [53**]		[104**^C], [18*], [51**], [138**]		[79**^R, 119**^R], [7**^C], [19**], [71*], [139**], [117**, 26*^C], [8, 39*^C]	14
Pharmacy		[97**], [17*]	[85**]		[61**^C]	4
Emergency Care						
Emergency Care	[63**^C], [129**^R], **[113**]**		[81**], [124**], [45**], [111*^C], **[24**]**, [74], [96*], [78**^R]			11
Ambulatory Care						
Ambulance Operations					[126**^C]	1
Home Health	[37^C], [82**^C]				[108***]	3
Surgical			[62**]			1
Clinic		[19*]	[14**^R], **[66]**		[2**, 56**, 3], [107**, 26**]	5
Administrative						
Billing	[117**^C], **[30**^C]					2
Employee Management			[117**]	[72**], [91*]		3

Legend: []articles in brackets relate to the same application, *beginning or ending metric, **beginning and ending metirc, ***beginning, ending and sustained metric, ^C cost savings,

^R revenue generated, **bold** Lean Six Sigma application

For those applications applying DMAIC, statistical analysis is sometimes used to identify the sources of process variation. Process and other changes typically are developed in response to the major drivers of variation, but are sometimes implemented without statistical or other proof that the change is valid. On the other hand, in some cases we have sufficient statistical power so that even miniscule differences of no practical value will be seen as significant. More attention needs to be directed to make certain that the changes implemented are of statistical and practice significance.

Generally, mathematical, computer, or statistical modeling are not used to help identify the best course of action, or to predict the change in behavior before implementation. There are opportunities to increase the use of modeling to improve the analysis phase of DMAIC.

Overall, six sigma has made important contributions in improving health care delivery, and should find increased application, often in conjunction with lean management.

ACKNOWLEDGEMENT

REFERENCES

[1]   Adams, R., P. Warner, B. Hubbard, and T. Goulding, "Decreasing Turnaround Time between General Surgery Cases: A Six Sigma Initiative," *The Journal of Nursing Administration,* 34(3), 2004, 140-8.

[2]   Ahmed-Jushuf, I., and V. Griffiths, A., "Identifying Hidden Capacity Through Modernization of Genitourinary Medicine Services,*" International Journal of STD and AIDS,* 18(5), 2007, 297-8.

[3]   Ahmed-Jushuf, I., and V. Griffiths. B., "Reducing Follow-Ups: An Opportunity to Increase the Capacity of Genitourinary Medicine Services across the UK," *International Journal of STD and AIDS,* 18(5), 2007, 305-7.

[4]   Albert, K., and D. Powers, "The Six Sigma Solution: This Data-Driven Approach to Consistent Patient Outcomes Improves Hospital Performance and Reduces Medical Error," *Nursing Spectrum,* 17(22), 2007, 26-7.

[5]   Allen, T. T., S. Tseng, K. Swanson, and M. A. McClay, "Improving the Hospital Discharge Process with Six Sigma Methods," *Quality Engineering,* 22(1), 2010, 13-20.

[6]   Antony, J., K. Downey-Ennis, F. Antony, and C. Seow, "Can Six Sigma Be the 'Cure' for our 'Ailing' NHS?," *Leadership in Health Services,* 20(4), 2007, 242-53.

[7]   Bahensky, J. A., J. Roe, and R. Bolton, "Lean Sigma--Will It Work for Healthcare?," *Journal of Healthcare Information Management: JHIM,* 19(1), 2005, 39-44.

[8]   Benedetto, A. R., "Adapting Manufacturing-Based Six Sigma Methodology to the Service Environment of a Radiology Film Library," *Journal of Healthcare Management* 48(4), 2003, 263-80.

[9]   Benitez, Y., L. Forrester, C. Hurst, and D. Turpin, "Hospital Reduces Medication Errors Using DMAIC and QFD," *Quality Progress,* 40(1), 2007, 38-45.

[10]   Bisgaard, S., and R. Does, "Quality Quandaries: Health Care Quality—Reducing the Length of Stay at a Hospital," *Quality Engineering,* 21(1), 2009, 117-31.

[11]   Black, K., and L. Revere, "Six Sigma Arises from the Ashes of TQM with a Twist," *International Journal of Health Care Quality Assurance,* 19(3), 2006, 259-66.

[12]   Burmahl, B., "Patient Care: Improved Discharge Planning Results in Bottom-Line Savings," *Hospital & Health Networks,* 82(8), 2008, 17.

[13]   Burrington-Brown, J., "Six Sigma Makes Sense as a Quality Initiative," *Journal of AHIMA /American Health Information Management Association,* 73(5), 2002, 56-7.

[14]   Bush, S. H., M. R. Lao, K. L. Simmons, J. H. Goode, S. A. Cunningham, and B. C. Calhoun, "Patient Access and Clinical Efficiency Improvement in a Resident Hospital-Based Women's Medicine Center Clinic," *The American Journal of Managed Care,* 13(12), 2007, 686-90.

[15]   Caldwell, C., J. Brexler, and T. Gillem, "Engaging Physicians in Lean Six Sigma," *Quality Progress,* 38(11), 2006, 42-6.

[16]   Carrigan, M. D., and D. Kujawa, "Six Sigma in Health Care Management and Strategy," *The Health Care Manager,* 25(2), 2006, 133-41.

[17]   Castle, L., E. Franzblau-Isaac, and J. Paulsen, "Using Six Sigma to Reduce Medication Errors in a Home-Delivery Pharmacy Service," *Joint Commission Journal on Quality and Patient Safety,* 31(6), 2005, 319-24.

[18]   Cavagna, E., R. Berletti, F. Schiavon, B. Scarsi, and G. Barbato, "Optimized Delivery Radiological Reports: Applying Six Sigma Methodology to a Radiology Department," *La Radiologica Medica*,105(3), 2003, 205-14.

[19]   Chan, A. L., "Use of Six Sigma to Improve Pharmacist Dispensing Errors at an Outpatient Clinic," *American Journal of Medical Quality,* 19(3), 2004, 128-31.

[20] Chan, W. P., W. T. Chiu, W. M. Chen, M. F. Lin, and B. Chu, "Applying Six Sigma Methodology to Maximise Magnetic Resonance Imaging Capacity in a Hospital," *International Journal of Healthcare Technology and Management,* 6(3), 2005, 321-30.

[21] Chassin, M. R., "Is Health Care Ready for Six Sigma Quality?," *The Milbank Quarterly* , 76(4), 1998, 565-91.

[22] Chassin, R., "The Six Sigma Initiative at Mount Sinai Medical Center," *Mount Sinai Journal of Medicine,* 75(1), 2008, 45-52.

[23] Chen, Y. K., J. Lin, and C. C. Chang, "Taiwan: Improving Radiography through Application of Six Sigma Techniques," *Journal for Healthcare Quality,* 27(3), 2005, 44-52.

[24] Christianson, J. B., L. H. Warrick, R. Howard, and J. Vollum, "Deploying Six Sigma in a Health Care System as a Work in Progress,*" Joint Commission Journal on Quality and Patient Safety,* 31(11), 2005, 603-13.

[25] Christopher D. A., A. Campbell, and J. Dateshidze, "Working Smarter with Intelligent Pumps," *Nursing Management Suppl.*, 2008, 9-12.

[26] Corn, J. B., "Six Sigma in Health Care,*" Radiologic Technology,* 81(1), 2009, 92-5.

[27] Courtney, B. A., J. B. Ruppman, and H. M. Cooper, "Save Our Skin: Initiative Cuts Pressure Ulcer Incidence in Half," *Nursing Management,* 37(4), 2006, 36-45.

[28] Daley, A. T., "Pro: Lean Six Sigma Revolutionizing Health Care of Tomorrow," *Clinical Leadership & Management Review: The Journal of CLMA,* 20(5), 2006, E2.

[29] Daniels, S., "Sharp Healthcare Gets to the Point in Managing Diabetes," *Quality Progress,* 41(4): 2008, 45-7.

[30] de Koning, H., J. P. Verver, J. van den Heuvel, S. Bisgaard, and R. J. Does, "Lean Six Sigma in Healthcare," *Journal for Healthcare Quality,* 28(2), 2006, 4-11.

[31] Dlugacz, Y. D., "Six Sigma Adds New Dimension to Quality Management Processes," *Journal for Healthcare Quality,* 26(5), 2004, 2.

[32] Does, R. J., T. M. Vermaat, J. P. Verver, S. Bisgaard, and J. van den Heuvel, "Reducing Start Time Delays in Operating Rooms," *Journal of Quality Technology* 41(1), 2009, 95-109.

[33] Drenckpohl, D., L. Bowers, and H. Cooper, "Use of the Six Sigma Methodology to Reduce Incidence of Breast Milk Administration Errors in the NICU," *Neonatal Network* 26(3), 2007, 161-6.

[34] DuPree, E., L. Martin, R. Anderson, N. Kathuria, D. Reich, C. Porter, and R. M. Chassin, "Improving Patient Satisfaction with Pain Management Using Six Sigma Tools," *Joint Commission Journal on Quality and Patient Safety,* 35(7), 2009, 343-350.

[35] Eckes, G., *Making Six Sigma Last: Managing the Balance Between Cultural and Technical Change,* 2001, New York, Wiley.

[36] Elberfeld, A., K. Goodman, and M. Van Kooy, "Using the Six Sigma Approach to Meet Quality Standards for Cardiac Medication Administration," *Journal of Clinical Outcomes Management,* 11(8), 2004, 510-6.

[37] Elberfeld, A., S. Bennis, J. Ritzius, and D. Yhlen, "The Innovative Use of Six Sigma in Home Care," *Home Healthcare Nurse,* 25(1), 2007, 25-33.

[38] Eldridge, N. E., S. S. Woods, R. S. Bonello, K. Clutter, L. Ellingson, M. A. Harris, B. K. Livingston, J. P. Bagian, L. H. Danko, E. J. Dunn, R. L. Parlier, C. Pederson, K. J. Reichling, G. A. Roselle, and S. M. Wright, "Using the Six Sigma Process to Implement the Centers for Disease Control and Prevention Guideline for Hand Hygiene in 4 Intensive Care Units," *Journal of General Internal Medicine,* 21(Suppl 2), 2006, S35-42.

[39] Elsberry, R. B., "Six Sigma: Applying a Corporate Model to Radiology," *Decisions in Imaging Economics,* 13(7), 2000, 56-66.

[40] Esimai, G., "Lean Six Sigma Reduces Medication Errors," *Quality Progress,* 38(4), 2005, 51-7.

[41] Ettinger, W. H., "Six Sigma: Adapting GE's Lessons to Health Care," *Trustee,* 54(8), 2001, 10-5.

[42] Ettinger, W. H., and M. Van Kooy, "The Art and Science of Winning Physician Support for Six Sigma Change," *Physician Executive,* 29(5), 2003, 34-8.

[43] "Facility Uses Six Sigma to Improve Quality," *Healthcare Benchmarks and Quality Improvement,* 13(9), 2006, 104-5.

[44] Fairbanks, C. B., "Using Six Sigma and Lean Methodologies to Improve OR Throughput," *AORN Journal,* 86(1), 2007, 73-82.

[45] "Focus on Process Slashes Average Cycle Time by 37%," *ED Management,* 16(11), 2004, 127-8.

[46] Franchetti, M., and K. Bedal, "Perfect Match," *Six Sigma Forum Magazine,* 8(4), 2009, 10-7.

[47]   Frankel, H. L., W. B. Crede, J. E. Topal, S. A. Roumanis, M. W. Devlin, and A. B. Foley, "Use of Corporate Six Sigma Performance-Improvement Strategies to Reduce Incidence of Catheter-Related Bloodstream Infections in a Surgical ICU," *Journal of the American College of Surgeons,* 201(3), 2005, 349-58.

[48]   Frings, G. W., and L. Grant, "Who Moved My Sigma Effective Implementation of the Six Sigma Methodology to Hospitals," *Quality and Reliability Engineering International,* 21(3), 2005, 311-28.

[49]   Gale, S. F., "Building Frameworks for Six Sigma Success," *Workforce,* 82(5), 2003, 64-69.

[50]   Geriaine, J., "Six Sigma Plan Delivers Stellar Results," *Materials Management in Healthcare ,* 16(4), 2007, 20-6.

[51]   Godin, E., D. Raven, C. Sweetapple, and F. R. Del Guidice, "Faster Test Results," *Quality Progress,* 37(1), 2004, 33-9.

[52]   Goedert, J., "Crunching Data: The Key to Six Sigma Success," *Health Data Management,* 12(4), 2004, 44-8.

[53]   Gorman, A., L. Donnell, H. Hepp, and T. Mack, "Improving Communication and Documentation Concerning Preliminary and Final Radiology Reports," *Journal for Healthcare Quality,* 29(2), 2007, 13-21.

[54]   Gras, J. M., and M. Philippe, "Application of the Six Sigma Concept in Clinical Laboratories: A Review," *Clinical Chemistry and Laboratory Medicine,* 45(6), 2007, 789-96.

[55]   Griffin, B., H. Cooper, C. Horack, M. Klyber, and D. Schimmelpfenning, "Best-Practice Protocols: Reducing Harm from Pressure Ulcers," *Nursing Management,* 38(9), 2007, 29-32.

[56]   Griffiths, V., and I. Ahmed-Jushuf, "Releasing Capacity in Sexual Health through Modernization," *International Journal of STD and AIDS,* 18(5), 2007, 299-304.

[57]   Guinane, S., and N. Davis, "The Science of Six Sigma in Hospitals," *American Heart Hospital Journal,* 2(1), 2004, 42-8.

[58]   Hagland, M., "Six Sigma Practices. A Strategy Based on Data Is Perfect Fit for Healthcare," *Healthcare Informatics,* 23(1), 2006, 27-8.

[59]   Hansen, B., "Reducing Nosocomial Urinary Tract Infections through Process Improvement," *Journal for Healthcare Quality,* 28(2), 2006, W2-9.

[60]   Harrington, H. J., and B. Trusko, "Six Sigma: An Aspirin for Health Care,"
         *International Journal of Health Care Quality Assurance Incorporating Leadership in
         Health Services,* 18(6-7), 2005. 487-515.

[61]   Higgins, T. J., J. C. Diener, T. Grunsfeld, and S. Wogen, "Aim for Perfection.
         Evaluate E-Prescribing Technology using Six-Sigma Techniques," *MGMA Connexion*
         2(9), 2002, 35-8.

[62]   "Hospital Cuts 30 Minutes Off Ambulatory Surgery LOS," *HealthCare Benchmarks
         and Quality Improvement,* 2009.
         http://findarticles.com/p/articles/mi_m0NUZ/is_2009_Jan_1/ai_n35534783/ . Retrieved
         July 20, 2010.

[63]   "How Much Does Poor Quality Cost Your ED?," *ED Management*, 18(11), 2006, 126-7.

[64]   Impellizzeri, F. M., M. Bizzini, M. Leunig, N. A. Maffiuletti, and A. F. Mannion,
         "Money Matters: Exploiting the Data from Outcomes Research for Quality Improvement
         Initiatives," *European Spine Journal,* 18(Suppl 3), 2009, 348-59.

[65]   Institute of Medicine, To Err Is Human: Building a Safer Health System, 1999.
         http://www.nap.edu/openbook.php?record_id=9728&page=1 . Retrieved July 18, 2010.

[66]   Jackson, J., and L. A. Woeste, "Using Lean Six Sigma to Reduce Patient Wait
         Times," *Labmedicine,* 39(3), 2008, 134-6.

[67]   Jacobson, J. M, and M. E. Johnson, "Lean and Six Sigma: Not for Amateurs,"
         Labmedicine, 37(3), 2006, 140-5.

[68]   James, C., "Manufacturing's Prescription for Improving Healthcare Quality,"
         Hospital Topics, 83(1), 2005, 2-8.

[69]   Jennings, A., "Bringing Six Sigma to the ICU Bedside," *RT: For Decision Makers
         in Respiratory Care*, 2006. http://www.rtmagazine.com/issues/articles/2006-06_09.asp .
         Retrieved July 20, 2010.

[70]   Johnson, T., G. Currie, P. Keill, S. J. Corwin, H. Pardes, and M. R. Cooper,
         "NewYork-Presbyterian Hospital: Translating Innovation into Practice," *Joint
         Commission Journal on Quality and Patient Safety,* 31(10), 2005, 554-60.

[71]   Kang, J. O., M. H. Kim, S. E. Hong, J. H. Jung, and M. J. Song, "The Application
         of the Six Sigma Program for the Quality Management of the PACS," *American Journal
         of Roentgenology,* 185(1), 2005, 1361-5.

[72]   Kaplan, S., S. Bisgaard, D. Truesdell, and S. Zetterholm, "Design for Six Sigma in Healthcare: Developing an Employee Influenza Vaccination Process," *Journal for Healthcare Quality,* 31(3), 2009, 36-43.

[73]   Kemper, B. P., M. Koopmans, and R. J. Does, "Quality Quandaries: The Availability of Infusion Pumps in a Hospital," *Quality Engineering,* 21(4), 2009, 471-7.

[74]   Kornecki, Z., "What Gets Measured Gets Managed," *Health Management Technology,* 29(2), 2008, 24-8.

[75]   "Lab Order to Results in 16 Minutes? You Heard Right! Six Sigma Helps ED Overcome Internal Squabbling," *ED Management,* 16(1), 2004, 89-90.

[76]   Landek, D., "Con: Six Sigma Not Always the Right Answer in the Clinical Laboratory," *Clinical Leadership & Management Review: the Journal of CLMA,* 20(5), 2006, E3.

[77]   Lanham, B., and P. Maxson-Cooper, "Is Six Sigma the Answer for Nursing to Reduce Medical Errors and Enhance Patient Safety?," *Nursing Economics,* 21(1), 2003, 39-41.

[78]   Lazarus, I. R., and B. Stamps, "The Promise of Six Sigma: Part 2," *Managed Healthcare Executive,* 12(1), 2002, 27-30.

[79]   Loree, A., M. Maihack, and M. Powel, "The Path of Least Resistance: Is There a Better Route?," *Radiology Management,* 25(5), 2003, 48-51.

[80]   Lucas, J. M., "The Essential Six Sigma," *Quality Progress,* 35(1), 2002, 27-31.

[81]   "Make These Changes to Cut Delays, Diversion Hours," *ED Management,* 15(6), 2003, 64-5.

[82]   Martin, N. C., and S. L. McLennan, "Six Sigma: Customer Driven Quality Improvement with the Bottom Line in Mind," *Caring,* 24(10), 2005, 38-40.

[83]   Merry, M. D., "Healthcare's Need for Revolutionary Change," *Quality Progress,* 36(9), 2003, 31-5.

[84]   Moorman, D., "On the Quest for Six Sigma," *American Journal of Surgery*, 189(3), 2005, 253-8.

[85]   Mukherjee, S., "A Dose of DMAIC," *Quality Progress,* 41(8), 2008, 44-51.

[86]   National Committee for Quality Assurance, "The State of Health Care Quality 2008." available at: http://www.ncqa.org/portals/0/newsroom/sohc/SOHC_08.pdf Accessed 23 July 2010.

[87]   Neff, M., "Demystifying 'QI,'" *Michigan Health and Hospitals,* 39(4), 2003, 28.

[88]   Neri, R. A., C. E. Mason, and L. A. Demko, "Application of Six Sigma/CAP Methodology: Controlling Blood-Product Utilization and Costs," *Journal of Healthcare Management,* 53(3), 2008, 183-95.

[89]   "New York Hospital Looks to Six Sigma for Culture Change," *Performance Improvement Advisor,* 7(11), 2003, 141-3.

[90]   Nimtz-Rusch K., and J. Thompson, "Nursing and Six Sigma: A Perfect Match for Quality and Safety Improvement," *Chart,* 105(3), 2008, 10-3.

[91]   Novicoff, W. M., and I. R. Lazarus, "Six Sigma Enters the Healthcare Mainstream: Performance-Improvement Technique Has Proven its Worth and Staying Power," *Managed Healthcare Executive,* 2004. http://managedhealthcareexecutive.modernmedicine.com/mhe/article/articleDetail.jsp?id=81361 Retrieved July 20, 2010.

[92]   O'Kane, M. E., "Roadmaps for Building a Better System," *Frontiers of Health Service Management,* 20(3), 2004, 25-8.

[93]   Pande, P. S., R. P. Neuman, and R. R. Cavanagh, *The Six Sigma Way: How GE, Motorola, and Other Top Companies are Honing Their Performance.* 2000, New York, McGraw-Hill.

[94]   "Paperless Hospital Uses Six Sigma to Improve Safety," *Quality Progress,* 37(9), 2004, 17.

[95]   Parker, B. M., J. M. Henderson, S. Vitagliano, B. G. Nair, J. Petre, W. G. Maurer, M. F., M. Weber, L. DeWitt, J. Beedlow, B. Fahey, A. Calvert, K. Ribar, and S. Gordon, "Six Sigma Methodology Can Be Used to Improve Adherence for Antibiotic Prophylaxis in Patients Undergoing Noncardiac Surgery," *Anestesia & Analgesia,* 104(1), 2007, 140-6.

[96]   Parks, J. K., J. Klein, H. L. Frankel, R. S. Friese, and S. Shafi, "Dissecting Delays in Trauma Care Using Corporate Lean Six Sigma Methodology," *The Journal of Trauma* 65(5), 2008, 1098-104.

[97]   "PASTE Plus Teams at Core of Award-Winning Program," *Healthcare Benchmarks and Quality Improvement,* 11(11), 2004, 125-7.

[98]   "Patient Satisfaction Planner. Six Sigma Project Slashes LOS for SNF Discharges. Initiative Targets Ambulation, Early Assessment," *Hospital Peer Review,* 32(12), 2007 167-9.

[99]   Pelletier, L., "Beth Lanham on Six Sigma in Healthcare. Interview by Luc R. Pelletier," *Journal for Healthcare Quality,* 25(2), 2003, 26-37.

[100]   Pellicone, A., and M. Martocci, "Faster Turnaround Time," *Quality Progress,* 39(3): 2006, 31-6.

[101]   Peltokorpi, A., and J. Kujala, "Time-Based Analysis of Total Cost of Patient Episodes: A Case Study of Hip Replacement," *International Journal of Health Care* 19(2-3), 2006, 136-45.

[102]   Perkins, J., R. Ambroson, and K. Kinsey, "Bring 'Order' to Diabetes Management," *Nursing Management,* 39(5), 2008, 24-30.

[103]   Proudlove, N., C. Moxham, and R. Boaden, "Lessons for Lean In Healthcare from Using Six Sigma in the NHS," *Public Money Management,* 28(1), 2008, 27-34.

[104]   "Radiology Lab Speeds Throughput with Six Sigma; Changes Include Shift in Personnel Positioning," *HealthCare Benchmarks and Quality Improvement,* 15(5), 2008, 44-5.

[105]   Revere, L., K. Black, and A. Huq, "Integrating Six Sigma and CQI for Improving Patient Care," *The TQM Magazine,* 16(2), 2004, 105-13.

[106]   Riebling, N. B., and L. Tria, "Laboratory Toolbox for Process Improvement: Six Sigma at North Shore-Long Island Jewish Hospital," *Labmedicine,* 39(1), 2008, 7-14.

[107]   Roesler, K., and D. Dydyk, "The Oncology Service Line's Use of Six Sigma in the Cowdery Patient Care Center: Creating Standardized Processes to Improve Patient Flow, Improving Staff and Patient Satisfaction," *Oncology Nursing Forum,* 34(2), 2007, 524.

[108]   Russell, K. G., and D. Cempellin, "Using Six Sigma to Improve Home Health Agency Operations," *Caring,* 27(9), 2008, 40-3.

[109]   Sagy, M., "Optimizing Patient Care Processes in a Children's Hospital Using Six Sigma," *Journal of Clinical Outcomes Management,* 16(9), 2009, 412-14.

[110]   San, D.C., "A Six Sigma Approach to Meeting Patient Demand. How to Optimize Inventory,*" Materials Management in Healthcare,* 15(2), 2006, 20-6.

[111]   Scalise, D., "Six Sigma in Action. Case Studies in Quality Put Theory into Practice," *Hospitals & Health Networks,* 77(5), 2003, 57-62.

[112]   Scalise, D., "Six Sigma. The Quest for Quality," *Hospitals & Health Networks,* 75(12), 2001, 41-6.

[113]   Schooley, J., "No Longer Waiting for Answers," *Quality Progress,* 41(11), 2008, 34-9.

[114]   Sehwail, L., and C. DeYoung, "Six Sigma in Health Care." *International Journal of Health Care Quality Assurance,* 16(4), 2003, 1-5.

[115]   Sherman, J., "Achieving REAL Results with Six Sigma," *Healthcare Executive,* 21(1), 2006, 8-14.

[116]   Shukla, P. J., S. G. Barreto, and M. S. Nadkarni, "Application of Six Sigma Towards Improving Surgical Outcomes," *Hepato-gastroenterology,* 55(82-83), 2008, 311-4.

[117]   Simmons, J. C., "Using Six Sigma to Make a Difference in Health Care Quality," *The Quality Letter for Healthcare Leaders,* 14(4), 2002, 2-10.

[118]   Simmons-Trau, D., P. Cenek, J. Counterman, D. Hockenbury, and L. Litwiller, "Reducing VAP with 6 Sigma," *Nursing Management,* 35(6), 2004, 41-5.

[119]   "Six Sigma Boosts Quality with Ongoing Analysis," *Hospital Peer Review,* 28(2), 2003, 19-22.

[120]   "Six Sigma Effort Paying Dividends for CT Hospital," *Data Strategies and Benchmarks,* 6(1), 2002, 1-6.

[121]   "Six Sigma Gives Leaders Tools for Improving Processes in OR," *OR Manager,* 6(1), 2004, 13-5.

[122]   "Six Sigma Improves Care, Reduces Hospitals' Costs: Projects Reduce Variation, Create Efficiency," *Hospital Case Management,* 2004. http://www.thefreelibrary.com/Six+Sigma+improves+care,+reduces+hospitals'+costs%3A+projects+reduce...-a0117040351 . Retrieved July 20, 2010.

[123]   "Six Sigma Projects Add Efficiency to Discharge Process," *Hospital Case Management,* 16(2), 2008, 23-5.

[124]   "Six Sigma Success Story: Reducing Variations Is Key," *ED Management,* 17(7), 2005, 82-3.

[125]   "Six Sigma Success: 100% Compliance in 3 Months," *Healthcare Benchmarks and Quality Improvement,* 11(5), 2004, 52-5.

[126] "Six Sigma System Saves County EMS Agency Big Bucks: Data-Driven Processes Slash Times & Employee Turnover," *EMS Insider,* 35(12), 2008, 9.

[127] Stankovic, A. K., and E. DiLauri, "Quality Improvements in the Preanalytical Phase: Focus on Urine Specimen Workflow," *Clinics in Laboratory Medicine,* 28(2), 2008, 339-50.

[128] Sunyog, M., "Lean Management and Six-Sigma Yield Big Gains in Hospital's Immediate Response Laboratory. Quality Improvement Techniques Save More Than $400,000," *Clinical Leadership & Management Review: The Journal of CLMA,* 18(5), 2004, 255-8.

[129] Taner, M. T., B. Sezen, "An Application of Six Sigma Methodology to Turnover Intentions in Health Care," *International Journal of Health Care Quality Assurance* 22(3), 2009, 252-65.

[130] Thompson, J., T. Miller, A. Krantz, and T. Thomas, "Six Sigma Engages Physicians in Achieving Strategic Performance Improvement at OSF Saint Francis Medical Center," *Global Business and Organizational Excellence,* 27(3), 2008, 6-20.

[131] Uphoff, M. E., "Six Sigma Contributes to Quality at Nebraska Health System," *Nebraska Nurse,* 36(2), 2003, 5.

[132] U.S. Department of Health and Human Services, Center for Medicare and Medicaid Services, "National Health Expenditures 2008 Highlights," 2010. Available at: http://www.cms.gov/NationalHealthExpendData/downloads/highlights.pdf   Retrieved 18 July 2010.

[133] van den Heuvel, J., A. J. Bogers, R. J. Does, S. L. van Dijk, and M. Berg, "Quality Management: Does It Pay Off?," *Quality Management in Healthcare,* 15(3), 2006, 137-49.

[134] van den Heuvel, J., R. J. Does, A. J. Bogers, and M. Berg, "Implementing Six Sigma in the Netherlands," *Joint Commission Journal on Quality and Patient Safety,* 32(7), 2006, 393-9.

[135] van den Heuvel, J., R. J. Does, and S. Bisgaard, "Dutch Hospital Implements Six Sigma," *Six Sigma Forum Magazine,* 4(2), 2005, 11-4.

[136] Van Kooy, M., L. Edell, and H. M. Scheckner, "Use of Six Sigma to Improve the Safety and Efficacy of Acute Anticoagulation with Heparin," *JCOM,* 9(8), 2002, 445-53.

[137] Veluswamy R., and R. Price, "I've Fallen and I Can't Get Up: Reducing the Risk of Patient Falls," *Physician Executive,* 36(3), 2010, 50-3.

[138] Viau, M., and B. Southern, "Six Sigma and Lean Concepts, A Case Study: Patient Centered Care Model for a Mammography Center," *Radiology Management,* 29(5), 2007, 19-28.

[139] Volland, J., "Quality Intervenes at a Hospital," *Quality Progress,* 38(2), 2005, 57-62.

[140] Wijma, J., A. Tripa, R.J. Does, and S. Bisgaard,"Quality Quandaries: Efficiency Improvement in a Nursing Department," *Quality Engineering,* 21(2), 2009, 222-8.

[141] Womack, J. P., and D. T. Jones, "Beyond Toyota: How to Root Out Waste and Pursue Perfection," *Harvard Business Review,* 74(5), 1996, 140-158.

[142] Womack, J. P., D. T. Jones, D. Roos, and D. S. Carpenter, *The Machine That Changed the World,* 1990, New York, Rawson Associates.

[143] Woodard, T. D., "Addressing Variation in Hospital Quality: Is Six Sigma the Answer?" *Journal of Healthcare Management,* 50(4), 2005, 226-36.

[144] Young, D., "Six Sigma Black-Belt Pharmacist Improves Patient Safety," *American Journal of Health-System Pharmacy,* 61(19), 2004, 1988-1996.

Community Organization Model and Applications: The Case of Falls Among the Elderly

Hengameh Hosseini

Assistant professor
Master of Health Care Administration Program
Political Science and Public Affairs Department
Steno Hall University
520 Jubilee Hall
400 South Orange
South Orange NJ
973-761-9212
hosseihe@shu.edu

ABSTRACT

The purpose of this study is to discuss falls among the elderly and their risk factors, the injuries caused by those falls, and the impact of fall injuries among the aged on the health care costs in the United States. As the paper argues, the aging of American society has various health and safety consequences, among them falls among the elderly and tremendous health care costs as the result of fall injuries.

My partner for the study of aging is Meadows Nursing Home in Dallas Pennsylvania, a nursing home with numerous older individuals. Since the author's goal for the study has been the development of a plan that can lead to decline of falls among the elderly, thus to lower health care costs, the study has utilized the Community Organization Model that has demonstrated its usefulness in the past. Utilizing this model and the data collected, the study has demonstrated strategies that if implemented will lead to fewer falls among the elderly, fewer fall injuries, and lower health care costs in the United States, particularly that part paid by the US government in the form of Medicare.

As the paper demonstrated, successful application of this model requires that its various suggestions be communicated to various stakeholders that include: nursing facilities, hospitals, older individuals and those in charge of their care, and other interested parties. Communication can use the media, social networks, health care community newsletters, and other such sources.

Introduction

Falls among the elderly is a topic of great interest within the health care/public health/medical communities. The purpose of this study, after analyzing these falls and their consequences, is to apply the Community Organization Model to the study of falls among the elderly which, to the author's knowledge, no one else has done before. My partner for the study is Meadows Nursing Home in Dallas Pennsylvania.

As implied above, a large sector of American society is aging, since Americans are living much longer. This has various consequences, since there are many health and safety – related problems associated with older individuals. One such problem is a substantial increase in the number of falls among this age group that often lead to injuries, which in turn cause substantial increases in health care costs , especially in communities with a large number of older persons such as Northeastern Pennsylvania . (Luuken, et al., Journal of Public Health). In the author's view, Community Organization Model can be utilized to develop a plan for reducing falls. This can be very useful, since fall injuries are expensive in terms of both cost and health care. On the basis of this model, and utilizing the data collected and statistical analysis of that data, the paper/ research develops strategies that, if implemented, will lead to fewer falls, fewer fall injuries, thus lower health care costs.

The researcher's plan is to communicate the strategies developed as a result of the study to various stakeholders that include: nursing homes, hospitals, individual elderly/their relatives through the media (printed and electronic), PA Department of Health, websites, and health care community newsletters. I will also present these findings in my health care class (s), and my conference presentations.

The study includes a descriptive component that presents a through overview of the state of research literature of falls prevention, in particular as it relates to local communities. It also includes the analysis of my qualitative survey and its results that leads both to a contingency model of how to prevent falls under different circumstances and identification of environmental and health factors that most strongly correlate (statistically) with falls in general. The questions asked in my qualitative research included, those such as: what caused falls among the elderly in Meadows Nursing homes, under what circumstances did falls occur, and so forth.

The Aging of U.S. Population and Falls Among the Elderly.

Americans are living longer, with older adults being the fastest growing segment in the United States' population. While older adults constituted 13% of the U.S. population in 1990, by 2050 this will be 23% of the United States Population. This rise in the older population has many consequences , among them are tremendous rise in the number of falls among the elderly. This rise in falls constitutes a serious problem. For example, in 2000 there existed some 1.8 million falls among the elderly in the United States which led to an emergency room visit with high costs. Furthermore, as suggested by Kochera (2002), falls constitute the leading cause of death from injuries among the elderly.

Older adults may fall due to two types of causes-intrinsic and extrinsic. The intrinsic type of falls among the elderly includes age related causes such as muscle weakness and diseases causing problems for balance, mobility, and daily human activities, cognitive impairment, hypertension ,living alone, poor foot care, unsafe footwear, impaired vision, hearing problem, dizziness , fear of falling, and neurological problems such as Parkinson's disease. (Guerlich, 1999).

The extrinsic type causes include environmental hazards such as poor lighting at home (or facility), irregular or slippery floor surfaces, unsafe stairways, crowded walkways , loose carpeting or throw rugs, furniture, and devices such as walkers, and inadequate support in bathrooms ,etc.   (Tinetti, Baker, and McAuy, 1994).

No matter which of these two cause the fall, falls among the elderly are costly. For example, the above mentioned 1.8 million falls among 34.7 million elderly in the year 2000 had a cost of 16.4 billion in direct medical and long term costs, the average direct cost per fall being some $9,400 in that year. (Kochera, 2002). Of these costs , 64% was paid by Federal government -48% of it by Medicare and 16% by Medicaid, 11% by state Medical funds and 25% by both private insurance and out of pocket payments. (Ibid). Given the rise of all medical costs in the United States, in particular for the elderly, since older persons need a great deal more medical attention than those who are younger, it is essential to find ways to reduce the very high cost of health are among the elderly, including those costs associated with falls among them. Since many of the falls associated with the environment (i.e. extrinsic) occur at home (including nursing home) or close to home, one way to reduce falls among the elderly is to modify the home/nursing home environment, since it would reduce or eliminate common hazards for frail elderly person. This modification can be conducted through the application of the Community Organization Model. Unitization of this model can assist us to reduce the direct health and long-term expenditures; it can help us improve the quality of life among the elderly. It can also help us reduce the burden among caregivers and family members, reducing the financial burden on them, saving them the time that can be lost from work as a result of falls, and avoiding problems related to the emotional and physical stress that can arise due to falls. Of course it would sustain the productivity of the older persons not falling.

Where Do Falls Among the Elderly Occur, and What Causes Those Falls?

A study was conducted by the National Health Interview Survey (NHIS) several years ago which was aimed at determining the nature and consequences of falls among the elderly, in particular falls that led to injury three months prior to the survey. Although the study was not focusing on nursing facilities, its findings are nevertheless useful for understanding the nature of falls no matter where they occur, including in nursing facilities. According to that study:

- Among the 65 and older age group, the majority (55%) of fall injuries took place inside the house, 23% occurred outside by near the house, and 22% occurred away from the house. To appreciate these statistics, it is use to compare them to those for a younger group of persons. For example, for those between the ages of 35 and 64, only 26% of the falls occurred inside the house, and 26% outside but near the house, but 48% of those falls occurred away from the house.

- Of those falls among the elderly which occurred both indoor and outdoor, 43 percent occurred at floor or ground level (i.e. not from a height); fourteen percent of those falls among the elderly occurred on stairs or steps; eleven percent from a curb or sidewalk, nine percent from a curb or sidewalk, and nine percent from a chair, bed or other types of furniture; and four percent at those falls occurred in bathtubs, showers, or on toilets. Of course, the locations of 26% of the falls were not specified.

- In terms of what caused those falls; fifty –nine percent of those falls among the elderly were caused by slipping, tripping, or stumbling. Among those falls, twenty-two percent were caused by a loss of balance, dizziness, fainting, or seizure. And, nineteen percent of those falls had other unspecified causes.

- In terms of the consequences of those fall injuries among the elderly, twenty-eight percent resulted in short-or long-term limitations in one or more activities required for daily living. This was fewer for the persons between the ages of 35 and 64, since it was only 16% for this age group.

  The above-mentioned activities limited as a result of fall injuries, according the NHIS, included: activities like eating, bathing, dressing, getting around the house, or routine activities such as performing everyday household chores, doing necessary business, shopping, or getting around for other purposes.

- In terms of the number of fall injuries leading to hospitalization: 20 percent of the injured elderly ended up to an overnight stay at a hospital, compared to only 10 percent for persons age thirty five to sixty four.

What is the Community Organization Model

As stated above, this study utilizes Community Organization Model. But what is this model? Community Organization Model is a model that "emphasizes active participation and development of communities that can better evaluate and solve health and social problems". (the National Cancer Institute, 2005, P.7). This model is model that provides a framework for understanding how social systems function and evolve and how organizations and communities can be motivated toward a particular goal, in this case to prevent or reduce the number of falls among the elderly. The Community Organization is a useful one because it can allow us to improve the health of the elderly, by preventing or reducing fall injuries, reducing fall injury risks, managing it, thus improving the well-being and self-sufficiency of the elderly. This model can complement individual actions with advocacy because it involves local community

development and action while emphasizing empowerment, community competence, and participation. In fact the model has been very used very successful in the past. It is useful to describe two such examples of the Community Organization Model.

The first application of this model is the Tenderloin Senior Organization Project (TSOP) which was a community organization effort that involved elderly residents of San Francisco's Tenderloin district a few decades ago. (Weschlser and Minkler, 1986). This 1979 effort involved graduate students and faculty members from the University of California, Berkeley, School of Public Health who had two goals: (a) improving physical and mental health among Tendeloin residents by reducing their social isolation and providing relevant health education, and (b) facilitating, through dialogue and participation, a process that encouraged residents to work together to identify common problems and seek solutions to these problems.

The application of Community Organization Model in the case of TSOP was very successful, since it gave rise to several positive outcomes. For example, the effort led to a decline of crime in Tendeloin district by 18 percent during the first year and by 26 percent during the second year. Additionally, changes in individual health and social behavior were also seen. As a result of the effort by TSOP, and an increased feeling of self-efficacy and social support, some residents decided to quit smoking or reduce alcohol consumption; an improvement in the mental health and self-esteem of the residents of Tendeloin was also observed (Minkler and Wallerstein, 2002),

A second example of the application of the Community Organization Model was the start of a youth-centered experimental program that began in New Mexico in 1986. (Weschlser and

Minkler, 1986). This example, named the Adolescent Social Action Program (ASAP) began, in 1982, as a youth-center experimental program involving collaboration among the University of New Mexico, University hospital, county detention center, and more than 30 multiethnic schools and communities in New Mexico. ASAP began with several goals: (a) to reduce excess morbidity and mortality among youths who live in high-risk environments, (b) to encourage these youths to make healthier choices, and (c) to empower them to take an active political and social role to improve their neighborhoods and communities. (Ibid).

The application of Community Organization Model in the case of ASAP too proved to be very successful. Wallerstein and Martinez (1994) conducted a qualitative research study of two participating ASAP high schools. As they argued: "the data uncovered the importance and value of talking as a way for youth to help other group members, patients, family and friends. Dialogue was a way to establish connections, disclose personal feelings, and adopt a caring stance, a precursor to social responsibility". (P. 133).

A participating student told the researchers that: "In ASAP everyone is open. They can talk and people listen… talking to these people at the hospital, they really got me to think; some of these people in my neighborhood might want to just talk. May be if we just talk to them and tell them "I did care about you", may be that would make an impression in their life." (Ibid, p.133). According to those two authors/researchers, application of the above model in the case of ASAP also resulted in change of behavior and self growth. As an example, the two authors quoted a young person who admitted that; "When I am partying, I think of the lady who died. Now I think that could happen. Now I control it, know when to stop". (Ibid).

The Meadow Nursing Center and Its Mission

The Meadow Nursing Center, an assisted living facility which is located on a 15-acres tract in Dallas Pennsylvania, is owned and managed by Ecumenical Enterprises, Inc. (EEI). Ecumenical Enterprises Inc. is a private, non-profit corporation sponsored by the First Presbyterian Church of Wilkes-Barre, the Roman Catholic Diocese of Scranton, the Jewish federation of Greater Wilkes-Barr, and the Metropolitan Lutheran Council of Wyoming Valley. In a statement by the EEI about its Meadows campus in Dallas, Pa we read: "Born out of concern for helping the ill and the infirmed, the Meadows Nursing center approaches its care in a compassionate manner. To us, our residents are truly special people with special needs". This nursing facility was established in 1983, as a Medicare and Medicaid eligible facility with 120 bed, and providing medical attention, physical therapy, occupational therapy, religious services, and specialized dietary menu. In 1993, it added ten new beds; it is now a 130 beds facility, and the staff of 23 persons.

Apparently, any type of community health care agency, including Meadows Nursing Center, is obligated to be insured. According to the documents provided by Meadows Nursing Center, the center is protected by Liability Insurance-from two insurance providers- American casualty Company of reading, PA, and Columbia Casualty Company.

For over twenty insurance providers-insurance-from two insurance providers –American Causally Company of Reading, PA, and Columbia Casualty Company.

For over twenty years, Meadows has also had a volunteer program, with some 100 members from the Back Mountain area in Luzern County to help the residents. The volunteers are retired persons –both male and female.

On the Summer time, Meadows Nursing Center also runs a volunteer program for kids ranging from 13 to 16 years old. Each Summer, about 25 young volunteers read for the residents, transport them, and help them in their various needs. Every summer, Meadows Nursing Center organizes fundraising events around the Meadows pond, helping to pay for various items needed by the residents.

What is the Mission of the Meadows?

As we read in the Mission Statement (of EEI, thus Meadows), the : "mission is to sustain and improve the quality of life of the older adults and those with physical disabilities that live in our diverse communities regardless of faith, backgrounds and income levels by providing housing and nursing services."

As the EEI's Mission Statement indicates, the Meadows Nursing Center provides the following services:

- 24 hour skilled nursing care, authorized by the resident's physician, and a full array of personal, dining, therapeutic, social, spiritual, and recreational services in Meadows Nursing Center.

Of course, EEI provides services as well. EEI provides subsidized independent living apartments in seven housing developments whose income level falls within the limits by U.S. Department of Housing and Urban Development and the U.S. Department of Agriculture. EEI provides associated living units in Meadows Manor for those residents that require some supervision and support with general daily activities such a bathing, dressing, dining, housekeeping and medical management. And, EEI also provides nursing and caring child day care at the Little Meadows learning Center for their employees and general public during week days, ranging from infants to kindergarten.

Meadow Nursing Facility and Its Fall Management Policy and Procedure

At the outset, it is worth mentioning that according to Meadow Nursing Facility a fall occurs "when an individual unintentionally comes to rest on the ground, floor, or other lower level that is not as a result of an overwhelming external force."

Obviously, Meadows Nursing Facility wishes to prevent or reduce falls among its early residents. This is obvious from its stated policy, or its stated purpose. According to its stated policy: "The falls Management Program is to prevent /reduce falls injury and ultimately improve the quality of life for our residents". According to those documents, the purpose of Meadows is "to limit and /or prevent the occurrence of falls within the parameters that can be controlled through structured program interventions. To minimize the severity of injuries sustained by a resident resulting from a fall."

The Meadows Nursing Facility in Dallas, as described by its documents, has a fifteen step Fall Management program. These include:

- A Fall Risk Assessment will be completed on admission and re-admission (to be completed by the admitting nurse), quarterly , with significant change and annually by the RNAC. Initial assessment is completed within 48-72 hours.

- Residents will b monitored on admission and readmission for five days and post fall for there days using an specific fall risk observation form.

- On admission and readmission, unless declined by the resident, on alarm unit will be applied to the resident's bed watched for five days. During these five days the resident will

be monitored for unsafe acts (ambulation, transfers). Via the fall risk observation form. If no unsafe acts are noted, the alarm will be discontinued. If unsafe acts are observed, appropriate interventions will be implemented, which may include continuation of this alarm.

- A Falling Leaf will be used to alert staff of fall risk status. Residents identified to be at high risk for falls, residents that have had a fall in the last 90 days, and residents with an alarm system in place will have a Falling Leaf placed outside the door (of their room) and the closet door of their room to heighten staff awareness to fall risk.

- Treatment Plan and Intervention will be individualized, at Meadows. These include: Use of devices such as alarms, protectors, room/bed modifications, behavior modifications, medication review and modification pain management, exercise or activity program, therapy, toileting programs, resident education, etc. as appropriate.

- Interventions will be put into place according to individual resident needs per the Supervisor and Charge Nurse. A resident specific action plan will be developed, which may involve therapies in the development of interventions. Interventions are in the least restrictive format. The charge nurse will receive and transcribe any interventions required and document on the treatment record, ACP and closet care appropriately.

- Interventions will be assessed for success. If unsuccessful, the resident will be re-evaluated for new interventions.

- Restorative nursing will assist when applicable in maintain resident functional status.

- The Care Plan will reflect the interventions post fall.

- Documentation of falls and interventions will be in the medical record. Post fall investigations will be completed by the RN Supervisor Utilizing at Incident/Accident form.

- Employees (of the Meadows) will evaluate facility environment to identify fall hazards on an ongoing basis. Any hazards indentified will be reported promptly to the maintenance department.

- Resident beds will be maintained in the lowest position and returned to this position if raised for care or cleaning.

- An Accident/Incident committee will review resident incidents, including falls, at least bi-monthly. Committee members will include the Administrator, Don, Quality Assurance Nurse, and may also include a therapy staff member, Behavior Nurse and Restorative Nurse. Data and interventions will be reviewed and evaluated and patterns and trends will be identified. Recommendations will be made to the Supervisor as needed in order to reduce the occurrence of accidents or incidents. Fall rates will be reported to the QA committee quarterly.

- Resident and /or Family education regarding falls and fall prevention strategies will be provided on admission and as needed.

- Nursing staff education regarding falls, fall reduction and prevention will occur at least annually and as needed.

Meadow Nursing Facility seems to have a relatively appropriate fall management policy and procedures. However, it is not perfect and can be improved. It seems to me that a very effective intervention can be a multifaceted fall risk assessment and its resulting fall management program. Multifactorial interventions conducted by health experts can prevent falls, in particular

if they are aimed at and concentrate on persons at risk and include several intervention approaches.

There are various types of interventions that can reduce serious falls among older persons. Examples are exercise, medication changes, education, and other types of therapies. Such interventions can include simple environmental corrections such as reducing clutter, eliminating electrical cords across floors, etc. to more permanent changes at home that include changes in handrails , grab bars, ramps, and others.

Of course, doing the above requires a thorough assessment of older persons at home or in a facility. These assessments should be done by credible health care professionals who can assess the fall history of the older persons at risk, their mobility, and their cognitive assessment. These professionals can also review the type and number of medications taken by the older persons at risk, their blood pressure, and the safety of the homes in which they are residing. Of course, these health professionals should communicate the results of these assessments to the older persons themselves, their family members, especially if they live at home, or those responsible for their care if they live in a facility. Health professionals should also make sure that home modifications do actually take place.

To make sure that we are effective in promoting falls management policies, we can , as suggested before, utilize the Community Organization Model, to further help reduce falls among the elderly, and thus fall injuries and health care costs. To apply the model we can do the followings:

1- Raise public awareness, at least among those concerned, that falls, thus fall injuries, among the elderly are preventable .

(at least can be reduced). This can be done by explaining risk factors and preventive methods and policies in booklets that are distributed to older persons, family members, health professionals, and employees of nursing facilities.

2- Organize and start community education projects that include exercise classes with the potential to improve the motor skills of the elderly.

3- Train nurses and other health professionals to serve as : (a) fall prevention advisors, (b) medication workshop leaders who understand the types and relationships among different types of medications taken by the elderly at the risk of falling, and (c) home safety advisors who can provide information about risk-free homes and nursing facilities. These individuals can provide direct instructions to the elderly themselves, or to those responsible for their cares.

4- Develop fall prevention programs that would enable local health professional to identify fall hazards and risks.

5- Address home hazard reduction by compiling a home safety checklist. This type of checklist can include information on the range of available home safety products in local hardware stores. Additionally, we can also educate hardware store owners and even employees about the importance of their help in preventing falls among the elderly.

6- Designate, in each locality/region, a general practitioner physician as a liaison officer to ensure that all physicians are aware of available fall prevention initiatives in the community and how to make referrals to such programs.

An important questions to ask that concern the above is : who should be the responsible person to make sure that these steps are taken? In response, we should try to involve various stakeholders in the community, particularly those who have something to gain from the reduction

or prevention of falls among the elderly- at-risk elderly. These include family members of those individuals, nursing facilities, et al. In other words, government (i.e. those in charge of Medicaid and Medicare), hospitals , nursing homes, insurance companies, schools of public health, and community activists interested in public safety should be involved in the above- mentioned activities.

References

Alexander, M.R., Ambre, J.J., Liskow, B.I., & Trost, D.C. (1979). Therapeutic use of albumin. *Journal of the American Medical Association, 241,* 2527-2529.

Alexander, M.R., Alexander, B., Mustion, A.L., Spector, R., & Wright, C.B. (1982). Therapeutic use of albumin: 2. *Journal of the American Medical Association, 246*, 831-833.

Binda, S.M., Culham, E.G., Brouwer, B. (2003). Balance, muscle strength, and fear of falling in older adults. *Experimental Aging Research, 29,* 205-219.

Buchner, D., Cress, M., de Lateur, B., Esselman, P., Margherita, A., Price, R., & Wagner, E. (1997a). A comparison of the effects of three types of endurance training on balance and other fall risk factors in older adults. *Aging, 9,* 112-119.

Carter, N.D., Kannus, P., & Khan, K.M. (2001). Exercise in the prevention of falls in older people: A systematic literature review examining the rationale and the evidence. *Sports Medicine*, *31*, 427-438.

Chiv, A.Y.Y., Au-Yeung, S.S.Y, & Lo, S.K. (2003). A comparison of four functional tests in discriminating fallers from non-fallers in older people. *Disability and Rehabilitation*, *25*, 45-50.

Democratic Progressive Party. (1999). White paper on China policy for the 21st century. *Taiwan Documents Project*. Retrieved June 4, 2000, from http://newtaiwan.virtualave.net/dpp01.htm.

Drummond, M. (1990). Allocating Resources. *International Journal of Technical Assessment of Health Care, 6*, 77-92.

Drummond, M.F., O'Brien, B., Stoddart, G. L., & Torrance, G. W. (1996). *Methods for the economic evaluation of health care programs (*2nd ed.). New York: Oxford University Press.

Friis, R.H. & Sellers, T.A. (1996). Epidemiology for public health practice.
     Aspen Publishers, Inc.

Gardner, M.M., Robertson, M.C., & Campbell, A.J. (2000). Exercise in preventing falls
     and fall related injuries in older people: A review of randomized controlled trials.
     British Journal of Sports Medicine, 34, 7-17.

Glick, T., Wientjes, H., & Rai, G. (1990). An evaluation of risk factors for inpatient falls
     in Acute rehabilitation elderly care wards. *Gerontology, 42,* 104-107.

Gorin, F. (1998). Falls in the nursing home: Preventable or Inevitable? *Journal of
     Gerontological Nursing, 18,* 20-35.

Guelich, M.M. (1999). Prevention of falls in the elderly. *Topics in Geriatric
     Rehabilitation, 15*, 15-25.

Hanlon, J.T., Landerman, L.R., Fillenbaum, G.G., Studenski, S. (2002). Falls in African
     & American and white community-dwelling elderly residents. *Journal of
     Gerontology A, 57,* 473-478.

Hill, K., Smith, R., Murray, K., Sims, J., Gough, J., Darzins, P., Clark, R., & Vrantsidis,
     (2000). *Report: An analysis of research on preventing falls and fall injury in
     older people: Community, residential and hospital settings.* Canberra: Department of
     Health and Aged Care.

Huang, H.C., Gau, M.L., Lin, W.C., & Kernohan, G. (2003). Assessing risk of falling in
     older adults. *Public Health Nursing, 20,* 399-441.

Joint Commission on Accreditation of Healthcare Organizations (2002). Healthcarecat the
Crossroads: Strategies for addressing the evolving nursing crisis. Retrieved February 19, 2005,
from http://www.jcaho.org/about+us/public+policy+initiatives/health+care+at+the+crossroads.

Joseph, M.A. & Tipton, D.J. (2005). Patient Falls: Getting to Excellence from a
     Nursing Home Administrator's perspective. Duquesne University.

Kempton, A., Van Beurden, E., Sladden, T., Garner, E., & Beard, J. (2000). Older
     people can stay on their feet: final results of a community-based falls prevention
     programme. *Health Promotion International, 15,* 27-33.

Lord, S.R., & Bashford, G.M. (1996). Shoe characteristics and balance in older
     women. *Journal of the American Geriatrics Society, 44*, 429-433.

Lord, S. R., Ward, J. A., Williams, P. & Zivavocic, E. (1999). The effects of a
     community exercise program on fracture risk factors in older women.
     Osteoporosis International, 6, 361-367.

Mientka, M. (2002). Nursing Shortage Affects Quality of Care. *US Medicine: News Highlights.* Retrieved June 29, 2005, from http://www.usmedicine.com/Dailynews.cfm?DailyID=125.

Mitchell, A., & Joes, N. (1996). Striving to prevent fall in an acute care setting: Action to enhance quality. *Journal of Clinical Nursing*, *5*, 215-220.

Mosenthal, A., Livingston, D., & Elcavase, P. (1995). Falls: Epidemiology and strategies for prevention. *Journal of Trauma*, *38,* 753-756.

Mosley, A., Morse, R., & Julko, S. (1989). Development of a scale to identify the fall-prone patient. *Canadian Journal on Ageing, 8*, 366-377.

National Cancer Institute. (2005). Theory at a Glance: A Guide for Health Promotion Practice. In *U.S. National Institutes of Health Publications*. Retrieved February 20, 2005, from http://cancer.gov/aboutnci/oc/theory-at-a-glance/page8.

Nyberg, L., & Gustafson, Y. (1995). Patient falls in stroke rehabilitation: A challenge to rehabilitation strategies. *Stroke, 26*, 838-42.

Parry, S.W., Steen, N., Galloway, S.R., Kenny, R.A., & Bond, J. (2001). Falls and confidence related quality of life outcome measures in an older British cohort. *Postgraduate Medicine, 77,* 103-108.

Peters, M., Kochanek, K.D., & Murphy, S.L. (1998). Deaths: Final Data for 1996. *National* Vital Statistics Systems Centers for Disease Control and Prevention, National Center for Health Statistics.

Rowell, D., Scott, D., Barnett, L., & Beard, J. (2004). Measuring the economic impact of falls prevention interventions. *Australasian Journal of Business and Social Inquiry, 2,* 1-13.

Salkeld, G., Cumming, R. G., O'Neill, E., Thomas, M., Szonyi, G., & Westbury, C. (2000). The cost effectiveness of a home reduction program to reduce falls among older persons. *Australia and New Zealand Journal of Public Health, 24,* 265-27.

Saw, S. M., & Ng, T. P. (2001). The design and assessment of questionnaires in clinical research. *Singapore Medical Journal*, *42,* 131-135.

Tinetti, M.E., Baker, D.I., & Mcaway, G. (1994). A multifactorial intervention to reduce the risk of falling among elderly people living in the community. *New England Journal of Medicine, 331,* 921-927.

United States Bureau of the Census (1989). *Census Report*. Washington, D.C.

How Do Parents Select Their Pediatrician?

A Multiple Criteria Decision Making Process

Dr. V. Miligkos, MD, PhD
Staff Pediatrician, Pediatrics Clinic, Mitera S.A. Greece

Dr. Konstantinos Nikolopoulos[*], PhD
Professor of Decision Sciences, Bangor Business School, U.K.
*for*TANK, the forecasting think tank

Dr. M. Miligkos, MD,
PhD Student, Medical School, University of Thessaly, Greece

Dr. Sally Sambrook
Professor of Human Resources Development, Bangor Business School, U.K.

*Corresponding Author: k.nikolopoulos@bangor.ac.uk; mob: 0044 7981 332913

How Do Parents Select Their Pediatrician?
A Multiple Criteria Decision Making Process

Abstract

In this study we investigate the process by which parents select their children's primary health care providers, i.e. their Pediatricians. For this purpose, a close-ended questionnaire has been administered to 320 parents and doctors in Greece with an approximately 100% response rate due to nature of the collection process. About 80% of the responses came from a group of parents that had already selected their pediatrician at the time of the survey; in fact, their own pediatrician was involved in the information collection process. Two complementary subgroups were also considered; a 13% of the sample came from the very Pediatricians themselves, aiming to obtain some insight on the potentially discrepancies between what specialists think is important and what actually matters. An additional 6% of the sample consisted of mothers in the first four days of their first childbirth. This subgroup enabled the examination of the evolvement of selection criteria in this decision making process. Our empirical results demonstrate that for all three groups, the selection priorities concerned primarily the pediatricians' communication skills, followed by their experience and lastly more generic issues such as the proximity of the surgery to their home. More specifically, accessibility over the phone and ability to explain the course and treatment of the illness constitute significant priorities for the parents. The analysis of the results also revealed some very interesting in-between subgroups differences: these relate, for example, to the perceived importance of doctors' formal qualifications and training in the course of selection, the actual cost-per-visit and the doctors being willing to pay home visits. Our study offers insights on the initial steps of obtaining child health care services, a relatively unexplored area of patient decision making. Furthermore, we offer a Human Resources (HR) perspective, through a discussion of the implications of our findings for HR Management (i.e. how private childcare clinics should recruit their specialists) and HR Development (i.e. how private childcare clinics should train their specialists). A natural extension of the work described in this paper is the construction of a Multiple Criteria Decision Analysis (MCDA) model, after a follow-up round of primary data collection via in-depth interviews with individuals that constituted our

original sample. Such a model will enable the sequential nature of this decision making process to be fully revealed.

INTRODUCTION

In this study we investigate the process by which parents select their children's primary health care, i.e. Pediatricians. This study was inspired – by and large by an earlier study in 1988 by Hickson, Stewart, Altemeier and Perrin in *Pediatrics*, where the main finding was the importance of communication skills in the profession of a pediatrician. We wanted to see if this is still the case, and for the first time in Greece; for this purpose a close-ended questionnaire was administered to 320 persons in Greece.

DATA COLLECTION – SURVEY

The data - in the form of a two-page questionnaire -have been collected in Greece during the period of November and December 2010. In the regional area that was sampled (region of Attica, Greece); 'Children healthcare' is offered through the following channels:

- *Primary care*: a network of private Pediatricians offering their services day and night including home visits, versus a network of Public Care Trusts offering outpatient appointments only in the mornings

- *Emergency Care*: in case of emergency parents can attend A&E units in one of the only two public children's Hospital in Athens (capital of Greece with population 4.000.000+) or any of the 5 private Children Hospitals (that typically the private Pediatricians are liaising with)

The questionnaire consists of 20 questions divided into three broad categories: *General, Experience, Communication* (see Appendix); and was collected via face-to-face contact and

through emails. We had three versions of the questionnaire so as to randomise the sequence of questions. People living in the sampling area are typically part of the 'labour class', with 25% of the responses coming over the internet – typically of University education with high IT skills. 53.8% of respondents wanted to be informed on the results of this research and provided their personal contact details.

Of all the responses, 80.10% came from a group of parents that have already made their decision/selection, as they were administered the questionnaire face-to-face from their own pediatrician; a further 13.1% of the sample came from Pediatricians themselves, a group of special interest, so as to get some insight on the discrepancy between what specialists think of being important for the parents. Finally a small part of the sample – 5.8%, consisted of mothers within 4-days from their first childbirth, a group that has not yet really selected their children child care – so again it is interesting to see their perspective as well.

FINDINGS

We have broken our sample into three subgroups:

- Group 1 : **DECISION** made

This subgroup stands for 80.1% of our data and includes parents sampled at Private Childcare Surgeries and at a Private Children Hospital; those are the parents that have already chosen a Pediatrician for their child's (children's)

- Group 2 : **Pediatricians**

This subgroup stands for 13.1% of our data and includes Pediatricians themselves, so as to see what criteria think that are the more important

- Group 3 : **NO Decision** made yet

This subgroup stands for 5.8% of our data and includes mothers within 4 days of their first childbirth in the maternity ward of a Private Children Hospital. These mothers have not made their mind yet of whom their future Pediatrician will be, and thus it is interesting to see how the criteria evolve over time

Group 1, Parents that already MADE a DECISION is the group that we are mostly interested in; the other two groups constitute only small part of our sample (in total 20%) and results are provided only for information purpose and to stimulate interest for further research. The most important category for all groups is: **COMMUNICATION.**

This is evidenced by the average of the scores given (per category) demonstrated in Table 1. (max=4.00)

	OVERALL		
Parents: Decision MADE	3,18	2,96	3,71
(80.1% of sample)	General	Experience	Communicattion
Pediatricians	3,12	2,69	3,39
13.1% of sample	General	Experience	Communicattion
'difference' to group 1	-	- - -	- - -
Parents: NO Decision yet	3,29	2,81	3,51
5.8% of sample	General	Experience	Communicattion
'difference' to group 1	+	- -	- -

Table 1. The most important category of Pediatrician's 'skills'

*RED: most important, **BLUE** second most important, **BLACK** third most important*

If we look into each category separately, on what counts most in the parents' decision, then we got the following three tables:

GENERAL									
Parents **Decision MADE** 2,92	3,38	3,87	3,69	2,79	3,26	3,11	2,69	3,21	2,87
(80.1% of sample) Proximity /Surgery	Word of mouth	Access /Phone	Access /Weekends	Appoint ments	Home visits	Substi tute	Access Priv Hospital	Cost	insurance cover
Pediatricians 2,74	3,38	3,74	3,57	2,83	3,21	2,75	2,67	3,38	2,93
13.1% of sample Proximity /Surgery	Word of mouth	Access /Phone	Access /Weekends	Appoint ments	Home visits	Substi tute	Access Priv Hospital	Cost	Insurance cover
'difference' to group 1 --	~	-	-	~	~	----	~	+ +	+
Parents **NO Decision yet** 3,28	3,50	3,89	3,72	2,39	3,72	3,44	2,61	3,33	3,06
5.8% of sample Proximity /Surgery	Word of mouth	Access /Phone	Access /Weekends	Appoint ments	Home visits	Substi tute	Access Priv Hospital	Cost	insurance cover
'difference' to group 1 ++++	+	~	~	----	+++++	+++	-	+	++

Table 2. GENERAL SKILLS

RED: most important, BLUE second most important, BLACK third most important

	EXPERIENCE				
Parents: Decison MADE	2,62	2,77	3,25	3,29	2,84
(80.1% of sample)	Work/Priv Hospital	Pg Degree	Attend Conf	Years in Service	Specialism
Pediatricians	2,79	2,12	2,74	3,02	2,81
13.1% of sample	Work/Priv Hospital	Pg Degree	Attend Conf	Years in Service	Specialism
'difference' to group 1	++	-------	-----	---	~
Parents: NO Decison yet	2,28	2,41	3,11	3,28	2,88
5.8% of sample	Work/Priv Hospital	Pg Degree	Attend Conf	Years in Service	Specialism
'difference' to group 1	---	----	-	~	~

Table 3. Skills associated with **EXPERIENCE**

RED: most important, BLUE second most important, BLACK third most important

	COMMUNICATION				
Parents: Decision MADE	3,67	3,65	3,71	3,81	3,69
(80.1% of sample)	Communication/General	Communication/Examination	Listening	Explaining	Persuading
Pediatricians	3,29	3,19	3,62	3,60	3,24
13.1% of sample	**Communication/General**	Communication/Examination	Listening	Explaining	Persuading
'difference' to group 1	- - - -	- - - - -	-	- -	- - - - -
Parents: NO Decision yet	3,56	3,67	3,56	3,61	3,17
5.8% of sample	**Communication/General**	**Communication/Examination**	Listening	Explaining	Persuading
'difference' to group 1	-	~	- -	- -	- - - - -

Table 4. COMMUNICATION SKILLS

RED: most important, BLUE second most important, BLACK third most important

SHORT DISCUSSION

The first striking finding is that for ALL three groups the most important skills are the COMMUNICATION ones.

For the **first group** of "PARENTS: Decision Made" the criteria ranks as follows (in parenthesis the category of each specific skill):

Very Important

  - (G) Accessibility over the phone + in the Weekends

  - (G) Word of Mouth

  - (E) Years in Service + Attending Conferences

  - (C) Explaining, Listening & Persuading

Not that important

  - (G) Cost

  - (G, E) 'Access to' and 'working for' a Private Children Hospital

For the **second** group of "Pediatricians" the criteria ranks as follows:

Very Important

- (G) Accessibility over the phone + in the Weekends

- (G) Word of Mouth

- (G) Cost

- (E) Years in Service + Specialism

- (E) Work for Private Children Hospital

- (C) Listening, Explaining & General Contact

Not that important

- (E) Attending Conferences, Pg Degrees

- (G) Providing contact information of a Substitute MD

For the **third** group of "PARENTS: NO Decision yet" the criteria ranks as follows:

Very Important

- (G) Accessibility over the phone + in the Weekends

- (G) Home Visits

- (G) Word of Mouth

- (E) Years in Service + Attending Conferences

- (C) Contact during Examination, Explaining, Listening

Not that important

- Appointments, Persuading

Our empirical results show that for all three groups, selection priorities ranked in order of importance, concerned the pediatricians' overall communication skills, and then experience or more generic issues such as the proximity of the surgery to their home; more specifically, accessibility over the phone, and ability to explain the course of the illness and the application of

the treatment are of essential importance for the parents. Interesting differences are also noted in between groups; with the Pediatricians rating relatively much lower the existence of Doctors' Postgraduate training, while rating relatively higher the importance of the cost-per-visit. On the other hand, mothers just after childbirth, rate relatively much higher the option for home visits from the doctors.

CONCLUSSIONS AND FURTHER RESEARCH

The study results provide insight about the first step in obtaining child health care services, a relatively unexplored area of patient decision making globally, and definitely in Greece.

All three groups do consider the communication skills as the most important ones overall. However as the highest ranked criterion is surfacing the **accessibility** of the doctor **over the phone**:

So... parents do need to be able to find a friendly and expert voice over the phone 24/7!

We leave for further research a follow up round of in-depth interviews where the sequential nature of this decision making process will be fully revealed.

REFERENCES

Hickson, G.B., Stewart, D.W., Altemeier, W.A. and Perrin, J.M. (1988) "First Step in Obtaining Child Health Care: Selecting a Physician", PEDIATRICS Vol. 81 No. 3 March 1988, pp. 333-338

CRITERIA FOR CHOOSING YOUR **PEDIATRICIAN**					
Categories					
GENERAL		NOT Important	LESS Important	**MORE** Important	VERY Important
1. How close is the surgery to your home		X			
2. Strong recommendation from someone (word of mouth)				X	
3. Accessible over the phone when you need him/her					X
4. Available in the weekends/bank holidays					X
5. Keeps a punctual appointments schedule				X	
6. Makes visits at home				X	
7. When not available, he/she suggests another Pediatrician to treat his/her patients					X
8. Has contacts and liaises with a private Pediatrician Clinique, in case of emergency		X			
9. Cost per visit					X
10. The cost of his services is partially/fully covered by Healthcare Insurance companies					X
EXPERIENCE					
1. Works for a private Pediatrician Clinique				X	
2. Holds a Postgraduate degree				X	

3. Attends Scientific Conferences				X
4. Years in service			X	
5. Further Specialism (e.g. expert in Child Asthma)				X
COMMUNICATION				
1. Communication with the children (in general)				X
2. Communication with children during the examination				X
3. Listening carefully to parent's concerns				X
4. Explaining clearly the diagnosis and treatment of the disease				X
5. Persuading of his/her scientific background/knowledge			X	

Which Category overall is more important for your Choice/Decision?

GENERAL	EXPERIENCE	COMMUNICATION
	X	X

You may choose a max of two categories

Any other important criterion for in your choice, not mentioned?

……………………………………………………………………………………………………

This piece of research will enhance the understanding of what is important for parent when choosing their children's healthcare, and will inform the respective healthcare management bodies

V. Miligkos, MD, PhD, Greece

K. Nikolopoulos, PhD, Professor of Decision Sciences, Bangor Business School, Wales, United Kingdom

Leave your contact details …………………………………………………………………………………………….

…………… if you want to be informed of the results of this research   Thank You.

STRATEGIC VISION OF THE HEALTH CARE GOVERNANCE IN THE I.R. OF IRAN

Ali Maher[1] , Behnaz Abolshams Asghari[2] , Mustafa Arda Altun, M. H. Salarianzadeh, M.H. Vaez Mahdavi, F. Sadr

INTRODUCTION

Iran is the seventeenth-largest country in the world and is located in the Middle East. Its borders are with Azerbaijan and Armenia to the northwest, the Caspian Sea to the north, Turkmenistan to the northeast, Pakistan and Afghanistan to the east, Turkey and Iraq to the west and finally the waters of the Persian Gulf and the Sea of Oman to the south. Iran's area is 1,648,000 km².

Iran's population increased dramatically during the latter half of the twentieth century, reaching about 70,049,000 in 2006. Population growth rate in Iran was 1.2% in 2002. In recent years, however, Iran's birth rate has dropped significantly. Studies show that Iran's rate of population growth will continue to slow down until it stabilizes. More than 66% of this population lives in urban areas. The life expectancy at birth rose from 69 years in 1995 to 70 years in 2004. The population is relatively young, with 29.6% below the age of 15 years (2005) and only 4.6% at or above 65 years old. The literacy rate for the population of 15 years of age and above was 82% in 2004 with the literacy rate of 76% for females of 15 years of age and 87% for the males of the same age group.

The Supreme Leader of Iran is responsible for delineation and supervision of the general policies. After the Supreme Leader, the Constitution defines the President as the highest state authority. The President is elected by universal suffrage for a term of four years. Administratively, Iran is divided into thirty provinces, each governed by an appointed governor. Each province is divided into areas, which are divided into smaller units and finally

---

[1]   Assistant professor: dralimaher@gmail.com
[2]   PhD student of Healthcare administration in Hacettepe University (behnaz09@hacettepe.edu.tr)

into villages, the smallest administrative units. Local councils are elected by public vote to four-year terms in all cities and villages of Iran. According to the Article 7 of Iran's Constitution, these local councils together with the Parliament are "decision-making and administrative organs of the State". This section of the constitution was not implemented until 1999 when the first local council elections were held across the country. Councils have many different responsibilities including electing mayors, supervising the activities of municipalities, studying the social, cultural, educational, health, economic and welfare requirements of their constituencies, planning and co-coordinating national participation in the implementation of social, economic, constructive, cultural, educational and other welfare affairs.

Iran's economy is a mixture of central planning, state ownership of oil and other large enterprises, village agriculture, and small-scale private trading and service ventures. Its economic infrastructure has been improving steadily over the past two decades. In the early 21st century, the service sector constituted the largest percentage of the GDP, followed by mining, manufacturing and agriculture. About 45% of the government budget came from oil and natural gas revenues and 31% came from taxes and fees. Government spending contributed to an average annual inflation rate of 14% in the period 2000-2004. In 2004, the GDP was estimated at $163 billion or $2,440 per capita ($8,100 at PPP). Because of these figures and the country's diversified but small industrial base, the United Nations classifies Iran's economy as semi-developed.[i]

The health care system in Iran is a unique case of structural integration between the health care delivery system, the medical education and the Ministry of Health and Medical Education (MOHME) which was established in 1985. The Ministry combines the responsibility for health service provisions and medical training. Both governmental and nongovernmental bodies play a role in the stewardship of the health care system. Resources

are generated by the government through its state funded medical education and construction of major health facilities. More than 50% of the financial resources for purchasing health services are paid out-of-pocket or through individual medical insurance, and the rest by the government.[ii] The government, through its regional enterprises known as Universities of Medical Sciences, delivers most of the care while the rest is supplied by the Social Security Organization, the private sector and different charities

During the 1960's and 70's, the country experienced the expansion of health care services through "Health Corps and Pilot Projects of Primary Health Care". However, it is since the Islamic Revolution in 1979 that the country's health care policy has been based on primary health care emphasizing the expansion of health care networks and plans in rural areas, with priority allotted to preventive over curative services.

During the 1980's and 90's, more attention was paid to the followings:

- Reduction of population growth through family planning;
- Control of diarrhea, respiratory and iodine deficiency diseases;
- Integration of mental health, tuberculosis, leprosy, diabetes, and malaria programs into the primary health care network system;
- Community-oriented medical education;
- Considerable increase of immunization coverage;
- Reduction of maternal and infant mortality;
- Increase of community participation;
- Increase of basic environmental sanitation and adequate safe water in rural communities;
- Expansion of health networks including the construction of district hospitals where needed.

National health care policies are decided at the meetings of the Council of the Undersecretary of MOHME, headed by the Minister. However, the initial information needed for planning purposes is collected from the bottom up. In each province of the country, there is at least one university of medical sciences and health services. The chancellor of this university is in charge of all health care affairs in that province, executing his duties through deputies for health, treatment and so on. The chancellor also works with the deans of different health-related schools. In each district, there is a district planning council, to which

each sector submits its planning needs in priority order. The plans approved by this council are referred for formal consideration to the provincial council, which coordinates the plans and ultimately sends them to the planning councils at the national level.

Public health and primary health care now account for 13% of the medical curriculum. In 2006, there were 12.96 physicians, 3.4 dentists and 3.1 midwifery personnel per 10,000 population. However, despite government incentives to attract physicians to rural deprived areas, the distribution of physicians to these areas has been very inequitable. Urban areas have 23 times more physicians than the rural areas.[iii] As to the physical resources, there were 520 university hospitals and 11.8 hospital beds per 10,000 populations in 2005. Available health care facilities in 2005 included 16939 health houses, 2424 rural health centers, 2256 urban health centers and 1067 health posts.[iv]

Definition of the health care system as stated in the World Health Organization (WHO) Report (2000) refers to the institutions, people and resources involved in delivering health care to individuals. Based on this reference, the health care system conceptual framework proposes four functions: Stewardship, Financing, Resource Creation and Service Provision.[v] The most elusive of these is the stewardship. In some ways, the concept of stewardship has been derived from a much widely understood notion of governance. Stewardship has been stated to be the careful and responsible management of the well-being of the population, the very essence of good governance. Just like governance, stewardship does not have a specific connotation to health, either.[vi] Governance is essentially a function of the state yet it influences all other actors in the health care system (i.e., the communities, private sector, civil society, international agencies, and nongovernmental public organizations). Governance can be seen as the exercise of economic, political and administrative authority to manage a country's affairs on all levels. It encompasses the

mechanism and processes through which citizens and groups articulate their interests, exercise their legal rights, meet their obligations and mediate their differences.

Many international development agencies, such as the World Bank, have delineated the characteristics and principles of governance that are well understood and have been used in measuring governance. One of these principles is "Strategic Vision". Leaders and the public have a broad and long-term perspective on good governance and human development along with a sense of what is needed for such development. There is also an understanding of the historical, cultural and social complexities in which this perspective is grounded.[vii]

The goal of this paper is to describe an analytical framework that would help to analyze the strategic vision of the governance in health care system at the national and sub-national levels in a simple yet reliable manner in Iran. The rationale underlying this effort is to assess formulation and implementation of the vision and strategic planning in health care system that would help the governors promote high quality health care governance.

DATA COLLECTION AND METHODOLOGY

This study is based on a methodological framework to assess the governance in health care system developed by the division of health systems and services development of the Eastern Mediterranean Regional Office (EMRO) of the WHO. The framework has adapted the widely accepted definition and the underlying principles of governance as recommended by the United Nations Development Program (UNDP). The assessment framework aims to explore the processes of the long-term strategic management rather than the political structures and outcomes. The framework elicits results that are relevant to stakeholders in their task of strengthening governance in health care system.

Overall, two broad approaches can be considered in assessing strategic vision in the health care system in Iran: 1) direct assessment based on the generally accepted domains of vision and 2) indirect assessment of the overall performance of the Ministry of Health

(MOH). The principle of vision has been disaggregated into different domains to capture as best as possible its full meaning and to be able to begin to express it in more operational terms. Each domain is divided into broad questions at three levels. These three levels are explored to assess the vision of the health care system governance.

It is assumed that Ministry of Health in any country is the principal governing body for health care as it has the mandate for such functions as policy making, planning, regulating, monitoring and evaluating of the health care system in addition to purchasing and ensuring provision of health care services. Thus, there are two levels: 1) policy formulation and 2) policy implementation. In many countries, these two levels are not clearly segregated while in other countries implementation of health care services falls under the jurisdiction of the sub-national agencies. In addition to MOH, there is a level above it that has a major influence on health care vision. The national government through its broad socioeconomic policies, legislative function, civil service reforms and political stability can heavily influence the health care vision.

This study was conducted in four phases as follows:

Phase 1: Reviewing the official and published national, regional and international documents and reports and related laws. These documents were gathered, classified and analyzed for collecting the essential data and information needed for each specific question/item. The most important documents were as follows:

- Constitution of the I.R. of Iran;
- Establishment Act and Performance reports of MOHME;
- Comprehensive Welfare and Social Security System Act;
- Five Year Development Plan (FYNDPs);
- Vision 2025 and general policies of the 4th Development Plan;
- Annual budget act of government;
- Socioeconomic report of the Management and Planning Organization (MPO);
- Official reports of the Statistical Center of Iran (SCI);
- Analytical reports on the health care system by the national and international organizations;
- Strategic plan in health sector in the 4th Development Plan;
- Related documents on international projects.

Phase 2: Preparing a primordial report. The gathered data and derived information were analyzed and the researchers prepared a primordial report. Then a questionnaire was prepared based on the report.

Phase 3: Structured interviews based on the questionnaire with a few key people. These direct interviews were designed, arranged and conducted with mid and senior managerial staff of the MOHME, MPO, universities and the parliamentarians to capture the different point of views to increase the validity and generalizability of the results.

Phase 4: Preparing the final report after analyzing the results of the interviews of the key people.

FINDINGS

Policy Formulation: National Level

In this section, we provide a brief description of the overall policy of the government in respect to market reform.

The governmental organizations have developed the following policies in order to achieve the higher economic indices while responding to public needs and people's welfare:

- Knowledge-based national production regarding human resources and social capital;
- Development of needed structures and infrastructures for growth of knowledge-based activities in public and private sector;
- Creating productive jobs and reduction of unemployment rate;
- Providing appropriate base for production of competitive goods and services in domestic and international markets;
- Abolishing the obstacles to the non-petroleum exports and providing appropriate mechanism for growth of total productivity;
- Creating compensational aids for vulnerable people through social security system in order to diminish the gap between the lower and upper income population;
- Control of inflation and increasing the purchasing power of low-income population;
- Enforcement of appropriate production infrastructure and diversification and expansion of complementary activities specifically transformation of small industries and expansion of economic activities in the fields with relative competitive advantages;
  - Empowerment of private and cooperative sectors as major initiatives for economic growth and providing direct financial support for them;
  - Commendation and creation of a culture for consumption of the internal products while increasing production and export of goods and services;

- Transformation of the assets gained through the petroleum revenues into other kinds of reserves and investments;
- Attempting to achieve social equity and providing equal opportunities.

Foreign Investments

The governmental organizations implement or will implement the following policies

in order to absorb more foreigner direct investments:

- Creating a secure atmosphere for investors;
- Developing confidence and supporting the investors while maintaining accountability;
- Legislating and implementing the Act of Commendation and Supporting External Investment (March 8, 2001) specifically in areas that create jobs in infrastructure;
- Improving banking and insurance structures in order to invigorate the capital market;
- Encouraging competition, preventing crises and combating the fiscal corruption.

Private Sector Involvement in Health Care System

The governmental organizations had or have in mind the following policies in order to

attract more cooperation from the private sectors providing health care services:

-Increasing the capacities and abilities of private sector through facilitating the process of achieving the resources, information and communication technologies;
-Expanding governmental-private technical, economic and financial links;
-Developing an environment to enhance the engagement of the private sector in providing services;
-Supporting public organizations, NGOs and small- medium sized private and cooperative companies;
-Deregulation for expansion of private ownership and participation;
-Establishing an effective and stable insurance system for improving financial relationship between people and private service providers.

Education and Public Provision of Service

The governmental organization had and has prepared the following policies in order

to provide the needed education and public services:

-Reviewing the structure and processes of education in order to empower people in developing social hypotheses at a global level;
-Enhancement of people participation in the optimal use of the country's educational capacities;
-Improving postgraduate education and technical training systems in order to make them more efficient;
-Providing the needed human resources for increasing the responsiveness of scientific and research centers country-wide;
-Compulsory education up to 8 years (secondary level);
-Expansion of second language learning and skills in computer use;

-Enhancement of the abilities and professional skills of teachers through establishing professional standards regarding knowledge, attitude and practice;
-Preparing and implementing strategic plans for teaching and the use of state-of-the-art technologies in educational courses and programs;
-Establishing a standardized system for evaluation of work force skills;
-Increasing the autonomy of the state owned universities, post graduate education and research centers.

Health Care System Rank in the Overall Developmental Framework of Resource Allocation

Table 1 indicates the percentage of health and nutrition act, social affaire and the total of the two expenditures versus the total government expenditure from 1986 to 2006.

Table 1. The Health Care Expenditure in Million of Dollars

Year	Health & Nutrition Act	Social Affairs	Total Gov Expenditure	of Total Social Affairs % to Gov Expd
1986	$0.17	$1.03	$2.41	16.7%
1991	$0.46	$2.70	$5.56	8.2%
1996	$3.11	$21.80	$37.57	8.2%
2001	$6.32	$51.73	$105.54	5.9%
2006	$14.61	$82.24	$416.40	3.5%

The total expenditure for social affairs shows a rise of 80 times when compared to the base year. During this period, total general governmental budget was raised from 2410 million Dollar in 1986 to 416400 million Dollar in 2006, revealing a rise equivalent to 172 times more than the base year. It should be noted that since 1995, after the legislation of the Public Insurance Act, spending a major portion of the health care budget has incurred through implementation of health insurance program. However, the growth rate for health care budget was slower than the growth rate for total general governmental budget.

Percentage of Health Care Expenditure

Table 1 data reveal that the allotment of social affairs expenditure to public governmental expenditure had a serious decline in 2006 comparing to 1986 (from 42.6% to 19.7%). During the same period, the allotment of health care sector expenditure versus the social affairs expenditure had a steady trend (about 17%). In this period (1986-2006), the allotment of health sector expenditure to total general governmental expenditure dropped to

half. However, a few budgetary adjustments (e.g. stabilizing the foreign exchange rate, energy subsidies to low-income households) have impacted the apparent growth of public expenditures. Also, there were some effects due to the expansion of health insurance program, such as spending about 30% of total health care credit through insurance program, which decreased the allotment of the health care sector's budget.

Finally, the data in Table 1 show that the rate of the growth of the budget for health care and nutrition during 2001-2006 was significant. However, it should be noted that a part of this high growth rate during the period 2001-2003 was due to clearance of the hidden subsidies in 2002-2003, (e.g. inclusion of 2424 million Dollar as subsidies in total health care sector credit).

Is Health Care Defined as a Basic Human Right in the Constitution?

According to the Constitution, Articles 3 and 29, health care is considered as a basic human right and the governments is supposed to offer free of charge basic health care services to all regardless of their language, race, color, ethnicity and religion.

Is There a Long Term Vision (Policy) for Health Care?

In the Vision 2025 for the Islamic Republic of Iran, the population is to enjoy health care and welfare, food security, secure and strong family unit, efficient social security, having equal opportunities, having fare distribution of income and resource and a high quality life. In addition, in the National 3rd and 4th Developmental Plans, it is noted that one of the main policies is to improve the health indicators. Among the most important strategies and middle range plans in health care sector, the followings may be noted:

-Food safety and security;
-Health care promotion;
-Reduction in life threatening hazards;
-Improvement of EMS (Emergency Medical Services);
-Health care sector reform according to demographic and epidemiologic changes;
-Prevention and treatment of AIDS, psychological disorders and drug abuse;
-Continuous improvement in quality of health services;
-Expanding health insurance system;

-Enforcing supervision and evaluation of the health care services.

Is the health care policy subject to regular review/revision processes?

The major policies and plans of the health care system, like other sectors, are revised every 5 years. The results of the review are used for revision of the next 5-year plan. In addition, when preparing the 4th National Development Plan, the 20-year plan for the country was under development at the same time, containing the long term health care sector policies.

Is there a national health policy/strategy plan explicitly stating objectives to be achieved within a time frame and specified resources?

In the 5-year development plans, there are descriptions of the current situation, sector trends, internal problems, external issues, future vision and policies and the strategies for health care sector accompanied by quantitative and qualitative indicators. For example, in the 1st and 2nd National Development Plans, the main health care policies were designed to achieve improved essential health indicators and health care services accessibility for the deprived rural population. During the 3rd National Development Plan, improved quality and efficiency of health care services was considered the main focus of the health care sector. These policies generated the necessity to remedy the impossibility of recruiting health care professionals trained by the government due to the shortage of financial resources. As a result, certain strategies were devised to remedy this situation, such as training more family physicians and enforcing the referral system.

IMPLEMENTATION

What programs are being implemented and how do they correspond to the policy objectives?

The following plans have been implemented or are under implementation by the government for the past 25 years:

-Expansion of free primary health care services for all and specifically for the rural population;
-Provision for free outpatient services for rural population;

-Provision for outpatient health services in public health centers of urban areas and its expansion with non-governmental sector assistance;
-Provision for inpatient services by government in non-developed or less developed regions;
-Delegation of some of the public centers to private sector and purchasing defined services from non-governmental sector;
-Actualization of health tariffs in accordance with the expenditure and inflation indicators;
-Enforcing the classification of inpatient services according to different specialties and burden of disease for each district for equitable distribution of resources, facilities and services;
-Regulating the national EMS network through organizing pre-hospital (urban and road) and hospital emergency care;
-Delegation of some of the emergency units to the private sector under certain conditions;
-Classifying the needed medications into two supported and free classes in order to attract insurance sector cooperation;
-Importation of raw materials for pharmaceuticals and supporting and importing drugs using formal price exchange to enhance the national pharmaceutical system;
-Allocation of subsidy for primary health care materials and drugs, long term use of drugs and drugs used for formidable diseases;
-Establishing quality control units in pharmaceutical factories and recruiting specialists to ensure quality of manufactured drugs;
-Setting priority for the pharmaceutical industry in terms of meeting currency needs, liquidity and banking facilities;
-Determination of appropriate nutritional food basket for population and dietary regimens for patients;
-Communication of nutritional knowledge and education of people in regard to the appropriate nutritional food basket;
-Planning and capacity building in order to diminish food wastes from production to consumption;
-Improving the processes to determine health tariffs, checking the bills and paying the insurers' obligation in a timely manner to attract insurance industry's cooperation;
-Improving the processes and eliminating the redundant regulations of issuing, renewing, canceling the contracts and management of the health centers;
-Clarifying the authorities and duties of all sectors related to the health care issues on providing, maintaining and improving health care for all levels of the society;
-Aggregation of all public, nongovernmental and international facilities in the way of providing health care for the population;
-Improving income and expenditure registry system and developing an appropriate accounting system meeting the needs of the health sector;
-Implementing the national health care accounting system and analyzing the health care sector economically;
-Continuous face-to-face and public health education via mass media;
-Promoting the social knowledge of high school students for control of population and family planning;
-Promoting social knowledge of women and empowering them in the national development processes;
-Improving the quality of family planning and standardizing the services in accordance to the economic, social and cultural levels;

-Improving the processes of supervising and inspecting the health and drug service providers.

Comparing these programs with the main objectives of the health care sector in the national development plans, one can find the consistency between these programs and the objectives set for MOH.

What is the package of services offered at different levels of care?

Health care service provision system covers a variety of services in accordance to the gender, age and other determinants in three levels:

1. **First-Line Services**: Primary health care:

   −Health education;
   −Pregnancy and children care;
   −Nutritional surveillance;
   −Family planning;
   −Health care in schools;
   −Mouth and tooth health;
   −Immunization;
   −Environmental and occupational health;
   −Case finding;
   −First curative aids;
   −Treatment and control of some diseases;
   −Injections;
   −Rehabilitation;
   −Assistant to the disabled and elderly persons;
   −Disease registry;
   −Obstetrics facilities;
   −Minor surgeries;
   −Basic lab tests;
   −Epidemiologic surveys.

2. **Second-Line Services:**

   −Special treatments;
   −Special lab tests;
   −Diagnostic imaging;
   −Special dental services;
   −Provision and distribution of special drugs;
   −Inpatient services for four specialties: (internal medicine, general surgery, obstetrics and gynecology, pediatrics);
   −Emergency services.

3. **Third-Line Services:**

   −Special and sub special diagnostic tests;
   −Radiotherapy and sub special diagnostic imaging;
   −Sub special dentistry;
   −Inpatient special non-general and sub special services like neurosurgery, ICU, CCU, etc.

How policies are formulated to definite priority programs?

There is no clear and defined process and structure to define priority programs. The Cabinet, Supreme Council for Health and/or the Minister of Health and Medical Education and technical committees make decisions as to the priority programs. However, this process can be bypassed from time to time through direct suggestion or direct notification by the minister or deputy ministers. This situation is due to the lack of formal and standardized processes.

How many policy objectives have definitive implementation plans?

It is not possible to have a clear answer to this question. Although, we can note that based on the documents of the 1st, 2nd and 3rd National Development Plans, there will be one to seven comprehensive and executive plans for all of the communicated policies, which will be developed with the cooperation of all experts and managers of the MOHME and MPO. The important issues are that in some cases these plans are not designed according to the existent resources. In addition, in some cases there is not a strong commitment for implementation of these plans. Furthermore, there is a significant lost time between notification of the policies and the development of the implementation plans.

CONCLUSION

Assessment of the health care governance in the Islamic Republic of Iran by the framework which has been developed by the WHO shows obvious strengths and a few specific weaknesses. These weaknesses are mostly in the process of implementation of the devised strategies and also in gathering the resources necessary for their implementation.

Another weakness is the lack of interaction between the stakeholders and the contribution of the social groups and entities involved in the decision-making processes. Although, the intersectional efficiency is good, but intersection cooperation is not suitable.

Furthermore, employment of non-medical personnel in management and planning of health care policies is not appropriate. The formation of social entities for defending the consumer rights in health care sector and transparency of decision-making and resource allocation is weak, too. Another problematic area is the appointment and promotion procedures of health care managers. In addition, the performance evaluation of such employees is not systematic.

In regard to the low level income population, even though there is a national commitment to this segment of population, but the mechanism for identification and prioritization of different income groups and resource allocation based on such patterns has not been established. However, the existence of health care network and health insurance program and also the expanded activity of the social protection organizations have made it possible for the poor and low level income groups to have access to the basic health care services. It is also important that the information system to be renovated and be equipped to generate systematic and up-to-date reports about different health issues and diseases.

Overall, the strategic vision of the health care governance of the I.R. of Iran has not been established based on a comprehensive and preplanned model, but it has been formed in response to different problems and challenges of current situation. Even though, this model has had a remarkable efficiency and major achievements during the past three decades, it should be reengineered based on a comprehensive model. The suggested framework for health governance by the WHO can be very useful for this purpose. It can be used as a blue print for the comprehensive health care governance for any country through assessing the strengths and weaknesses and devising strategies to alleviate the weaknesses.

The executive team of this project will strive to present the results of this research to the senior managers of the MOHME to try to plan for improving of the health care governance in the I.R. of Iran.

Acknowledgement:

REFERENCES

1. Health system achievements in previous 8 years, MOHME, 2003
2. Performance report of Ministry of Health during the first FYNDP, MOHME, 1993
3. The most important activities of Ministry of Health in 1993
    4. The most important activities of Ministry of Health in 1994
    5. Performance report of Ministry of Health during the third FYNDP, MOHME, 2001
    6. The world health report, WHO, 1995 to 2005
    7. Act of Budget for Islamic republic of Iran, MPO, 1991 to 2001
    8. Subtraction report of Budget, parliament library, 1995 to 2001
    9. Country Economical report, MPO, 1991 to 2001
    10. National Account report, SCI, 1991 to 2001
    11. Yearly Statistical report, SCI, 1991 to 2001
    12. Trend of two decades of Health system budgets, MOHME, 2004
    13. Trend of resources allocation in Health system from 1986 to 2004, MOHME, 2004
    14. First Five year national development plan, MPO, 1982
    15. Second Five year national development plan, MPO, 1987
    16. Third Five year national development plan, MPO, 1998
    17. Fourth Five year national development plan, MPO, 2004

---

[i] Yearly Economic report 2000 to 2004- National Management & Planning Organization-

[ii] Kazemianejad, M.- National Health Account report 1380- Ministry of Health & Medical Education-

[iii] Yearly Economic report 2000 to 2004- National Management & Planning Organization

[iv] Physical Environment Office , Management Development Deputy, Hospital Bed Distribution Report , 2006

[v] World Health Organization- World Health Report 2000- Health systems: Improving performance. Geneva

[vi] WHO EMRO- Good governance for improving health system performance: A methodological framework to assess governance in health- non-distributed technical paper- 2005

[vii] Same reference

DO MEDICAL MALPRACTICE PAYMENTS SHORTCHANGE WOMEN?

Vasanthakumar N. Bhat, Lubin School of Business, New York, NY 10038 (vbhat@pace.edu)

ABSTRACT

In this paper, we analyze medical malpractice payments by gender and probe whether mean payments to women are lower than those of men. Our analysis accounts for the state where malpractice happened, age and years of experience of physicians, outcomes, age of patients, and nature of allegations. We examine the data for the period from 2004 to 2009. Our analysis indicates that the average medical malpractice payments to women are lower by 4.34 percent than those of men.

Introduction

This paper examines whether there are differences in malpractice payments made to men and women. We use the data from National Practitioners Data Bank for the years from 2004 and 2009. Our analysis accounts for the state where the malpractice happened, age and years of experience of physicians, outcomes, age of patients, and nature of allegations. In this paper, we first provide mean payments by states, mean payments by graduation years of physicians, mean payments by nature allegations, mean payments by malpractice consequences or outcomes, mean payments by the nature allegations, and patient's age. We also present multiple regression analysis output with states as fixed variables and years as random variables.

Payment by State : Except in 6 states, average payments to men exceed those of women.

State	Mean Payment -Men	Mean Payment – Women	Ratio
SOUTH DAKOTA	**224812**	**274622**	**0.818623**
NEVADA	**295451**	**332607**	**0.888289**
TENNESSEE	**282612**	**298119**	**0.947984**
NEW HAMPSHIRE	**348287**	**362165**	**0.96168**
NEW MEXICO	**235876**	**244418**	**0.965052**
NORTH CAROLINA	**348482**	**356155**	**0.978456**
WEST VIRGINIA	254130	252550	1.006256
MASSACHUSETTS	513359	504847	1.016861
MISSISSIPPI	288310	283106	1.018382
OHIO	349122	340717	1.024669
KANSAS	172794	168097	1.027942

DISTRICT OF COLUMBIA	448267	435346	1.02968
TEXAS	206458	198283	1.041229
OKLAHOMA	274119	261409	1.048621
MAINE	351307	334603	1.049922
KENTUCKY	294837	277286	1.063296
INDIANA	282237	261873	1.077763
UTAH	246347	224309	1.098248
PENNSYLVANIA	373056	337949	1.103883
NEW YORK	473613	426766	1.109772
ARIZONA	367029	329955	1.112361
MISSOURI	339058	301964	1.122842
ILLINOIS	638668	565001	1.130384
GEORGIA	393667	347193	1.133856
SOUTH CAROLINA	206031	181032	1.138092
NEW JERSEY	433685	379393	1.143102
NORTH DAKOTA	319837	279674	1.143606
WASHINGTON	350086	302971	1.15551
IDAHO	276370	235909	1.171511
ARKANSAS	356367	300909	1.184302
VIRGINIA	391478	330398	1.184868
NEBRASKA	193125	162371	1.189406
LOUISIANA	211129	176905	1.19346
ALABAMA	401272	333728	1.202392
MARYLAND	406446	331978	1.224316
FLORIDA	301137	242219	1.243243
OREGON	399925	320164	1.249125
MICHIGAN	184388	146470	1.258879
COLORADO	369638	291429	1.268364
RHODE ISLANDS	397205	310561	1.278992
CALIFORNIA	252244	196914	1.280986
MONTANA	327759	252376	1.298693
PUERTO RICO	84223	63090	1.334966
HAWAII	519503	382148	1.359429
MINNESOTA	487770	356786	1.367122
CONNECTICUT	716488	510183	1.404375
IOWA	365405	258811	1.41186
DELAWARE	597381	421122	1.418546
WISCONSIN	615584	407657	1.510054
ALASKA	420723	277244	1.517519
VIRGIN ISLANDS	414609	243272	1.704302
WYOMING	528690	264607	1.99802
VERMONT	438246	200668	2.183936

Payments by Graduation Year of Physicians: In all cases, average payments to men exceed those of women by all physicians graduated in all decades.

Graduation Year	Men	Women	Ratio
1930-39	503604	338706	1.49
1940-49	267008	231254	1.15
1950-59	327344	292931	1.12
1960-69	338562	307554	1.10
1970-79	347250	299504	1.16
1980-89	359456	317083	1.13
1990-99	346070	308689	1.12
2000-current	289369	245202	1.18

Payments by nature allegations – Average payments to men exceed those of women for all types of allegation.

Nature of allegation	Mean	Mean	Ratio
Diagnostic	355312	342140	1.04
Anesthetic	429332	382807	1.12
Surgery	303490	262739	1.16
Medication	257524	244336	1.05
Blood	302800	200698	1.51
Obstetrics	708426	497645	1.42
Treatment	286514	250083	1.15
Monitoring	349012	310701	1.12
Equipment	163957	150226	1.09
Misce	234030	170360	1.37

Payments by outcome – Except in emotional, major temporary, and minor permanent injuries, payments to men exceed those of women.

OUTCOME	Men	Women	Ratio
Emotional	**82964**	**111623**	**0.74**
Insignificant	55766	48901	1.14
Minor Temporary	98249	93548	1.05
Major Temporary	**175784**	**193928**	**0.91**
Minor Permanent	**198039**	**204431**	**0.97**
Significant Permanent	382171	383143	1.00
Major Permanent	555980	513271	1.08

Quadriplegic, Brain Damage, Lifelong	886244	831132	1.07
Death	324533	316731	1.02

Mean Payments by Patient's age– Except for age 1 to 19, mean payments to men exceed of those of women.

Patient's age	Men	Women	Ratio
>1	375802	357757	1.05
1 to9	**321125**	**327841**	**0.98**
10 to 19	**293919**	**308695**	**0.95**
20 to 29	335471	329841	1.02
30 to 39	352804	306734	1.15
40 to 49	338380	293334	1.15
50 to 59	290870	264791	1.10
60 to 69	216724	200173	1.08
70 to 79	145004	143467	1.01
>80	106481	89239	1.19

Putting it all together, we perform a regression analysis using the years as a random variables and states as fixed variables and using logarithm (base e) of malpractice payment as a dependent variable. The result is as follows:

VARIABLE	Estimate	t Value
Intercept	8.2777	0
Physician Age Age 30-39	0.09951	1.6
Age 40-49	0.1005	1.59
Age 50-59	*0.1163*	*1.82**
Age over 60	0.05784	0.9
Patient Age 1-19	0.1971	12.96***
20-39	0.2623	19.22***
40-59	0.2916	23.72***
Physician Experience Less than 20	0.1123	0.8
21 to 30	0.1347	0.96
Over 30	0.1271	0.9
Nature of Allegation Diagnostic	0.3891	13.76***

Anesthetic	0.3686	9.74***
Surgery	0.5119	17.94***
Medication	0.1553	4.65***
Blood	0.02157	0.2
Obgyn	0.6097	18.93***
Treatment	0.1894	6.55***
Monitor	0.1802	4.86***
Settled out of court	-0.6415	24.57***
Outcome Emotional	-0.4718	12.7***
Significant permanent	1.0841	78.97***
Major permanent	1.4563	91.13***
Life long	1.9	86.65***
Death	1.0274	86.93***
Metro	-0.00156	0
Periodic Payment	0.000136	0
FEMALE	**-0.04442**	**4.87***

***significant at 0.01 level * significant at 0.10 level.

Based on the regression output, it can be concluded that the average medical malpractice payments to women are lower by 4.34 percent than those of men. This result is statistically significant at the .01 level.

ORGANIZATIONAL CULTURE, TENURE AND WILLINGNESS TO PARTICIPATE IN CONTINUOUS IMPROVEMENT PROJECTS IN HEALTHCARE

Marco Lam, York College of PA
441 Country Club Road, York, PA 17402, 717-815-1585, mlam@ycp.edu

Dan Robertson, WellSpan Health
1001 South George Street, York, PA 17405, 717-812-4076, drobertson10@wellspan.org

ABSTRACT

In this paper we investigate the factors that impact employees' willingness to participate in continuous improvement projects. We surveyed 226 professionals employed by a large healthcare organization in the northeastern United States. Results indicate that tenure, prior experiences, and perceived organizational culture are the main drivers in predicting willingness to participate in continuous improvement projects. Consistent with continuity theory, we do not find any differences based on age. While we find that employees are more willing to participate in an organization that supports change, of less importance is whether employees perceive that approach to be systematic.

Keywords: quality management, organizational culture, survey results

TYPE OF PERFORMANCE MATTERS: ANOTHER
LOOK AT THE WORK ATTITUDE—JOB PERFORMANCE RELATIONSHIP

MARK JOHN SOMERS
NEW JERSEY INSTITUTE OF TECHNOLOGY AND RUTGERS BUSINESS SCHOOL

DEE BIRNBAUM
RHODES COLLEGE

ABSTRACT

This study explored relationships between organizational commitment, occupational commitment, job involvement, work group commitment, and job satisfaction and three measures reflecting incidents of unacceptable job performance: verbal warnings, written warnings, and suspensions with pay. There was no relationship between work attitudes and verbal warnings or work attitudes and suspension with pay. However, both job satisfaction and occupational commitment were negatively related to supervisor issued written warnings. These findings were discussed in terms of theory development and with respect to their practical implications.

INTRODUCTION

Although there has been consistent progress in many domains within human resources management and organizational behavior, areas where research has not progressed are also evident. One area that has remained problematic is the work attitude---job performance linkage in that hypothesized relationships have been inconsistent and much less robust than expected. Put simply, theory based predictions that work attitudes are meaningfully related to job performance have not been supported despite repeated tests over a long period of time.
A wide range of work-related attitudes are hypothesized to be positively related to job performance including organizational commitment [21], job involvement [19], job satisfaction [11] occupational commitment [[4] and work group commitment [6]. In general, empirical testing has yielded disappointing results characterized by modest, inconsistent relationships between these two variables [6, 8, 9, 12, 16].

This overall disappointing pattern of findings has led to a rethinking by some scholars of the attitude—performance relationship with attention directed toward both variables. With respect to work attitudes, the general expectation of a robust positive relationship with job performance has been tempered such that relationships are seen as more modest and more targeted [7]. That is, rather than operating from the general view that attitudes influence performance, the question

has been restated to identify which work attitudes and under which conditions they influence job performance.

Similarly, the notion of job performance has been rethought to go beyond supervisory-rated task proficiency. Indeed, job performance has been viewed in broader terms to include exemplary performance that includes extra-role behaviors, and incidents of unacceptably poor performance that are detrimental to organizations and to individuals [3]. Although supervisors might be involved with and assess such incidents of job performance, they are not part of the normal performance appraisal process, but often have a much greater effect on individuals and organizations than does the "standard" annual performance review [20].

Empirical research is indicative of a constricted criterion space that is exacerbated by a limited conceptual frame with respect to predictor variables. With regard to the latter, most research on the work attitude—job performance relationship is restricted to studies of only one work attitude (e.g.. job satisfaction). It is, therefore, not clear if and how constellations of work attitudes affect job performance. With regard to the former, most studies use supervisory-rated job performance as the criterion variable so that relationships between work attitudes and other facets or types of job performance remain largely unexplored [20].

This study addresses both of these gaps in the literature by studying several work attitudes in relation to non-conventional measures of job performance reflecting incidents of unusually poor job performance. We chose this dimension of job performance because it is under researched and for the practical benefits of indentifying predictors of unacceptable job performance.

WORK ATTITUDES

The attitude—performance relationship is grounded in the comprehensive process models that typify research on work attitudes. These multi-stage process models begin with identification of relationships between a given work attitude and its hypothesized antecedents and end with hypothesized relationships between that work attitude and work outcomes [21,22].   Job performance is invariably included as an outcome variable and hypothesized relationships are justified on the characteristics of the psychological state that defines each work attitude. For example, the strong emotional ties that characterize affective organizational commitment are thought to translate into job performance [18] although these relationships have proven to be weak in practice [16]. Similarly, psychological immersion in work stemming from job involvement is thought to be lead to positive relationships with job performance, although these relationships have proven to be even weaker than were those observed for organizational commitment [19]. Similar findings are evident for other work attitudes including job satisfaction [12] occupational commitment [4] and work group commitment [6].

Not surprisingly, these findings were seen as cause for concern especially since they lead to questions about the efficacy of the discipline [7]; that is, it is seen by some scholars as embarrassing that theory based predictions that are fundamental to the fields of human resources management and organizational behavior are consistently disconfirmed. It is especially problematic because job performance sits at the nexus of several critical human resources

management areas including employee training and socialization, leadership, productivity, and goal attainment.

Attempts to remedy this state of affairs led to some rethinking of the outcome variables that work attitudes are more likely to affect such as employee turnover and organizational citizenship behavior [17] and generated greater emphasis on establishing boundary conditions for the work attitude—job performance relationship [7].

JOB PERFORMANCE

Although job performance is widely viewed in terms of supervisory-rated task proficiency at the operational level, at the conceptual level, job performance is seen as an amalgam of distinct, but related constructs [2].. The multi-faceted character of job performance has been recognized for some time, but research in human resources management is characterized by an emphasis on task proficiency usually as rated by one's supervisor despite ongoing concerns about the validity of this performance metric [2, 3].

Recent developments in theory and in the practice of human resources management have led to what has been termed by Borman [3] as an enlarged criterion space; that is, dimensions of job performance other than task proficiency are increasingly seen as important to understanding human performance at work. One element of this enlarged criterion space is focused on work behavior that extends beyond proficiency in the tasks that comprise one's job [3]. This broader notion of performance has its roots in prosocial behavior in organizations and is characterized by activities that benefit individuals, work groups or the entire organization.

In addition to behavior that is beneficial to organizations, interest has also been expressed in behavior that is detrimental to individuals and to organizations; that is, unacceptable performance that is harmful to individuals, groups or to the entire organization [3]. It has been suggested that when jobs are structured and have clearly defined procedures, disciplinary actions are an acceptable index of this dimension of performance and there is evidence of discrimimant validity for this measurement model [3]. Specifically, disciplinary actions were correlated more strongly with supervisory ratings related to reprimanding employees than they were with respect to supervisory ratings of task proficiency.

THE STUDY

This study explores the relationship between a constellation of work attitudes and performance that is detrimental to both individuals and to organizations. In so doing, research on work attitudes and job performance is reframed by exploring a broader array of work attitudes in relation to a more focused criterion. Specifically, the study is focused on unacceptable job performance as indicated by specific behavioral incidents that require disciplinary action. These behaviors, in turn, represent a dimension of o job performance that can be used to gain new insights into the attitude—performance relationship.

As such, this study is relevant to theory development and has practical applications as well. The multi-dimensional nature of job performance has not been fully recognized in human resources management research in general and in studies of work attitudes in particular [3]. Boundary conditions seem to be a useful concept in understanding attenuated attitude—performance relationships, but for this modification of current theory to be meaningful, it is necessary to identify their parameters; that is, to gain some insights into the conditions under which relationships between work attitudes and job performance might be more pervasive or more robust than what is likely to be indicated by a more general conceptual model.

This study, in turn, explores poor job performance as a potential boundary condition. It is possible that a sense of alienation or lack of engagement stemming from poor work attitudes might manifest itself in incidents of poor job performance rather than by affecting task proficiency. That is, the non-significant relationships observed in prior studies might be a function of the type of job performance that was studied.

To get a sense of which work attitudes might be relevant to understanding incidents of poor job performance, a wide range was included in this study. Specifically, we were interested in relationships between organizational commitment, job involvement, work group commitment, occupational commitment, and job satisfaction and job performance.

The practical implications of capturing relationships between work attitudes and poor performers are also evident. Examining patterns of relationships between work attitudes and incidents of poor job performance is useful in capturing predictors of exceptionally poor job performance which can then be used to develop programs and policies to limit such behavior in future.

METHOD

Sample

The sample was comprised of 228 staff nurses in a large hospital in the southern region of the United States. All participants were full-time employees. The study was sponsored by management and was conducted on-site during normal working hours using a self-report questionnaire. Participation was voluntary and 100 percent of those nurses asked to complete the survey agreed to do so. The sample as 92 percent female with a mean age 39.1 years, and an average organizational tenure of 9.01 years. It was 63 percent white and 32 percent black. Seventy one percent of the sample held bachelor's or graduate degrees in nursing.

Measures

Affective, continuance and normative commitment to the organization were measured with Allen & Meyer's [1] scales. (α = .83, .68, and .85 respectively). Occupational commitment was measured with Blau, Paul and St. John's [5] scale. This measure captures the affective component of occupational commitment. (α = .83) Job involvement was measured with Kanungo's [13] scale as adapted by Blau, Paul and St. John [5] which taps psychological immersion in one's work. (α = .62) Work group commitment was assessed with an 6 item *a priori* measure focused on commitment to the work group. The scale taps affective attachment to the work group. A sample question was "I feel that I am part of my work group." (α = .77) Job satisfaction was measured with Quinn & Staines [10] five item measure (α = .87)

Responses were along 5-point Likert-type scales for each commitment measure except for work group cohesion which was measured with a 7-point Likert-type scale.

Job performance was measured by tallying the number of verbal and written warnings and the number of incidents where employees were suspended with pay that were issued by nursing supervisors during the preceding twelve months. Data were taken from employee personnel records. Data were coded as binary (presence/absence) so that means reflect the percentage of the sample that had received verbal and/or written warnings or who were suspended. Each of these events fell outside of the normal performance appraisal process and was triggered by a discrete incidence of poor performance that was deemed unacceptable (e.g., medication error) or by a pattern of behavior that was persistent and unacceptable because it affected the quality of care delivered (e.g., failure to follow procedures, excessive, unexplained absenteeism).

RESULTS

Descriptive Statistics

Means, standard deviations among study variables is presented in Table 1.

Table 1. Descriptive Statistics

Variable	Mean	sd
Work Attitudes		
Affective Org. Comm.	3.30	.72
Continuance Org. Comm.	3.30	.62
Normative Org. Comm.	3.61	.80
Job Involvement	2.63	.63
Work Group Comm.	4.05	1.20
Occupation Comm.	3.64	.66
Job Satisfaction	3.56	1.04

Job Performance:

Verbal Warnings	.08	2.69
Written Warnings	.07	.258
Suspension With Pay	.08	.279

Work Attitudes and Job Performance

Data were analyzed with binary logistic regression for each of the three performance variables. Regression models were built such that all of the predictor variables were entered in one step. Results indicated that there was no relationship between work attitudes and job performance for two of the three indices of job performance. However, two significant predictors emerged with respect to written warnings, job satisfaction and occupational commitment (See Table 3). In addition, consistent with attitudinal theory the sign of these relationships was negative indicating that positive work attitudes were associated with less likelihood of receiving a written warning. Although the level of explained variance was comparatively low, it was consistent with observed relationships between work attitudes and job performance [12]. Further, given that the low incidence of supervisory warnings and suspensions reduces statistical power, our results most likely underestimate the true relationship between work attitudes and job performance.

Table 2. Results From Logistic Regression

Verbal Warnings

Predictor	B	S.E	Sig.
Job Involvement	.094	.594	.875
Occupational Commitment	-.027	.581	.963
Affective Commitment	-.184	.651	.777
Continuance Commitment	.000	.556	.000
Normative Commitment	-.238	.534	.656
Work Group Commitment	-.274	.273	.314
Job Satisfaction	-.477	.344	.166
Nagelkerke R Square			.059

Written Warnings

Predictor	B	S.E	Sig.
Job Involvement	-.217	.529	.682
Occupational Commitment	-1.456	.595	.014
Affective Commitment	.188	.607	.756
Continuance Commitment	.392	.507	.439

Predictor	B	S.E	Sig.
Normative Commitment	-.534	.537	.320
Work Group Commitment	.337	.257	.190
Job Satisfaction	-.634	.310	.040
Nagelkerke R Square			.105

Suspension With Pay

Predictor	B	S.E	Sig.
Job Involvement	-.446	.450	.322
Occupational Commitment	-.308	.461	.505
Affective Commitment	.573	.528	.277
Continuance Commitment	-.315	.419	.453
Normative Commitment	.280	.475	.555
Work Group Commitment	-.091	.203	.655
Job Satisfaction	-.161	.269	.550
Nagelkerke R Square			.039

DISCUSSION

The somewhat disappointing pattern of results with respect to the relationship between work attitudes and job performance has led to calls for researchers to think differently; that is, rather than seeking to confirm broad, omnibus relationships between these two variables, researchers have been asked to take a different view and try to identify the conditions under which certain work attitudes might influence job performance. For example, Guion [7] has suggested using different methodologies to explore the possibility that relationships might be nonlinear. Others [3, 20] have looked to the type of job performance in question as a potential boundary condition.

This study took the latter approach and used incidents of unusually poor performance to study the attitude—performance relationship. This type of performance obviates much of the leniency error that typifies supervisory ratings of task proficiency [2]. Incidents of poor job performance are more suited to extending theory on the attitude—performance relationship than is task proficiency because they represent a new look at an old problem. Since this dimension of job performance is tied to factors that go beyond task proficiency such as one's degree of socialization by the organization, it captures elements of performance that have gone largely unstudied. Further, conceptual definitions of work attitudes align more closely with these properties of poor job performance than they do with task proficiency [16] thereby opening a new window to study the attitude—performance relationship. For example, based on theory, highly committed employees are expected to be less likely to act in ways that are detrimental to the organization than they are to have higher levels of task proficiency.

Although the overall incidence of our three indicators of job performance was comparatively low, significant relationships were observed between job satisfaction and occupational commitment for supervisory issued written warnings. Given that our sample was of staff nurses, it is not surprising that occupational commitment emerged as a predictor of poor job

performance. Further, our findings provide some support for the hypothesis that job satisfaction does affect job performance, despite a long history of spotty results at best.

Given that both occupational commitment and job satisfaction share conceptual underpinnings associated with positive affect, the overall pattern of results suggests that positive affect in organizations tends to mitigate behaviors that might be detrimental to them. Conversely, the sense of alienation that can stem from disaffection or lack of engagement might be expressed as behaviors that take the form of unacceptable job performance. Whether this is a cry for help or an attempt to damage the organization is an area for future study, but our findings suggest that the it might be more productive to refocus theory development on the negative consequences of poor work attitudes in relation to certain elements of job performance rather than on the implied spillover of positive work attitudes in relation to task proficiency.

Our results are a bit surprising in there was no relationship between organizational commitment [20] indeed, none of the three components of organizational commitment was related to job performance. At the very least, a negative relationship between affective organizational commitment and incidents of poor job performance might be expected, but apparently commitment to the occupation played a larger role in influencing job performance. Perhaps the vocational nature of the sample was a factor and different results might have been observed for a more general employee population.

The practical implications of these findings are evident and they suggest that workers with positive work attitudes are assets to organizations, while those who are disaffected, can engage in behavior that is detrimental to individuals and to organizations. Thus, while some managers might be inclined to view this type of job performance as isolated events restricted to a few employees, our findings suggest that these are not random events, but rather are indicative of employees who are poorly socialized as indicated by low levels of occupational commitment and job satisfaction. Organizations and managers would do well to take these factors into account in hiring decisions and in managing employees on a day-to-day basis.

These results are, of course, very preliminary, but they are encouraging given the comparatively low incidence of poor performance. We were also unable to make comparisons with supervisor-rated task proficiency which is a second limitation of the study, and results must be interpreted with these limitations in mind. Regarding future research, this type of performance data is difficult to gather so that studies of poor job performance are relatively uncommon. It is, however, in the interest of organizations to make these data available for research. Perhaps once they see the advantages of doing so, studies of this type will be more common.

REFERENCES

[1] Allen, N. & Meyer, J. (1990). *The Measurement and Antecedents of Affective, Continuance and Normative commitment to the Organization*. Journal of Occupational Psychology, 63, 1-18.

[2] Astin, W. (1964). *Criterion Centered Research*. Educational and Psychological Measurement, 24, 807-822.

[3] Borman, W. (1990). Job Behavior, Performance and Effectivness. In Dunnette, M. & Hough, L. (eds.) Handbook of Industrial-Organizational Psychology. Palo Alto, CA: Consulting Psychologists Press.

[4] Blau, G. (1994). *Testing for a Four Dimensional Structure of Occupational Commitment*. Journal of Occupational and Organizational Psychology, 76, 469-488.

[5] Blau, G, Paul, A. & St. John, N. (1993) *On Developing a General Index of Work Commitment.* Journal of Vocational Behavior, 42, 298-314.

[6] Fornes, S., Rocco, T. & Wollard, K. (2008). *Workplace Commitment: A Conceptual Model Developed From Integrative Review of the Research.* Human Resource Development Review, 7, 339-357.

[7] Guion, R (1992) Agenda for Research and Action. in P. Smith (Ed.), Job Satisfaction: How People Feel About Their Jobs and How It Affects Their Performance. (pp. 257-281). New York: Lexington Books.

[8] Iaffaldano, M. & Muchinsky. P. (1985) *Job Satisfaction and Job Performance: A Meta-Analysis. Psychological Bulletin, 97, 251-273.*

[9] Mathieu, J. & Zajac, D. (1990). *A Review and Meta-Analysis of the Antecedents, Correlates and Consequences of Organizational Commitment*. Psychological Bulletin, 108, 171-194.

[10] Quinn, R. & Staines, G. (1979) The 1977 Quality of Employment Survey. Ann Arbor, MI: University of Michigan, Institute for Social Research.

[11] Locke, E.A. (1970) *Job Satisfaction and Job Performance: A Theoretical Analysis.* Organizational Behavior and Human Performance, 10, 375-412.

[12] Judge, T., Bono, J., Thoresen, C., & Patton, G. (2001). *The Job Satisfaction-Job Performance Relationship: A Qualitative and Quantitative Review*. Psychological Bulletin, 127, 376-407.

[13] Kanungo, R. (1982) *Measurement of Job and Work Involvement.* Journal of Applied Psychology, 67, 341-349.
[14] Lee, K., Carswell, J. & Allen, N (2000). *A Meta-Analytic Review of Occupational Commitment: Relations with Person and Work-Related Variables*. Journal of Applied Psychology, 85, 799-811.

[15] Mathieu, J. & Zajac, D. (1990). *A Review and Meta-Analysis of the Antecedents, Correlates and Consequences of Organizational Commitment.* Psychological Bulletin, 108, 171-194.

[16] Meyer, J. & Allen, N. (1997). Commitment in the Workplace: Theory, Research and Application. Thousand Oaks, CA: Sage.

[17] Meyer, J. & Herscovitch, L. (2001). *Commitment in the Workplace: Towards a General Model.* Human Resources Management Review, 11, 299-326.

[18] Meyer, J., Paunonen, S., Gellatly, I., Goffin, R., & Jackson, D. (1989). *Organizational Commitment and Job Performance: It's the Nature of Commitment that Counts*. Journal of Applied Psychology, 74, 152-156.

[19] Rabinowitz, S. & Hall, D. (1977*). Organizational Research on Job Involvement.* Psychological Bulletin, 84, 265-288.

[20] Somers, M. & Birnbuam, D. (1998). *Work-Related Commitment and Job Performance: It's Also the Nature of the Performance That Counts*. Journal of Organizational Behavior, 19, 621-634.

[21] Steers, R. (1977). *Antecedents and Outcomes of Organizational Commitment*. Administrative Science Quarterly, 22, 46-56.

[22] Wiener, Y. & Yardi, Y. (1980). *Relationships between Job, Organization, and Career Commitments and Work Outcomes: An Integrative Approach*. Organizational Behavior and Human Performance, 26, 81-96.

THE ORGANIZATIONAL CONTEXT AND ECONOMIC BENEFIT FOR WORKSITE HEALTH PROMOTION PROGRAM

Hsin-Chih Kuo, I-Shou University,
No. 8, Yida Rd., Yanchao Township, Kaohsiung County 824, Taiwan, R.O.C.
886-7-6151100 ext.7411
e-mail: simon@isu.edu.tw

Chi-Hsing Tseng, National Pingtung Institute of Commerce,
No. 51, Min Sheng E. Rd., Pingtung 900, Taiwan, R.O.C.
886-8-7238700 ext. 6215
e-mail: tseng@npic.edu.tw

ABSTRACT

In order to improve the health and productivity of employee and reduce the medical expense, enterprises pay attention to worksite health promotion programs. When worksite health promotion programs are carried out, enterprises will face several inconsistencies among individual, departmental, and organizational levels. Enterprises also need to coordinate the activities among different functional departments. Otherwise they perhaps lose the expected benefits and cause the disturbance of related departments. This research adopted the case study method and took the medical staff in a hospital as the target of this study. The researcher interviewed several medical professionals and used secondary data as the multiple sources. This study found the influential factors and the economic benefit of the worksite health promotion program. The influential factors mainly contained the coordination among individual, departmental, and organizational levels. The economic benefit is obvious for the people who execute the primary activity of value chain in the organization and the worksite health promotion program could be regarded as an important practice of human resource management.
Keyword: worksite health promotion, organizational context, economic benefit

INTRODUCTION

According to the government's regulation, enterprises must setup a safe and healthy working environment for employees. For example, employee assistance programs improve the health status and reduce health risk for employee. The attitude of enterprise has changed from the consideration of the amount of money to the evaluation of the worksite health promotion program. The worksite health promotion program can improve the productivity of employees, reduce the medical expenditure of company, and increase the net profit of company (Hollander & Lengermann, 1988). The worksite health promotion program will be accomplished through the interdisciplinary integration and the determinants of health contain health risk and employee target behavior (Chu, Driscoll, & Dwyer, 1997). The success of worksite health promotion program needs to go through organizational development and will be influenced by physical environment, organizational structure, and administration.

THEORETICAL BACKGROUND

The health promotion plan can be performed in the school, community, hospital, or worksite. The worksite has the advantage to implement the health promotion program for a lot of advantage in worksite (Fielding & Piserchia, 1989). For example, many people spend a large amount of time in work place and the employers have economic incentives to invest and to improve employees' health.

The implementation of worksite health promotion program
The goal of occupational safety and health (OSH) was to prevent the accident and occupational disease. Based on the occupational safety and health, the worksite health promotion (WHP) explored a new opportunity to maintain the occupational health and safety and to become part of business strategy (Breucker & Schroer, 1999). The characteristics of OSH is top-down and that of WHP is bottom-up respectively.

A survey of U.S.A. nationwide worksite health promotion programs in 1985 is as follows (Fielding & Piserchia, 1989): 65.5% enterprises which have over 50 employees will provide more than one health promotion programs. These programs included the evaluation of health risk, smoking cessation, blood pressure control and treatment, exercise/fitness, weigh control, nutrition education, stress management, back problem prevention and care, and off-the-job accident prevention. Another survey of U.S.A. worksite health promotion program in 1999 is as follows (Wilson, DeJoy, Jorgensen, & Crump, 1999): (1) 25% small-scale enterprises which employee's number from 15 to 99 persons would offer health promotion plans. (2) 44% enterprises which employee's number more than 100 employees would offer health promotion program. (3) Most worksite health promotion programs were concerned with occupational safety and health, back injury prevention, and CPR. The scope of worksite health promotion program implementation in Taiwan is more extensive. For example, the topic of worksite health promotion programs contained worksite health psychology, work stress, healthy hospital, program evaluation, and so forth.

The organizational context of worksite health promotion program
From the organizational view, comprehensive worksite health promotion program would be more suitable. The reasons are as follows (Weiner, Lewis, & Linnan, 2009): (1) Organizational structure and division will cause the inconsistencies among program decision maker, program executive, and program user when worksite health program is been implemented. (2) The division of labor produces the differentiation which needs more interdependence among different tasks. (3) Senior managers expect that comprehensive health promotion program could produce collected benefits, such as improvement of employee's health, increase of productivity, decrease of health care cost, and so forth. When organizations implement the comprehensive worksite health promotion programs, they must face two key tasks: (1) Program should be fit with the organizational context. (2) The employees should participate in the health promotion programs (Weiner, et al., 2009).

The case study of a Swedish enterprise showed that employee's health is concerned with organizational structure, culture, and leadership (Eriksson, Jansson, Haglund, & Axelsson, 2008). Some scholars tried to find the determinants of health promotion program implementation in the social ecology and organizational theory. These four

determinants included organizational predisposition, organizational capacity, internal organizational factors, and external system factors (Riley, Taylor, & Elliott, 2001).

Economic benefit of worksite health promotion program
The implementation of worksite health promotion program is usually assessed by the improvement of health related indexes such as lifestyle, target health behavior (diet, exercise, smoking, drinking, weigh loss), or physical examination value (body mass index, blood pressure, cholesterol, triglyceride). However, the assessment is gradually shifted to economic benefit.

Sick leave. The reduction of sick leave day could be converted into monetary unit for the assessment of economic benefit. A comparison of sick leave between participation and non-participation in worksite health promotion programs during six years was held in a manufacturing company which has 4189 male employee. The result showed that the reduction of sick leave day could save 623,040 dollars per year (Schultz et al., 2002).

Productivity. The relevant cost of employee's worksite health promotion program could be measured by individual productivity. For example, the absence or disability resulted in the lost time which would reduce the productivity of employee (Burton, Conti, Chen, Schultz, & Edington, 1999) .

Return on investment. Some study used the return on investment (ROI) to measure the economic benefit of employee worksite health promotion program. Comprehensive worksite health promotion program could reduce 14% disability days of blue collar in two years and would gain 2.05 dollars per one dollar invested (Bertera, 1990). ROI could also be measured the benefit for business health management program (Ozminkowski et al., 1999), for the medication and behavior treatment in smoking cessation program (Javitz et al., 2004), for the employee's health risk reduction and health care cost saving (Goetzel, Ozminkowski, Baase, & Billotti, 2005)

Cost-effectiveness. Another measurement of worksite health promotion program is to use the cost-effective analysis. In this way, a study compared the cost-effectiveness of four interventions in a worksite wellness program for cardiovascular disease and found that health education plus follow-up consult would be better than health education or fitness equipment offered only (Erfurt, Foote, & Heirich, 1992).

No matter worksite health promotion programs measured by lifestyle, target health behavior, ROI, or cost-effectiveness, the measurement reflected the diverse assessment and had the trend of economic benefit amount.

METHODOLOGY

This study adopted a case study design and attended to find the influential factors and economic benefit of worksite health promotion program. The target of the case study is the physicians within a hospital which is a large scale hospital and has more than 1000 beds in Taiwan. Triangulation was obtained from multiple sources which come from top management team, human resource unit, and physicians. Moreover,

the multiple sources were collected into a single database for the case study (Yin, 2003).

FINDING

Generally speaking, influential factors of worksite health promotion programs included the individual health cognition, available time, and expenditure. This study indicated that the effectiveness of health promotion program in the hospital was impacted by individual health cognition and available time. From the view of the hospital, the object of worksite health promotion program is to develop and to maintain the human resource within the organization. The hospital had a strong motive to implement the worksite health promotion program for physicians since they were the most important human resources for the value activities in the hospital. To be the leader of medical staff, the physician's health status is more emphasized by the hospital. This target population of this study was physicians and could be viewed from different perspectives.

Individual level

When physicians faced the health or medical problems, the attitudes of physicians would solve it by oneself, consult other physicians, or just delay. It's interesting for the health behavior of physicians. The health behavior of a physician may be based on locus of control, cognition, past experience, or demographic variables. When a physician is diagnosed to suffer from some disease, the disease is often very serious or in the late stage rather than the early stage. The reasons may be as follows:

- Physicians who provide medical services may have overconfidence on their health status and diseases progress.
- When physicians give patients medical information or treatment, they usually develop isolation mechanism between patients and themselves.
- Physicians may have some opportunism for their occurrence of some diseases.
- Physicians also have occupational or life pressure that would cause insomnia, alcohol dependence, or suicide.

The health examination checks blood pressure, triglyceride, or ultrasound scan. It is simple and has a good value for early diagnosis without an expensive fee. The health examination should be the most available thing for a physician's health promotion program, but most physicians didn't take the routine health examination. The reasons may be as follows:

Health cognition. The health cognition of physicians may be the important factor that affects the implementation of worksite health promotion programs for physicians. Owing medical background, physicians may have overconfidence and believe that they can handle out their health by themselves. Most physicians could quickly approach health care resource because the medical resources are available in the hospital. Physicians usually ignore their own health and don't have a good plan for their health status.

Available time. Physicians, especially surgeons, usually have emergency operations that would disturb their health examinations. For the surgeons, there are a lot of activities to be done before, during, and after the surgery. If the worksite health

promotion program could be expected to be implemented effectively, the program should be incorporated into the schedule of ordinal activities for surgeons. For example, the time period between morning meeting and ward visiting could be used to take the simple health examination.

Expenditure. The expenditure on health examination for physicians could be viewed as the benefit for physicians in the hospital. It also could be viewed as an investment in human resource for maintenance of the most precious human capital. The human resource of physicians is valuable, scare and non-substitute. The experienced core human resource in the hospital should be maintained and developed. It is obvious that the hospital could benefit from the investment in the worksite health promotion of physicians.

Departmental level
According to the specialty, the physicians in a hospital could be divided into internal medicine, surgery, and other specialties. No matter internal medicine or surgery, most physicians need to carry out some operations in clinical. The clinical processes are different between internists and surgeons. However, the difference between internists and surgeons mainly lies on the operation technical complexity, the uncertainty of the operational result, and the percentage of operation time spent in clinical practices.

The surgeons have more complicated operations, high operation result uncertainty, and high percentage of clinical working time than internists. Therefore, the surgeons are hard to implement a regular activity of health promotion since the operations of surgery time are usually urgent. If the worksite health promotion program wants to be implemented, the coordination among different department should be considered such as outpatient department, inpatient department, operation room, emergency room, and intensive care unit.

Organizational level
If the organization has a valuable, rare, inimitable human resource, the organizational would implement the worksite health promotion program to maintain the human resource. From the view of the hospital, the physician who owns high skill is difficult to be substituted. The hospital would attend to improve the health status of physicians by the way of preventive medicine since the productivity of the hospital would be affected by physicians. For example, the hospital would encourage physicians to accept the regular health examination and to shape their health behavior.

To avoid interrupting the clinical care and operation, the hospital won't aggressively implement health promotion program for physicians. If the organization wants to implement the worksite health promotion program effectively, it would consider the coordination among different departments such as outpatient department, inpatient department, operation room, emergency room, and intensive care unit. The worksite health promotion program should be modified to match the process for physicians and it would be implemented more effectively.

Economic benefit
From the view of economic benefit, the health examination is just the beginning of a worksite health promotion program for physicians. The cost of health examination

program can be divided into direct medical cost, direct non-medical cost, and indirect cost.

- *Direct medical cost.* Direct medical cost includes all cost produced by the medical staff such as health examination fee.
- *Direct non-medical cost.* Direct non-medical cost includes all cost produced by the non-medical staff such as travel fee or other related expense.
- *Indirect cost.* Indirect cost includes time cost produced by the non-working time of employees.

The benefit of health examination program resulted from the reduction of sick leave day. If the sick leave day would be reduced one day per month, the productivity could be gained twelve working days per year. The benefit minus the cost is equal to net gain.

Benefit (the reduction of sick leave and the promotion of productivity) － Cost (direct and indirect cost) directly = Net gain

When the worksite health promotion program was implemented for the physicians, the economic effect was obvious. The hospital would view the worksite health promotion program as the most important program for human resource maintenance and gain the most economic benefit.

CONCLUSION

If the worksite health promotion program is expected to be implemented effectively, the government can make a policy to facilitate the worksite health promotion program. Worksite health promotion that is bottom-up can be accomplished by the empowered employees. From the policy view, the successful implementation will be achieved by the combination of private and public interests. Government can ask organizations to take part of medical expense for employees' sickness. If organizations want to reduce total costs, they will pay attention to the disease prevention and intend to implement worksite health promotion program. Through the way of cost reduction for the private interest, the goal of public interest could be achieved simultaneously.

REFERENCES

Bertera, R. L. (1990). The effects of workplace health promotion on absenteeism and employment costs in a large industrial population. *Am J Public Health, 80*(9), 1101–1105. doi: 10.2105/ajph.80.9.1101

Breucker, G., & Schroer, A. (1999). Concept and principles of workplace health promotion. *Promotion & Education, 6*(3), 3–8. doi: 10.1177/102538239900600302

Burton, W. N., Conti, D. J., Chen, C.-Y., Schultz, A. B., & Edington, D. W. (1999). The role of health risk factors and disease on worker productivity. *Journal of Occupational and Environmental Medicine, 41*(10), 863–877.

Chu, C., Driscoll, T., & Dwyer, S. (1997). The health-promoting workplace: An integrative perspective. *Australian and New Zealand Journal of Public Health, 21*(4), 377–385.

Erfurt, J. C., Foote, A., & Heirich, M. (1992). The cost-effectiveness of worksite wellness programs for hypertension control, weight loss, smoking cessation, and exercise. *Personnel Pyschology, 45*(1), 5–27.

Eriksson, A., Jansson, B., Haglund, B. J. A., & Axelsson, R. (2008). Leadership, organization and health at work: a case study of a Swedish industrial company. *Health Promot. Int., 23*(2), 127–133. doi: 10.1093/heapro/dan015

Fielding, J. E., & Piserchia, P. V. (1989). Frequency of worksite health promotion activities. *American Journal of Public Health, 79*(1), 16–20. doi: 10.2105/ajph.79.1.16

Goetzel, R. Z., Ozminkowski, R. J., Baase, C. M., & Billotti, G. M. (2005). Estimating the return-on-investment from changes in employee health risks on the dow chemical company's health care costs. *Journal of Occupational and Environmental Medicine, 47*(8), 759–768.

Hollander, R., & Lengermann, J. (1988). Corporate characteristics and worksite health promotion programs: survey findings from Fortune 500 companies. *Social Science & Medicine, 26*(5), 491–501.

Javitz, H., Swan, G., Zbikowski, S., Curry, S., McAfee, T., Decker, D., . . . Jack, L. (2004). Return on investment of different combinations of bupropion SR dose and behavioral treatment for smoking cessation in a health care setting: an employer's perspective. *Value in Health, 7*(5), 535-543.

Ozminkowski, R. J., Dunn, R. L., Goetzel, R. Z., Cantor, R. I., Murnane, J., & Harrison, M. (1999). A return on investment evaluation of the Citibank, N.A., health management program. *American Journal of Health Promotion, 14*(1), 31–43.

Riley, B. L., Taylor, S. M., & Elliott, S. J. (2001). Determinants of implementing heart health promotion activities in Ontario public health units: a social ecological perspective. *Health Educ. Res., 16*(4), 425–441.

Schultz, A. B., Lu, C., Barnett, T. E., Yen, L. T.-c., McDonald, T., Hirschland, D., & Edington, D. W. (2002). Influence of participation in a worksite health-promotion program on disability days. *Journal of Occupational and Environmental Medicine, 44*(8), 776–780.

Weiner, B. J., Lewis, M. A., & Linnan, L. A. (2009). Using organization theory to understand the determinants of effective implementation of worksite health promotion programs. *Health Educ. Res., 24*(2), 292–305. doi: 10.1093/her/cyn019

Wilson, H. S. D. M. G., DeJoy, P. D. M., Jorgensen, D. C. M., & Crump, B. S. C. J. (1999). Health promotion programs in small worksites: results of a national survey. *American Journal of Health Promotion, 13*(6), 358–365.

Yin, R. (2003). *Case study research: Design and methods*: Sage Publications, Inc.

The Impact of Social Health on the Level of Job Satisfaction: A Case Study of Iran Khodro Leasing Company

Maryam Sam Aram

Allameh Tabatabaie University
Tehran, Iran

ABSTRACT

A major factor contributing to the survival of any organization is the existence of healthy workers, not only physically healthy, but also healthy in a holistic manner. The holistic health of workers in an organization has been the focus of numerous researches by scholars of management science along different dimensions including, physical, psychological and spiritual health. These studies have examined the impact of such dimensions on various aspects of the work environment. However, social health, as one of the most important dimensions, has not been considered as a significant factor by previous literature in organizational science. According to Keyes (1984), a factor that can impact an overall health of an individual is the sense of social well-being or social health. Keyes defines social well-being along different dimensions: *"Social integration, social acceptance, social contribution, social actualization* and *social coherence"* (Keyes, 1998).

This study builds upon Keyes' theory and presents a model for testing the impact of social health on job satisfaction. The social health construct is operationalized along several dimensions: Social integration, social acceptance, social contribution, social actualization and social coherence. In addition, several hypotheses exploring the impact of social health on job satisfaction are developed.

The model developed for this research is tested in an Iranian corporation, Iran Khodro Leasing Company. Iran Khodro Leasing Company is a financial and credit institution specialized in leasing automobiles. Currently, the company is trying to expand its market share nationally and internationally. The social health and the level of job satisfaction are measured through a survey questionnaire provided to the employees of the company. Multiple regression analysis is used to examine the collected data. The results of the statistical analysis supports several hypotheses developed in this research.

Key words: Social Health, Social Integration, Social Acceptance, Social Contribution, Social Actualization, Social Coherence, Job Satisfaction.

EVOLUTIONARY GAME ANALYSIS ON CYBER SECURITY INVESTMENT

Rong Pan, School/College of Business, Hohai University & Bryant University, rpan@bryant.edu
Chen Zhang, Ph.D., College of Business, Bryant University, Smithfield, RI, USA, czhang@bryant.edu
Changxin Xu, Ph.D., School of Business, Hohai University, Nanjing, China, xuchxin@hhu.edu.cn

ABSTRACT

In this paper, we propose an evolutionary game model to analyze the investment decision making process of the cyber offender-defender interaction from a microeconomic perspective. Our evolutionary game model shows that the cyber offender-defender game can possibly reach two stable equilibrium points after a long-term evolution. One of them implicates a defender-dominant game while the other implicates an offender-dominant game. We found that an offender-dominant game can be avoided by maintaining the security investment above a threshold level determined by the system vulnerability among other environmental parameters such as residual risk and potential loss. Hence with an appropriate level of security investment, the defender can manipulate the game effectively to discourage attacking attempts. Our numeric calculation implicates the goal of leading the cyber security environment towards a benign defender-dominant one may not be beyond the possible.

Keywords: Cyber Security, Investment Decision, Evolutionary Game Theory

INTRODUCTION

In recent year, the prevalence, frequency and severity of cyber attacks are growing as indicated in annual CSI Computer Crime and Security Survey [19]. Major cyber attacks are also growingly costly to the victim organizations. Recent survey conducted by McAfee shows that a 24-hour down time from a major attack was estimated to cost US$6.3 million on average and as much as US$8.4 million a day in the oil/gas sector [15]. The actual frequency of security incidents and associated financial loss might be even higher since organizations may not report all cyber security incidents due to reputational and other concerns as Gordon and Loeb [9] point out.

Cyber security investment, therefore, becomes a critical decision in organizational IT investment. Practitioners currently apply Return on Security Investment (ROSI) as a cyber security investment metric --- 67.8 percent in 2009, over 44 percent in 2008, trying to find out the appropriate level of investment to prevent the recurring events of security breaches [19].

In this paper, we propose an evolutionary game model to analyze the investment decision making process of the cyber offender-defender interaction from microeconomic perspective. The offender-defender recursive interaction is considered to decide the threshold and expected levels of cyber security investment in self-protection. In making this decision, the defender needs to balance the investment in self-protection with the reduction in the risk of becoming the victim of a compromise.

Our Evolutionary Game Theory (EGT) model shows that the offender-defender game can possibly reach two stable equilibrium points after recursive interactions. One of them implicates a defender-dominant game while the other implicates an offender-dominant game. Hence we believe, with an appropriate level of security investment, the defender can manipulate the game effectively to discourage attacking attempts.

The rest of the paper is organized as following: In Section 2, we describe the existing work on cyber security investment. Section 3 introduces the background of our research initiatives. In Section 4 we develop an EGT model to discuss the offender-defender interaction. In Section 5, we discuss how defender can reverse the

offender-dominant game by appropriately investing in self-protection and provide some quantitative examples. Lastly we present the conclusion and future research.

RELATED WORK

Security researchers have generally reached a consensus that the existing security problems cannot be solved by technological means alone. Since technological means cannot eliminate all risks, attention is increasingly paid to quantifying the cyber security investment [7] [8] [10] [13] [16], employment of cyber insurance to transfer residual risks [1] [2], the impact of information sharing on cyber security and how interdependency influence investment incentive of agents [7] [12] [18]. However, most existing research is from the defender or defenders dependency perspective. Very little research has been conducted from the offender-defender game perspective.

Gordon and Loeb [10] presented an economic model to determine the optimal information security investment amount. Their research shows that information assets with midrange vulnerabilities rather than the highest vulnerability are worthier of protection and a firm should invest only a relatively small portion of the expected loss due to a security breach. Hausken [11] analyzed how income, interdependency, and substitution effects impact security investment for organizations of different sizes. Ogut and Menon [18] holds the view that interdependency of cyber-risk reduces firms' incentives to invest in security technologies and to buy insurance coverage. Cavusoglu and Mishra [3] proposed a comprehensive analytical model to balance cost and utility of IT security measures for investment decision support.

Due to the strategic nature of security issues, some researchers also investigate the problem from game-theoretic perspective. Cavusoglu and Raghunathan [4] compared game theory and decision theory methods from several dimensions including investment levels, vulnerability and investment payoff. Garcia and Horowitz [6] established a game-theoretic model to analyze the economic motivations for investment in improving cyber security. Their research found that if the ratio of social value to revenue at stake of Internet service providers continues to grow, the likelihood of underinvestment in security becomes higher. Kunreuther and Heal [12] established a parametric game-theoretic model to address the situation when the security choices by one agent affect the risks faced by others.

Next, we describe the background of our research initiatives.

BACKGROUND OF OUR RESEARCH INITIATIVES

Cavusoglu's paper [4] discussed the limitations of other traditional models. Because of the strategic nature of security issues, we believe game theory is more suitable to model the offender-defender interaction.

In addition, classical game theory requires perfect rationality, which may not be practical in all cyber attacks. This is partially due to the asymmetry of game information and partially due to the automation of many cyber attacks. If one player adopts irrational behavior, then the classical game theory might fail. Hence, we adopt the EGT, which is better equipped to deal with the bounded rationality assumptions and is able to capture the recursive strategic interaction between the offender and the defender.

Another distinction of EGT from traditional game theory is that it allows evolving strategies which is more practical for cyber security investment decisions because of the information asymmetry. Evolutionarily Stable Strategy (ESS) and Replicator Dynamics (RD) are two vital concepts in evolutionary game theory. ESS is a mathematical definition for an optimal choice of strategy, which can resist invasion by any mutant strategy. RD is a set of deterministic difference or differential equations, which describes the proportion of a population of players adopting a strategy.

Hypotheses

We assume there are two types of risk-neutral players with bounded rationality in the cyber security investment decisions. One is the population of defenders who decide whether to invest in improvement of cyber infrastructure self-protection and the other is the profit-driven population of offenders who needs to weigh cost against return to decide whether to conduct cyber attacks. Their decisions are discrete, the defender chooses whether or not to invest in self-defense and the offender chooses whether or not to perform an attack.

Organizations victimized from a cyber compromise suffer damages. These damages can be tangible or intangible: financial loss due to fraudulent transactions, interruption of business operations, loss of sensitive data, loss of reputation and customer confidence, to name a few. We assume the total financial loss is L.

The ability of an organization to defend against cyber attacks depends on its level of self-protection investment with the assumption that the appropriate defense technology is applied. Before the defender makes additional cyber security investment, the probability of a successful attack completely depends on the inherent vulnerability, v, of the original system. We define $p(v)$ to be the probability of a compromise before investment in self-protection. After the defender invests I in self-protection of its cyber infrastructure, the offender has a probability r of success. Obviously, $r<p(v)$. Kunreuther [12] assumes that damages can only occur once, i.e., damages resulting from multiple security failures are no more severe than that of a single failure. We assume the expected financial loss of defender before and after the cyber security investment are $p(v)L$ and rL respectively, with the impact of the severity of the vulnerability v but without the impact of the frequency of attacks.

Defender invests in self-protection with capital I while offender incurs a cost of C to attack. Assume the offender can derive return R out of a successful attack, then $rR-C$ represents his payoff. The profit-driven offender will not perform attack if he cannot profit from the action. Therefore, $p(v)R-C>0$. With the black market of cyber attacks maturing, the cost of offender is constantly reducing with the high return rate encouraging the observable rapid growth of such a cyber black economy [14] [19].

It is an appointed task for defender to protect its cyber infrastructure and information, therefore making cyber security investment cannot bring any immediate profit to him (ignore the long-term security reputation gain). Thus, if the defender invests to protect its cyber infrastructure, when the attacker succeeds, the defender will face a loss of rL. So his expected payoff is $-rL$. On the contrary, if the defender does not invest in cyber security, he can save the additional cyber security investment I and therefore when suffering attack, his payoff is $I-p(v)L$.

Obviously, the offender can gain nothing if he does not perform attack. In that case, the payoff of the offender is 0. Likewise, if the offender never attacks, the defender's expected payoff is 0 and I respectively when he does and does not invest in self-protection. Hence, we introduce the payoff matrix of the offender-defender game (in normal form) as:

TAB.1 EXPECTED PAYOFF MATRIX OF THE EVOLUTIONARY GAME BETWEEN THE POPULATION OF OFFENDER AND THE POPULATION OF DEFENDER

Offender \ Defender	Invest	Do Not Invest
Attack	$rR-C$, $-rL$	$p(v)R-C$, $I-p(v)L$
Do Not Attack	0 , 0	0 , I

Evolutionary Game Analysis

Assume that in the population of offender, the proportion who adopts the "Attack" strategy is α. Thus, the proportion who adopts the "Do not Attack" strategy is $(1-\alpha)$. Meantime, we denote by parameter β the proportion in the population of defender who adopts the "Invest" strategy while by $(1-\beta)$ the proportion who adopts the "Do not Invest" strategy.

Hence, the expected utility of offender population when the "Attack" and the "Do not Attack" strategy are adopted U_{oa}, U_{od} as well as the average utility of the population $\overline{U}_o$, respectively, are

$$U_{oa} = -\beta R[p(v)-r] + p(v)R - C \qquad (1)$$

$$U_{od} = 0 \qquad (2)$$

$$\overline{U}_o = \alpha\{-\beta R[p(v)-r] + p(v)R - C\} \qquad (3)$$

The expected utility of defender population who adopt the "Invest" and the "Do not Invest" strategies are U_{di}, U_{dd} respectively, together with the average utility of the population $\overline{U}_d$ are

$$U_{di} = -\alpha r L \qquad (4)$$

$$U_{dd} = -\alpha p(v)L + I \qquad (5)$$

$$\overline{U}_d = -\alpha\beta r L + (1-\beta)[-\alpha p(v)L + I] \qquad (6)$$

Replicator dynamics (RD) of this evolution system can be obtained by applying RD equation to the two populations of players - offender and defender:

$$\left.\begin{array}{l} \dfrac{d\alpha}{dt} = \alpha(1-\alpha)\{-\beta R[p(v)-r] + p(v)R - C\} \\[2mm] \dfrac{d\beta}{dt} = \beta(1-\beta)\{\alpha L[p(v)-r] - I\} \end{array}\right\} \qquad (7)$$

By the stability principle of differential equation and the definition of Evolutionarily Stable Strategy (ESS), it is known that when $\beta=(p(v)R-C)/(R(p(v)-r))$, $d\alpha/dt$ always equals to 0. That is, all α are in the stable state. When $\beta>(p(v)R-C)/(R(p(v)-r))$, $\alpha^*=1$ and $\alpha^*=0$ are two stable states of α. At this time, $\alpha^*=0$ is in the stable state. When $\beta<(p(v)R-C)/(R(p(v)-r))$, $\alpha^*=1$ and $\alpha^*=0$ are still two stable states of α. But $\alpha^*=1$ is in the stable state at this time. Figure 1 depicts phase diagrams and stable states when α evolves under the aforementioned three conditions.

Likewise, when $\alpha=I/(L(p(v)-r))$, all β are in the stable state. When $\alpha> I/(L(p(v)-r))$, $\beta^*=1$ and $\beta^*=0$ are two stable states of β. At this time, $\beta^*=0$ is in the stable state. When $\alpha< I/(L(p(v)-r))$, $\beta^*=1$ and $\beta^*=0$ are two stable states of β. But $\beta^*=1$ is in the stable state at this time. Figure 2 provides phase diagrams and stable states when β evolves under the aforementioned three situations.

The RD relationships between these two evolving populations can be represented in the following coordinate plane, as Figure 3 shows.

$(a)\beta = \dfrac{p(v)R - C}{R[p(v) - r]}$ $(b)\beta > \dfrac{p(v)R - C}{R[p(v) - r]}$ $(c)\beta < \dfrac{p(v)R - C}{R[p(v) - r]}$

FIG.2 REPLICATOR DYANMAICS PHASE DIAGRAM OF THE DEFENDER POPULATION IN THE ASYMMETRIC OFFENDER-DEFENDER GAME

$(a)\alpha = \dfrac{I}{L[p(v) - r]}$ $(b)\alpha > \dfrac{I}{L[p(v) - r]}$ $(c)\alpha < \dfrac{I}{L[p(v) - r]}$
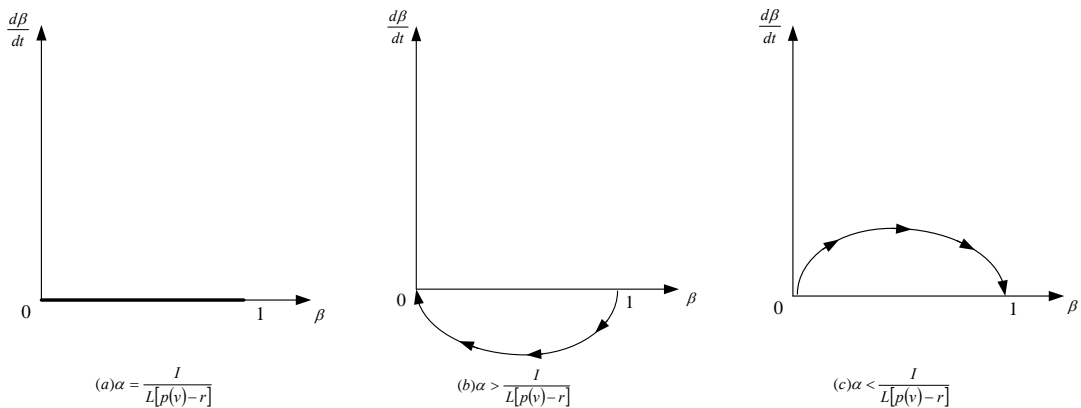
FIG. 3 REPLICATOR DYANMAICS AND STABILITY FOR TWO PLAYER POPULATIONS IN THE ASYMMETRIC OFFENDER-DEFENDER GAME
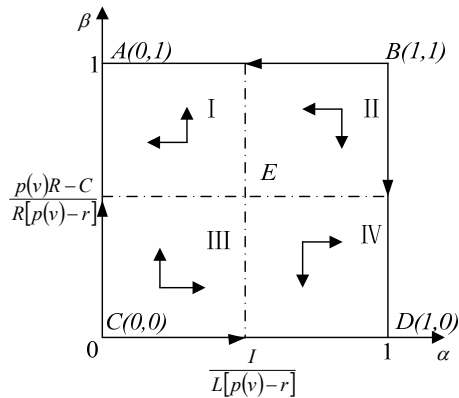
According to flow directions of arrows in Figure 3, it can be found that $\alpha^*=1$, $\beta^*=0$ and $\alpha^*=0$, $\beta^*=1$ are stable states in this asymmetric offender-defender game. It can be observed from Figure 3 that, when the initial state falls into District Ⅰ, it will converge into the stable state $\alpha^*=0$, $\beta^*=1$. That is, the population of offender will all adopt the "Do not Attack" strategy while population of defender will all choose the "Invest" strategy. This is a defender-dominant game. When the initial state falls into District Ⅳ, it will converge to the stable state $\alpha^*=1$, $\beta^*=0$. That is, the population of offender will all adopt the "Attack" strategy while the population of defender will all adopt the "Do not Invest" strategy. This is the offender-dominant game. When initial conditions fall into District Ⅱ and District Ⅲ, the game will converge to either stable state $\alpha^*=\beta^*=0$ or $\alpha^*=\beta^*=1$.

Results

Obviously, the evolutionary strategy of $\alpha^*=0$ and $\beta^*=1$ is the ideal outcome, that is, the population of defender tends to insist on investment in cyber security and this group strategy effectively discourages the population of offender to attack. Therefore this offender and defender game reaches an equilibrium beneficial to the defender. In practice, what can we do to impel the evolutionary strategy $\alpha^*=0$ and $\beta^*=1$ instead of a black hat community dominant game? What is the determinant to reverse the situation to a white hat community dominant game? Let's first examine the problem based on RD equation of the offender group.

In compliance with the aforementioned analysis, when the proportion of defender population choosing the "Invest" strategy reaches *(p(v)R-C)/R(p(v)-r)*, the offender-defender interplay ends up with a stable state no matter which strategy the population of offender takes. Thus, this equilibrium is the key to achieve the defender-dominant game. We denote $\bar{I}=\beta I$ as the average investment level of the entire defender group to facilitate benchmarking with the offender-dominant game. Thus,

$$\beta \times I = \bar{I} = \frac{p(v)R-C}{R[p(v)-r]} \times I \tag{8}$$

r represents the probability of being compromised after defender has invested *I* in self-protection. In other words, it is the residual risk that the defender cannot eliminate within reasonable investment in cyber security. The expression of *r* can be obtained from Equation(8), which reflects the trend that, given a certain average white hat community investment level, the increase of a specific defender's investment will reduce his residual risk.

$$r = p(v) - \frac{I[p(v)R-C]}{\bar{I}R} \tag{9}$$

In order to reverse the offender-dominant game, defender can take action in self-protection with additional cyber security investment. As discussed earlier, when $\beta>(p(v)R-C)/R(p(v)-r)$, the expected payoff of the offender proportion that adopts the "Attack" strategy is less than the average payoff of the entire offender population. Therefore, the population of offender will adjust its strategy toward "Do Not Attack" gradually. So the defender can induce the offender to adopt the "Do Not Attack" strategy by appropriately setting self-protection investment in cyber security. Because of Equation (8), the defender is expected to reach the following to obtain the dominance in the game. We name it as the expected value of cyber security investment:

$$I = \frac{\bar{I}R[p(v)-r]}{p(v)R-C} \tag{10}$$

Since the expected value of security investment will be higher, the investor might also be interested in the bottom line of avoiding an offender-dominant game. As we know, when the proportion of offender that adopts the "Attack" strategy is $\alpha=I/(L(p(v)-r))$, the game becomes offender-dominant. Since the parameter value of α is restricted within [0,1], forcing its value to be more than 1 can effectively avoid the formation of an offender-dominant game. Therefore this threshold value of defender investment should be:

$$I > L[p(v)-r] \tag{11}$$

Numeric Examples

Let's illustrate the aforementioned solution by the following numeric example. Equation (11) indicates that investment of the defender *I* rises linearly with the increase of the chance of system compromise *p(v)*. So the

increase of $p(v)$ should lead to more discouragement for offender. We denote the ratio of $d(v)=C/R=p(v)e$ as the discouragement of attack (or effectiveness of defense, both phrases include letter d) which is proportional to $p(v)$ where $(0<e<1)$. And the expected value of self-defense investment is

$$I = \frac{p(v)-r}{p(v)-p(v)e}\bar{I} \tag{12}$$

According to Gordon-Loeb model [10], the optimal investment in information security should be less than or equal to 36.79% of the loss that would be expected in the absence of any investment in security. Hence, we assume the average white hat community investment $\bar{I}=\lambda L$ is less than one third of the average loss $(\lambda<1/3)$. Hence the self-protection threshold equals to

$$I = \frac{1-\dfrac{r}{p(v)}}{1-e}\lambda L \tag{13}$$

Statistics on vulnerabilities from 1997 to 2010 are retrieved from the United States National Vulnerability Database with severity rubric of one to four and the weighted total is calculated before normalization and application to our model. Figure 4 shows the 10-year weighted total severity of vulnerability, with 2010 as a partial year at the time this paper is submitted [17].

FIG. 4 WEIGHTED TOTAL VULNERABILITY (1997-2009) IN U.S.
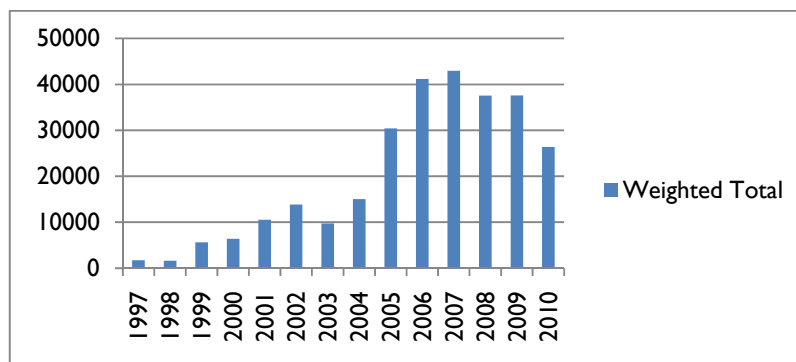
With the assumption that $p(v)$ and $d(v)$ are linear and the linear relationship between threshold value and $p(v)$ in Equation(11), we perform linear scaling of the data in Figure 4 and obtained the normalized investment threshold and expected values of the years between 1997 and 2009, with the partial year of 2010 omitted. Results are provided in Table 2 and charted in Figure 5. "Danger Zone" indicates that defender's self-protection investment is under the safety threshold and the offender-defender game might turn out to be a dangerous offender-dominant game. "Grey Zone" corresponds to the investment amount exceeding the investment threshold but has not reached the expected value to obtain the defender dominance. "Safety Zone" is when defender's investment has reached the expected value and therefore has dominated the game.

Furthermore, the real world data also indicate that it may not be a remote story for us to win the cyber war. McAfee's survey shows that 24 hours of down time from a major attack was estimated to cost U.S. $6.3 million on average which is an estimate of loss L. It is estimated by the Department of Homeland Security that the U.S. cyber security market should grow steadily with a cumulative market value of U.S. $55 billion from 2010 to 2015 [5]. Hence total cyber security investment of the country is expected to reach U.S. $9 billion in 2010 to encourage a healthy development of the defense. Hence the number of investments is estimated to reach approximately 4,886 with U.S. $1.842 million investment on average to protect against exploits upon each major zero-day vulnerability, i.e., published severe vulnerability without vendor patch yet. This estimate implicates achieving a defender dominant game may not be beyond the possible.

Tab.2 DATA AND CALCULATIONS FOR FIGURE 5 (r=0.05, e=0.3 and λ=0.25).

Year	Total Vulnerability	Chance of Compromise	Investment Threshold	Investment Expected Value
1997	1710.5	0.104812	0.054812	0.186769781
1998	1633	0.104443	0.054443	0.186167848
1999	5628	0.123467	0.073467	0.212511961
2000	6,388	0.127086	0.077086	0.216630583
2001	10,521	0.146767	0.096767	0.235472844
2002	13,832	0.162533	0.112533	0.247275059
2003	9,708	0.142895	0.092895	0.232175973
2004	15,023	0.168205	0.118205	0.250979884
2005	30,435	0.241595	0.191595	0.283229312
2006	41,165	0.292691	0.242691	0.296132635
2007	42,931	0.3011	0.2511	0.297836504
2008	37,562	0.275533	0.225533	0.292333405
2009	37,604	0.275733	0.225733	0.292380414

CONCLUSION

We have proposed an evolutionary game-theoretic model which addresses the decision making in the offender-defender interaction from a microeconomic perspective. We found that an offender dominant game can be avoided by maintaining the security investment above the threshold level determined by the system vulnerability among other environmental parameters such as residual risk and potential loss. In order to obtain defender dominance, the cyber security investment needs to further reach the expected value determined by the system vulnerability among other environmental parameters such as the residual risk and effectiveness of the defense in reducing attackers' return to cost ratio. Finally, we argue that the prevalence of recession-driven information security investment cut [15] is not wise since it might lead towards an offender dominant game with quickly deteriorating security environment for ecommerce. Our numeric calculation implicates the goal of leading the cyber security environment towards a benign defender dominant one may not be beyond the possible.

For future research, we hope to improve our model to consider more factors such as defender interdependency and find out the feasibility of applying our model in cyber insurance investment decisions to address the residual cyber security risks of organizations.

REFERENCES

[1] Bolot, J. & Lelarge, M. *Cyber Insurance as an Incentive for Internet Security*. Seventh Workshop on the Economics of Information Security, 2008, 6, 25-28.

[2] Böhme, R. & Schwartz, G. *Modeling Cyber-Insurance: Towards A Unifying Framework*. Workshop on the Economics of Information Security, 2010, 6, 1-36.

[3] Cavusoglu, H. & Mishra, B. *A Model for Evaluating IT Security Investments*. Communications of the ACM, 2004, 47, 87-92.

[4] Cavusoglu, H. & Raghunathan, S. *Decision-Theoretic and Game-Theoretic Approaches to IT Security Investment*. Journal of Management Information Systems, 2008, 25(2), 281-304.

[5] Department of Homeland Security. U.S. Federal Cybersecurity Market Forecast 2010-2015. available at http://www.marketresearchmedia.com/2009/05/25/us-federal-cybersecurity-market-forecast-2010-2015/, last accessed Nov. 2010.

[6] Garcia, A. & Horowitz, B. *The Potential for Underinvestment in Internet Security: Implications for Regulatory Policy*. Journal of Regulatory Economics, 2007, 31(1), 37-55.

[7] Gordon, L. & Loeb, M. *Budgeting Process for Information Security Expenditures*. Communications of the ACM, 2006, 49(1), 121-125.

[8] Gordon, L. & Loeb, M. *Sharing Information on Computer Systems Security: An economic analysis*. Journal of Accounting and Public Policy, 2006, 22 (7), 461-485.

[9] Gordon, L. & Loeb, M. *A Framework for Using Insurance for Cyber-Risk Management*. Communications of the ACM, 2003, 46(3), 81-85.

[10] Gordon, L. & Loeb, M. *The Economics of Information Security Investment*. ACM Transactions on Information and System Security, 2002, 5(4), 438–457.

[11] Hausken, K. *Income, Interdependence and Substitution Effects Affecting Incentives for Cyber Investment*. Journal of Accounting and Public Policy, 2006, 25, 629-665.

[12] Kunreuther, H. & Heal, G. *Interdependent Security: The Case of Identical Agents*. NBER Insurance Project Workshop, 2002, 2, 1-26.

[13] Liu, W. & Tanaka, H. *Empirical Analysis Methodology for Information Security Investment and Its Application to Reliable Survey of Japanese Firms*. Information and Media Technologies, 2008, 3(2), 464-478.

[14] Locasto, M. & Stavrou, A. *Dark Application Communities*. New Security Paradigms Workshop, Schloss Dagstuhl, Germany, 2006, 9, 11-18.

[15] McAfee. In the Crossfire: Critical Infrastructure in the Age of Cyber War, available at http://newsroom.mcafee.com/, last accessed Nov. 2010.

[16] Negrete-Pincetic, M. & Yoshida, F. *Towards Quantifying the Impacts of Cyber Attacks*. IEEE PowerTech Conference, Bucharest, Romania, 2009, 1-8.

[17] NIST Computer Security Division. United States National Vulnerability Database, available at http://web.nvd.nist.gov/view/vuln/statistics, last accessed Nov. 2010.

[18] Ogut, H. & Menon, N. *Cyber Insurance and IT Security Investment: Impact of Independent Risk*. Fourth Workshop on the Economics of Information Security, 2005, 6, 1-30.

[19] Peters, S. (2009). 2009 CSI Computer Crime and Security Survey, available at http:// pathmaker.biz/whitepapers/CSISurvey2009.pdf, last accessed Nov. 2010.

Faculty and Student Perceptions about Encryption and Data Storage

Cynthia L. Knott

G. Steube

Northeast Decision Science Institute (NEDSI)

Montreal, Canada

April 14th-16th, 2011

Abstract

The importance of securing data and information is a critical issue in today's world. These are no longer stored on a central system that is easy to protect and secure. Now everyone carries around small storage devices, which make guaranteeing that the information is guarded is much more complex and uncertain. This paper builds on the previous research of Knott & Steube's in the paper *Encryption and Portable Data Storage*, to be published in the Spring of 2011. In the previous work we identified the potential security issues that arise from using a portable storage device such as a USB flash drive. TrueCrypt software was introduced as an option that allows users to encrypt and hide data. The TrueCrypt software, which is publically available, is particularly useful for safeguarding data on USB flash drives that are easily compromised. A survey of undergraduate students was administered which focused on their practices and attitudes about security. It was found that there were strong associations between the use of flash drives, security, and the use of passwords.

This research adds another group of interest, the faculty at Marymount University. Again, their practices and attitudes about security were gathered in a survey. The major difference in this research is that, as opposed to the previous study, the faculty group was first given an informational session and tutorial about encryption and data storage before they were asked to fill out the survey. After they were given the tutorial about the encryption process, their responses were gathered. This paper includes the results and analysis of the faculty survey that determined what habits and practices they followed with respect to securing their personal data and files. Some of the questions included in the analysis are the following:

- Do you encrypt your USB flash drive?

- Do you use any type of security for your USB flash drive?

- How important do you think security is for a flash drive?

- Do you use passwords to protect your USB flash drive?

- Do you think it is important to use security when using a USB flash drive?

- Do you backup your work?

- Do you use a flash drive while teaching in the classroom or other presentations?

- So you use a flash drive for transporting data?

- After seeing how to use the encryption software, are you more likely to use it to secure your USB flash drive?

The findings indicate that faculty members are concerned about their private data. They also indicated that after seeing the tutorial on the TrueCrypt software that they are more likely to encrypt their USB flash drives. We have also found that many of the faculty members have followed up with inquiries about how to ensure that their private data is secure. This paper also includes an exploratory use of discriminant analysis to determine if the questions from the survey could be used to successfully distinguish membership in the faculty and student groups based on the answers to the first six questions of the instrument. Further research could be performed to determine if the answers to the six questions from another group such as technology professionals could be used to distinguish among the faculty, student and professional groups.

The growing use of portable data storage devices is an accepted reality in today's society (GFI Software, 2010). One type of portable data storage device in common use today is the USB flash drive or thumb drive or memory stick (PCTechGuide, 2009). Because these drives can hold an increasing amount of data and are easily ported from one location to another, many businesses consider them to be their greatest security threat (EzineArticles.com, 2010). McAfee Labs (2010) in its 2010 threats report state:

> One of the most active categories of malware this quarter was AutoRun worms (malware found on removable storage, mainly USB drives). Due to the widespread adoption of USB drives by both consumer and enterprise users around the world, this infection vector continues to be a leading source of pain. (p. 11)

This paper builds on the previous research of Knott & Steube's in the paper *Encryption and Portable Data Storage*, to be published in the Spring of 2011. In the previous work we identified the potential security issues that arise from using a portable storage device such as a USB flash drive. TrueCrypt software was introduced as an option that allows users to encrypt and hide data. The TrueCrypt software, which is publically available, is particularly useful for safeguarding data on USB flash drives that are easily compromised. A survey of undergraduate students was administered which focused on their practices and attitudes about security. It was found that there were strong associations between the use of flash drives, security, and the use of passwords.

This research adds another group of interest, the faculty at Marymount University. Again, their practices and attitudes about security were gathered in a survey. The first two sections of this paper summarizes the work that was initially conducted. The third section of this

report presents faculty responses about security and encryption and the fourth section compares

student and faculty attitudes as expressed in the questionnaire.

Using Encryption to Secure a USB Flash Drive

Typically flash drives are missing important encryption and authentication safeguards to

protect the data (IronKey, 2007). The methods recommended in this report provide encryption

and authentication through the use of a password. The solution suggested is the open source

software provided by TrueCrypt Foundation (2010b); TrueCrypt's website (www.truecrypt.org)

provides complete documentation for the process.

TrueCrypt software has the following advantages:

Creates a virtual encrypted disk within a file and mounts it as a real disk

Encrypts an entire partition or storage device such as USB flash drive or hard drive

Encrypts a partition or drive where Windows is installed

Encryption is automatic, real-time and transparent

Parallelization and pipelining allow data to be read and written as fast as if the drive was

not encrypted

Encryption can be hardware-accelerated on modern processors

Provides plausible deniability, in case an adversary forces you to reveal the

password (TrueCrypt Foundation, 2010b)

The step by step procedure for creating an on-the-fly TrueCrypt disk is fully described on

their website in the documentation section (TrueCrypt Foundation, 2010a). It is worth noting that

this approach can also be used on non-USB disk drives to secure other portable and non-portable

devices. As an open source solution, the software is free and readily available for download and

use. After encrypting a folder on the USB flash, the existence of this folder is not visible and requires a password to mount and unhide the device. If the USB flash drive with TrueCrypt software were lost or stolen, the drive with the secured data would not be visible. The folder with the hidden information can only become visible by mounting the information through a password.

This paper recommends the use of the TrueCrypt solution for securing portable data devices and in particular its use with USB flash drives is essential. This solution provides a versatile approach for safeguarding such devices. In addition, this software is regularly updated to ensure its currency and usability. TrueCrypt is a reasonable answer to both individuals who wish to secure their information as well as organizations that provide portable storage devices to their employees.

The following section of this report provides data about the perceptions of users of portable storage devices in today's world. Although the risks in using unprotected portable storage devices are manifestly clear and a solution through TrueCrypt readily available it is useful to examine behaviors and attitudes toward securing portable storage units because it is these behaviors and attitudes that will ultimately determine the actions that people will take to protect their data.

Student Perceptions and Attitudes Towards Protecting Portable Storage Devices

To investigate current attitudes toward protecting portable data devices, a survey was administered to 63 undergraduate students in business courses at Marymount University in the Fall 2010 semester. A copy of this instrument is available in Appendix A. The overall results of the survey have been tabulated by question are in Appendix B.

The relationships among the questions used in the survey were analyzed using contingency tables with a phi coefficient to indicate the relationship among the categorical data. Phi is a chi-square based measure of association; the chi-square coefficient depends on the strength of the relationship and sample size. Since phi has a known sampling distribution it is possible to compute its standard error and significance (Howell, 2002). The *PASW 19* package was used for the significance level of the computed phi value. Questions 3 and 5 had no variation in response and were not included in the phi analysis. Question 3 indicated that all participants thought that security was extremely important for a flash drive and Question 5 revealed that all participants backup their work. Consequently contingency tables were developed for Questions 1, 2, 4, and 6. Each of these questions can be represented by the variable labels listed in Table 1.

Table 1

Questions and Labels

Question Number	Corresponding Variable Label
1	Use Flash
2	Use Security
4	Use Passwords
6	Attitude Toward Security

For this analysis the strength of the association will be assessed through a rule of thumb which provides a range of values for Phi and verbal assessment. Strong negative and strong positive associations are represented by Phi values between -1.0 to .-7 and .7 to 1.0, respectively.

Weak negative and positive associations are between -.7 to .-3 and .3 to .7, respectively. Values of Phi indicating little or no association are between -.3 to .3 (Simon, 2005).

Use Flash by Use Security

The relationship between using flash and security is provided in Figure 1. The Phi value was .688 and significant at the p=.05 level. Using flash security is strongly associated with the use of security.

Figure 1. Use Flash by Use Security
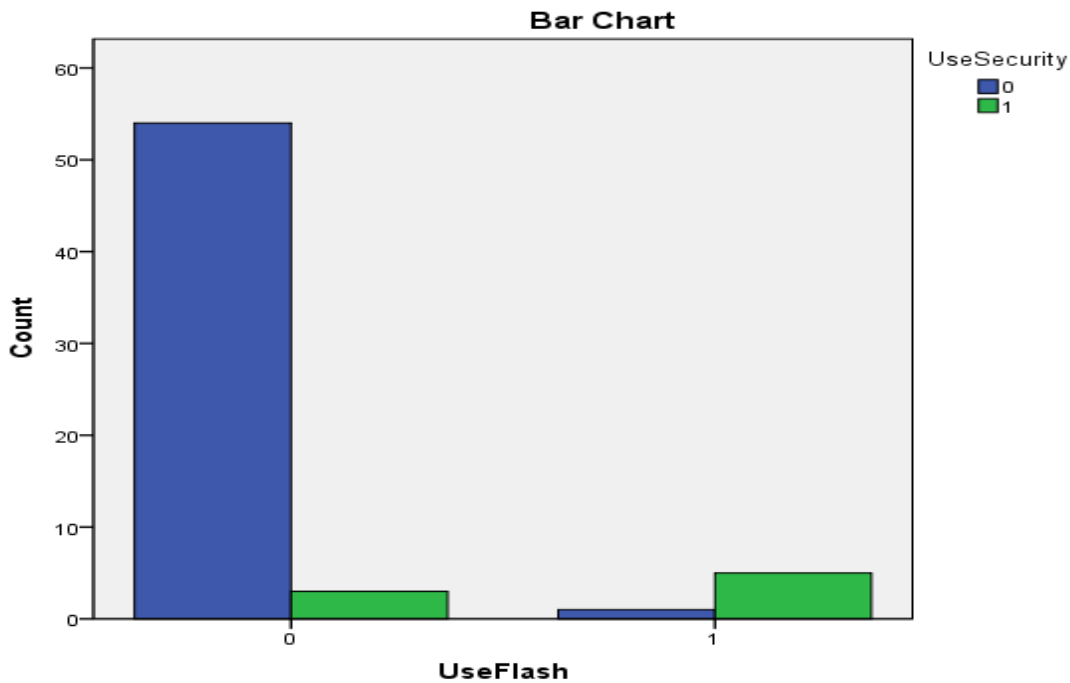
Use Flash by Use Passwords

The association between using flash and using passwords is presented in Figure 2. The Phi value was .574 and significant and the p=.05 level. The use of flash is strongly related to the use of passwords.

Figure 2. Use Flash by Use Passwords
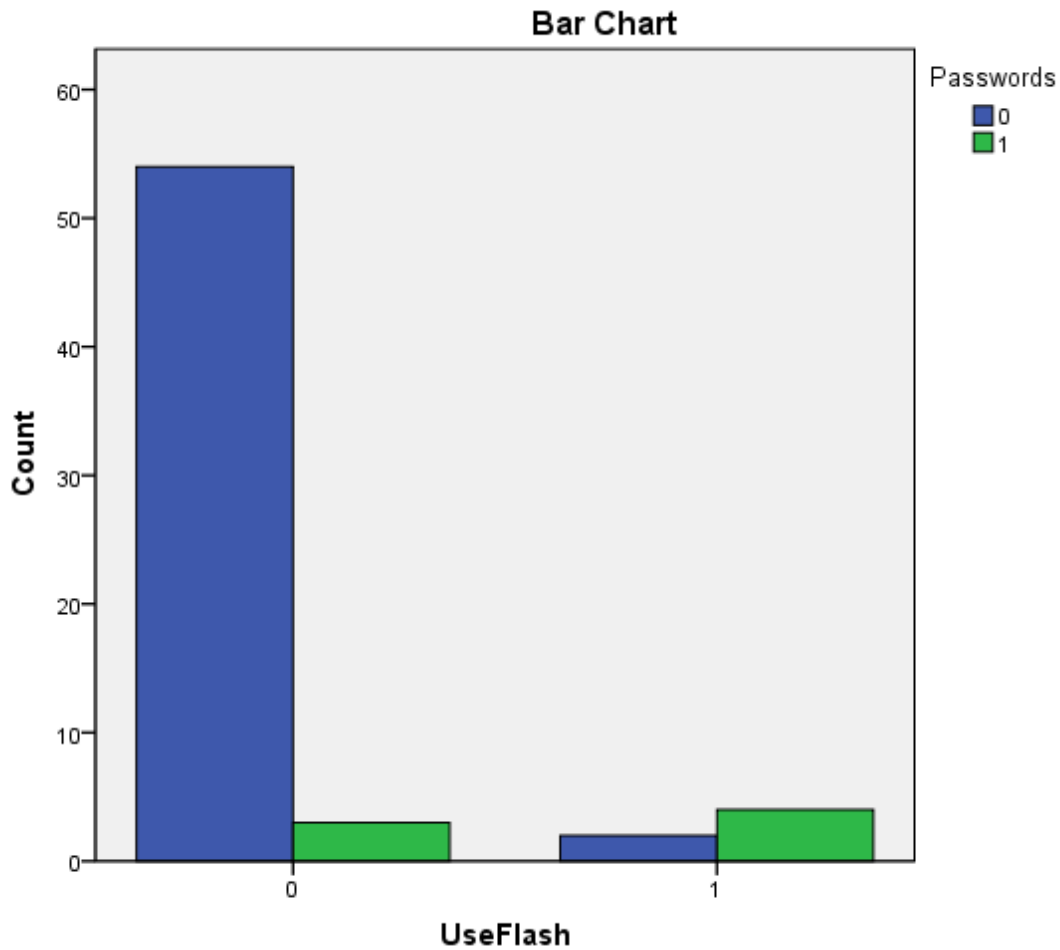
Use Flash by Attitude Toward Security

The relationship between using flash and attitude toward security is presented in Figure 3. The Phi coefficient was .229 and not significant at the p=.05. The relationship between using flash drive and the attitude toward security is a weak association.

Figure 3. Use Flash and Attitude Toward Security
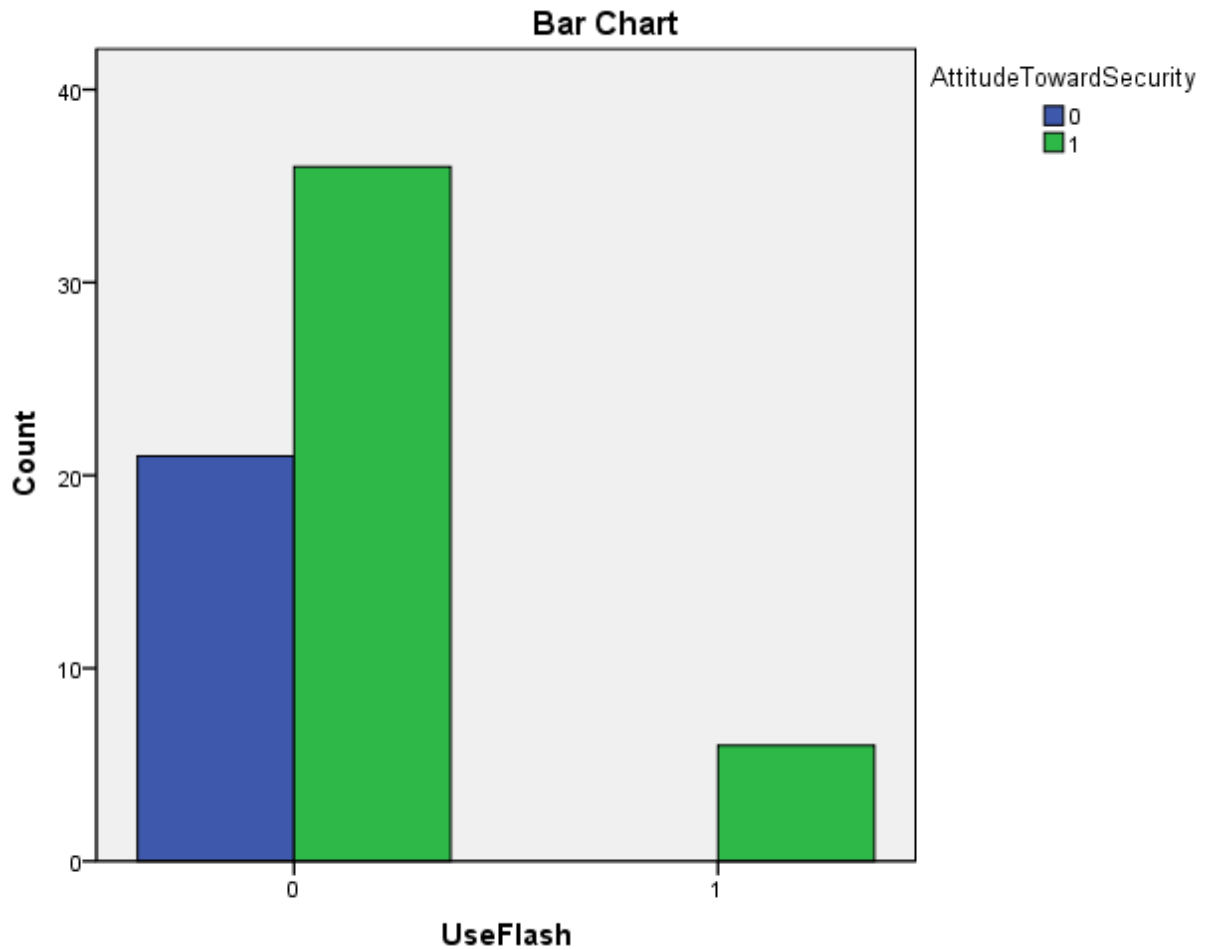
Use Security by Use Passwords

The relationship between using security and using passwords is displayed in Figure 4. The Phi coefficient was .624 and significant at the p=.05 level. Using security is strongly related to the use of passwords.

Figure 4. Use Security and Use Passwords
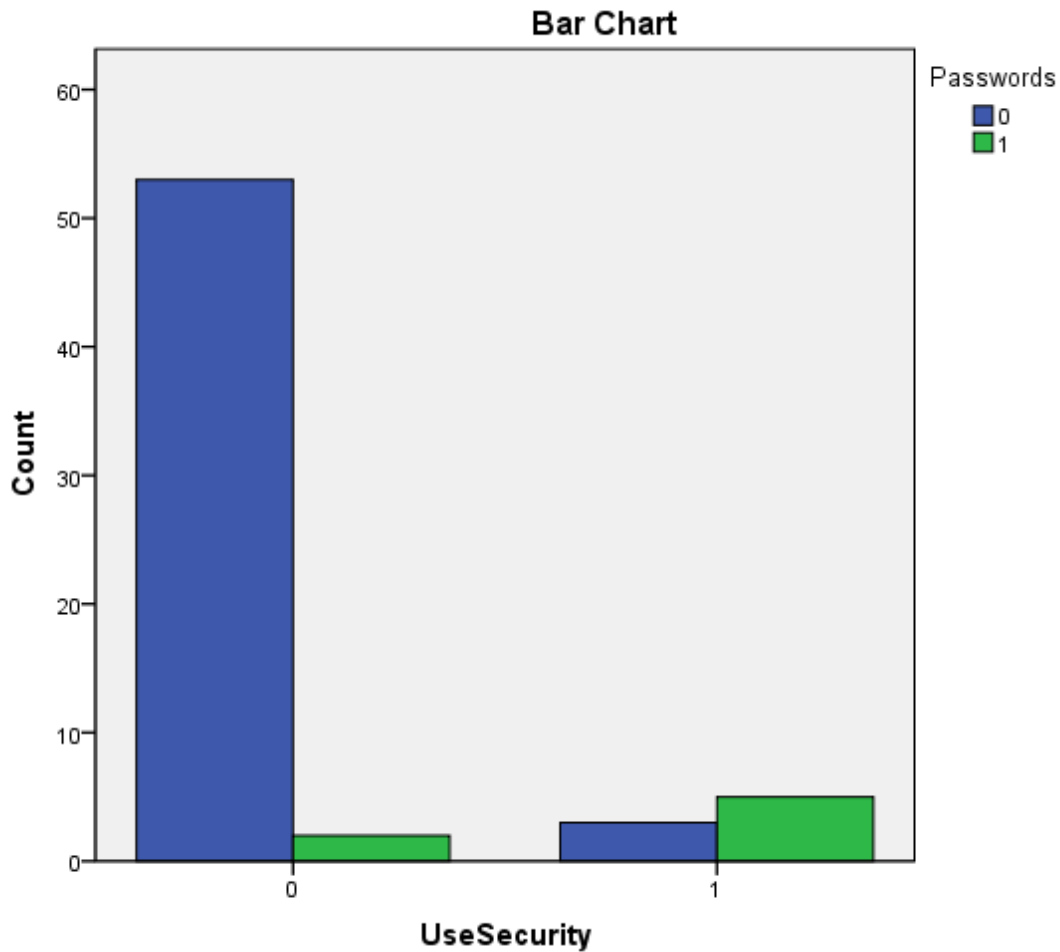
Use Security and Attitude Toward Security

The association between using security and attitude toward security is presented in Figure 5. The Phi coefficient was .169 and not significant at the p=.05 level. The use of security is only weakly related to attitude toward security.

Figure 5. Use Security and Attitude Toward Security
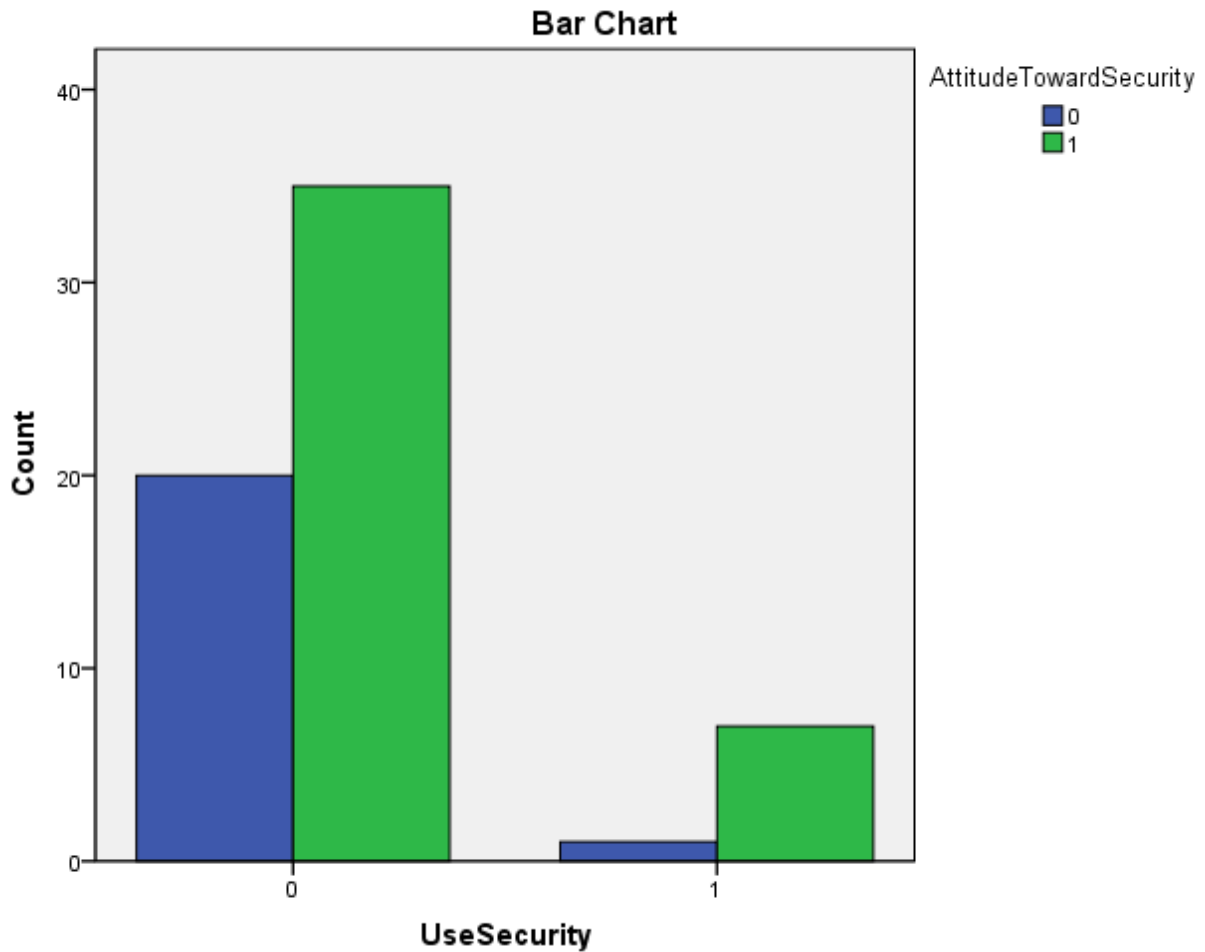
Use Passwords and Attitude Toward Security

The association between using passwords and attitude toward security is provided in Figure 6. The Phi coefficient was .143 and not significant at the p=.05 level. Using passwords is weakly associated with attitude toward security.

Figure 6. Use Passwords and Attitude Toward Security
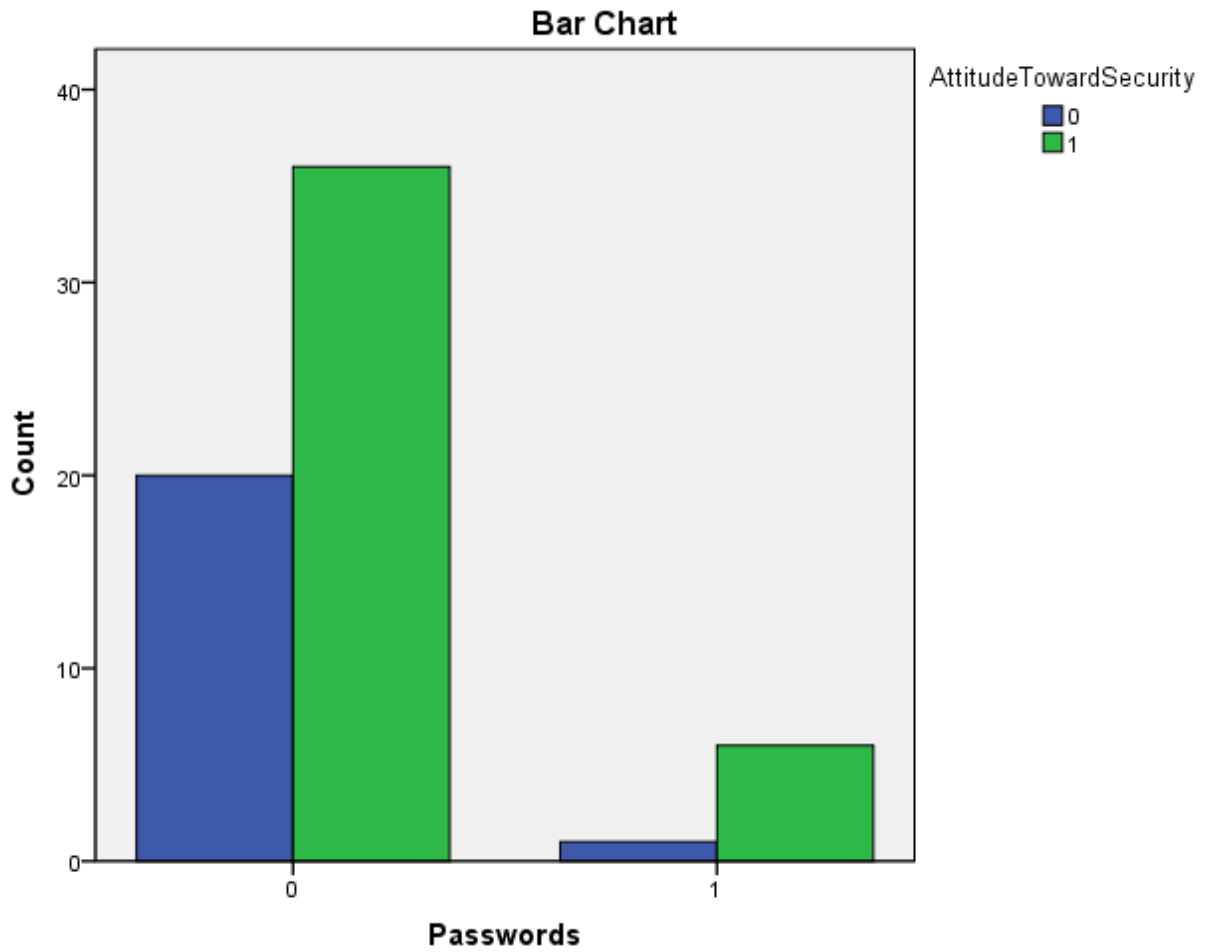
Summary of Phi Coefficients

The associations for all the variables are summarized in Table 2.

Table 2

Summary of Phi Coefficients

	Use Flash (UF)	Use Security (US)	Use Passwords (UP)	Attitudes Toward Security (ATS)
UF		.668*	.574*	.229
US	.624*		.624*	.169
UP	.574*	.624*		.143
ATS	.229	.169	.143	

* = significance at the .05 level

Faculty Perceptions and Attitudes Toward Protecting Portable Storage Devices

Faculty at Marymount University were surveyed about their practices and attitudes about security for portable storage devices. This group of faculty received a presentation and tutorial about the use of TrueCrypt for securing their USB flash drives before they completed the survey. Many of the questions used in this survey were the same as the questions posed in the student survey. The faculty survey is included in Appendix C and the results of the survey are available in Appendix D.

The first six questions in the faculty survey were the same as the first six questions in the student survey. The analysis of these questions was also performed using a descriptive summary of the results because the number of respondents was 13. There was no variation in responses for questions 1 and 2. None of the participants used encryption or any other method to secure their flash drives. The importance of security for flash drives (Question 3) is available in Table 3. About 36% felt that security was either extremely or somewhat important. Only two (15.3%) of

participants used passwords to protect their flash drives as indicated in question 4. Question 5 revealed that 12 (92.3%) of the respondents back up their flash drives. Responses to question 6 indicated that 10 (76.9%) participants believed that security for their USB drive was important.

Table 3

Question 3: How important do you think security is for a flash drive?

Importance	Number of Responses	Percent of the Total
Extremely important	4	30.8%
Somewhat important	2	15.4%
Neutral	4	30.8%
Not very important	2	15.4%
Not important at all	1	7.7%

The remaining questions in the survey were unique to the faculty participants. Table 4 displays the distribution of faculty based on their school of association in the university based on the data from question 7. The responses to question 8 showed that eleven (84.6%) use a flash drive for classroom teaching. All respondents indicated in question 9 that they transport data with a USB device. All participants agreed that they are more likely to encrypt their portable drive after viewing the presentation on TrueCrypt.

Faculty and Student Perceptions and Attitudes Toward Protecting Storage Devices

This section compares student and faculty perceptions with regard to security for their flash drives. The comparison is done in two ways: a descriptive approach and a predictive method. The descriptive analysis examines the responses of faculty and students to the first six questions in the survey that were identical for both groups. The predictive assessment explores

using discriminant analysis to investigate if the responses to questions one thru six could be used

to indicate whether the respondent belongs to the faculty or student group. Can the responses be

used to predict group membership?

Table 4 provides the descriptive analysis for the first six questions for both faculty and

student groups. Many of the questions in this table indicate differences in attitude and practice

between faculty and students. Whether this divergence is sufficiently large enough to distinguish

between faculty and student group membership based on the responses is explored in the next

section that explores the use of discriminant analysis.

Table 4

Descriptive Analysis of Faculty Versus Student Responses

Question	Faculty Responses	Student Responses
1 (Encrypt Flash)	No one encrypts their drive	9.5% encrypt their drive
2 (Use Security)	No one uses security	12.6% uses security
3 (Importance of Security)	(See Table 3 for details) 30% felt that security was extremely important	100% felt that security was extremely important
4 (Use Passwords)	15.3% use passwords	11.1% use passwords
5 (Backup)	92.3% backup their flash drives	100% backup their flash drives
6 (Attitude Toward Security)	76.9% believe that security is important for their flash drives	66.6.6% believe that security is important for their flash drives

The purpose of discriminant analysis is to predict group membership based on predictor

variables (Tabachnick & Fidell, 2007). Discriminant analysis is used to predict membership in

naturally occurring groups rather than groups developed through random assignment. The use of

faculty and student groups fit this requirement. In addition, the fact that these groups differ in

size is not an issue for the analysis (Tabachnick & Fidell). In this analysis one classification

function was used that included responses to questions one thru six on both the faculty and

student questionnaires. The *PASW 19* statistical package used to conduct the analysis. Table 5

reveals that the Wilks's Lambda was statistically significant and the function was useful in

discriminating the faculty and student group membership. Table 6 identifies the standardized

discriminant function coefficients. These standardized discriminant function coefficients are

equivalent to the standardized betas in regression analysis (Field, 2009); the higher the value of

these coefficients the greater is their contribution to separating the groups. In this case question

three contributed the most in separating the two groups. Finally the overall success in using the

function to distinguish between faculty and student responses can be seen in Table 7. This table

shows that the six questions were successfully in 94.7% of the cases. The cases that were

classified incorrectly can be found in Appendix E shows how every instance from the 63 student

and 13 faculty replies to the questions were classified.

Table 5

Wilks's Lambda

Test of Function(s)	Wilks's Lambda	Chi-square	df	Sig.
1	.416	62.306	6	.000

Table 6

Standardized Canonical Discriminant Function

	Function 1
UseFlash	-.167
UseSecurity	-.237
ImportanceOfSecurity	.950
Passwords	.332
Backup	-.187
AttitudeTowardSecurity	.227

Table 7

Classification Results [a]

			Predicted Group Membership		
		FacultyStudent	0	1	Total
Original	Count	0	63	0	63
		1	4	9	13
	%	0	100.0	.0	100.0
		1	30.8	69.2	100.0

[a] 94.7% of original grouped cases correctly classified.

Conclusion

This paper identified portable data storage devices as potential security victims to a variety of security issues. One solution advocated in the report is the TrueCrypt software that encrypts and hides data. The TrueCrypt software is particularly useful for safeguarding data on USB flash drives that are easily compromised. Whether or not individuals opt to secure their portable data devices is a function of their attitudes toward security and their subsequent behaviors. Data was collected from undergraduate students in business classes to examine some

of the attitudes and practices toward USB Flash Drives and from university faculty. It was found that there were strong associations between the use of flash drives, security, and the use of passwords for the student data. All student participants indicated that security for a flash drive was extremely important; all respondents also indicated that using security with a flash drive was important. Faculty did not use encryption or passwords to secure their flash drives. All students backup their drives but only 92.3% of faculty performs this action. It was also found that the first six questions in the questionnaire were useful for predicting group membership. The faculty received a presentation on how to encrypt their USB drives using TrueCrypt. Faculty in this sample unanimously agreed that they would be likely to encrypt their drives after having been shown TrueCrypt. The six questions used in both surveys appear to successfully discriminate between faculty and student responses. One suggestion for further research to administer the six questions to another group such as technology professionals to determine if these questions would useful in distinguishing among faculty, student, and professional groups.

References

EzineArticles.com. (2010). Usb drives a security threat? Retrieved November 1, 2010, from http://ezinearticles.com/?USB-Flash-Drives-a-Security-Threat?&id=1375217

Field, A. (2009). *Discovering statistics using spss* (3rd ed.). Thousand Oaks, CA: SAGE Publications.

GFI Software. (2010). The threats posed by portable storage devices. Retrieved November 1, 2010, from http://www.gfi.com/whitepapers/threat-posed-by-portable-storage-devices.pdf

Howell, D. C. (2002). *Statistical methods for psychology* (5th ed.). Pacific Grove, CA: Duxbury/Thomson Learning.

IronKey, Inc. (2007). Benefits of secure usb flash drives. Retrieved November 1, 2010, from http://www.discovery.net.au/content/downld/IronKey%20Whitepaper%20-%20Benefits%20of%20Secure%20USB%20Flash%20Drives.pdf

McAfee Labs. (2010). Mcafee threats report: First quarter 2010. Retrieved November 1, 2010, from http://www.mcafee.com/us/local_content/reports/2010q1_threats_report.pdf

PCTechGuide. (2009). Definition of usb flash drive. Retrieved November 1, 2010, from http://www.pctechguide.com/glossary/WordFind.php?wordInput=USB+Flash+Drive&input=Look+it+up%21&searchType=MatchWord

Simon, S. (2005). What is a phi coefficient? Retrieved November 2, 2010, from http://www.childrens-mercy.org/stats/definitions/phi.htm

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston: Pearson/Allyn & Bacon.

TrueCrypt Foundation. (2010a). Introduction. Retrieved October 1, 2010, from http://www.truecrypt.org/docs/

TrueCrypt Foundation. (2010b). Truecrypt. Retrieved October 1, 2010, from

http://www.truecrypt.org/

EzineArticles.com. (2010). Usb drives a security threat? Retrieved November 1, 2010, from

    http://ezinearticles.com/?USB-Flash-Drives-a-Security-Threat?&id=1375217

GFI Software. (2010). The threats posed by portable storage devices. Retrieved November 1,

    2010, from http://www.gfi.com/whitepapers/threat-posed-by-portable-storage-devices.pdf

Howell, D. C. (2002). *Statistical methods for psychology* (5th ed.). Pacific Grove, CA:

    Duxbury/Thomson Learning.

IronKey, Inc. (2007). Benefits of secure usb flash drives. Retrieved November 1, 2010, from

    http://www.discovery.net.au/content/downld/IronKey%20Whitepaper%20-

    %20Benefits%20of%20Secure%20USB%20Flash%20Drives.pdf

McAfee Labs. (2010). Mcafee threats report: First quarter 2010. Retrieved November 1, 2010,

    from http://www.mcafee.com/us/local_content/reports/2010q1_threats_report.pdf

PCTechGuide. (2009). Definition of usb flash drive. Retrieved November 1, 2010, from

    http://www.pctechguide.com/glossary/WordFind.php?wordInput=USB+Flash+Drive&in

    put=Look+it+up%21&searchType=MatchWord

Simon, S. (2005). What is a phi coefficient? Retrieved November 2, 2010, from

    http://www.childrens-mercy.org/stats/definitions/phi.htm

TrueCrypt Foundation. (2010a). Introduction. Retrieved October 1, 2010, from

    http://www.truecrypt.org/docs/

TrueCrypt Foundation. (2010b). Truecrypt. Retrieved October 1, 2010, from

    http://www.truecrypt.org/

Appendix A

Survey About Encryption and Portable Data Storage

1) Do you encrypt your USB flash drive?

Yes   No

2) Do you use any type of security for your USB flash drive?

Yes   No

3) How important do you think security is for a flash drive?

1   2   3   4   5

(1-extremely important, 2-somewhat important, 3-neutral, 4-not very important, 5-not important at all)

4) Do you use passwords to protect your USB flash drive?

Yes   No

5) Do you backup your work?

Yes   No

6) Do you think it is important to use security when using a USB flash drive?

Yes   No

7) What year in school are you?

Freshman
Sophmore
Junior
Senior

APPENDIX B

Student Survey Results

Q1	Q2	Q3	Q4	Q5	Q6	Q7
0	0	1	0	1	1	3
0	0	1	0	1	1	4
0	0	1	0	1	1	4
0	0	1	0	1	1	2
0	0	1	0	1	0	4
0	0	1	0	1	0	4
0	0	1	0	1	1	4
0	0	1	0	1	1	3
0	0	1	0	1	1	4
1	0	1	0	1	1	3
0	0	1	0	1	1	4
0	1	1	1	1	1	3
0	0	1	0	1	1	3
0	0	1	0	1	0	2
0	0	1	0	1	0	4
0	0	1	0	1	0	3
0	0	1	0	1	0	4
0	0	1	0	1	0	3
0	0	1	0	1	1	1
0	0	1	0	1	1	4
0	0	1	0	1	0	4
0	0	1	0	1	1	4
1	1	1	1	1	1	2
1	1	1	1	1	1	2
1	1	1	0	1	1	1
0	0	1	0	1	0	2
0	0	1	1	1	1	4
0	0	1	0	1	1	3
0	0	1	0	1	1	3
0	0	1	0	1	1	3
0	0	1	0	1	1	3
0	0	1	0	1	0	2
1	1	1	1	1	1	2
0	0	1	0	1	1	2
0	0	1	0	1	1	3
0	0	1	0	1	1	3
0	0	1	0	1	1	3
0	0	1	0	1	1	3
0	0	1	0	1	1	3
0	0	1	0	1	1	3

APPENDIX B (Continued)

Q1	Q2	Q3	Q4	Q5	Q6	Q7
0	0	1	0	1	0	3
0	0	1	0	1	1	3
0	0	1	0	1	1	3
0	0	1	0	1	0	3
0	0	1	0	1	1	4
0	1	1	0	1	1	3
0	0	1	0	1	1	4
0	0	1	0	1	0	3
0	0	1	1	1	0	4
0	1	1	0	1	0	3
0	0	1	0	1	1	3
0	0	1	0	1	0	4
0	0	1	0	1	0	4
0	0	1	0	1	1	2
0	0	1	0	1	1	4
0	0	1	0	1	1	3
1	1	1	1	1	1	4
0	0	1	0	1	0	3
0	0	1	0	1	1	3
0	0	1	0	1	1	3
0	0	1	0	1	0	3
0	0	1	0	1	0	3
0	0	1	0	1	0	3

APPENDIX C

Faculty Survey About Encryption and Portable Data Storage

1) Do you encrypt your USB flash drive?

Yes   No

2) Do you use any type of security for your USB flash drive?

Yes   No

3) How important do you think security is for a flash drive?

1   2   3   4   5

(1-extremely important, 2-somewhat important, 3-neutral, 4-not very important, 5-not important at all)

4) Do you use passwords to protect your USB flash drive?

Yes   No

5) Do you backup your work?

Yes   No

6) Do you think it is important to use security when using a USB flash drive?

Yes   No

7) What school are you in?

Arts & Science   Business   Health Professions   Library

8) Do you use a USB flash drive while teaching in the classroom or other presentations?

Yes   No

9) Do you use a USB flash drive for transporting data?

Yes   No

10) After seeing how to use the encryption software, are you more likely to use it to secure your USB flash drive?

Yes   No

Appendix D

Faculty Survey Results

Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
0	0	1	0	1	1	1	1	1	1
0	0	1	0	1	1	3	0	1	1
0	0	1	0	1	1	2	1	1	1
0	0	4	0	1	1	3	1	1	1
0	0	5	0	1	1	3	1	1	1
0	0	3	1	1	0	4	1	1	1
0	0	2	0	1	1	1	1	1	1
0	0	4	0	1	1	2	0	1	1
0	0	3	0	0	0	3	1	1	1
0	0	2	0	1	1	3	1	1	1
0	0	1	1	1	1	3	1	1	1
0	0	3	0	1	0	2	1	1	1
0	0	3	0	1	1	3	1	1	1

Appendix E

Classification Statistics

Casewise Statistics (0= student and 1= faculty)

Case	Actual Group	Predicted Group
1	0	0
2	0	0
3	0	0
4	0	0
5	0	0
6	0	0
7	0	0
8	0	0
9	0	0
10	0	0
11	0	0
12	0	0
13	0	0
14	0	0
15	0	0
16	0	0
17	0	0
18	0	0
19	0	0
20	0	0
21	0	0
22	0	0
23	0	0
24	0	0
25	0	0
26	0	0
27	0	0
28	0	0
29	0	0
30	0	0
31	0	0
32	0	0
33	0	0
34	0	0
35	0	0
36	0	0
37	0	0

Appendix E (Continued)

Case	Actual Group	Predicted Group
38	0	0
39	0	0
40	0	0
41	0	0
42	0	0
43	0	0
44	0	0
45	0	0
46	0	0
47	0	0
48	0	0
49	0	0
50	0	0
51	0	0
52	0	0
53	0	0
54	0	0
55	0	0
56	0	0
57	0	0
58	0	0
59	0	0
60	0	0
61	0	0
62	0	0
63	0	0
64	1	0**
65	1	0**
66	1	0**
67	1	1
68	1	1
69	1	1
70	1	1
71	1	1
72	1	1
73	1	1
74	1	0**
75	1	1
76	1	1

** Misclassified case

Examination of the State of Digital Forensics Education Programs in the United States

Doug White, John McQuilkin
Gabelli School of Business
Roger Williams University

ABSTRACT

This paper reports on data collection efforts to document the status of digital forensics [DF] programs currently in the United States. As Digital Forensics continues to grow as a discipline, more programs will develop at both the undergraduate and graduate level. This paper represents an initial attempt to begin documentation of those programs, which degrees they offer, where they are housed administratively, and what they are called. This will begin the process of developing a published resource for reference to these programs.

A DEA BASED MULTIDIMENSIONAL FRAMEWORK FOR ANALYZING COMMERCIAL BANKS

Rashmi Malhotra, St. Joseph's University, (610) 660-3497, rmalhotr@sju.edu
Raymond Poteau, Philadelphia University, (215) 951-2811, PoteauR@philau.edu

ABSTRACT

This study develops a multidimensional framework using data envelopment analysis as a benchmarking tool to assess the performance of the commercial banks in India. Using data envelopment analysis approach, this study compares the relative performance of 35 banks against one another with eight variables as the benchmark parameters. This study finds that most of the banks are consistently performing well over a period from 2005 to 2009. The study also shows the areas in which inefficient banks are lagging behind and how they can improve their performance to bring them at par with the efficient commercial banks.

INTRODUCTION

Data Envelopment Analysis (DEA) is a very popular technique that aids managers in making decisions in a multidimensional framework. Although, there are many methods and techniques to design decision processes that rank-order alternatives, but DEA also illustrates how an alternative can be improved. Thus DEA is an excellent tool for practitioners who can use this mathematical modeling technique to benchmark different decision-making units. This study illustrates the use of Data Envelopment Analysis to study the efficiency of commercial banks in India. For the past three years, the developed world has been experiencing many financial problems such as subprime lending, toxic assets, stricken banks, and over leverage. Usually the global financial turbulence hits the emerging markets harder than the developed market counterparts. However, the recent financial crisis has seen a reversal of roles between the developed markets and emerging markets. The emerging markets rebounded much more quickly, while the United States of America, the world's biggest economy, along with other developed markets continue to struggle. Therefore, evaluating the performance of the commercial banks in an emerging market like India is an important issue that can offer further insight into the progress made by these nations using Data Envelopment Analysis.

Data Envelopment Analysis is a technique that assesses the productive efficiency of homogenous operating units such as schools, hospitals, banks or utility companies. DEA approach is a powerful technique for performance measurement, because of its objectivity, ability to handle multiple inputs and multiple outputs that can be measured in different units. Also, in contrast to non-parametric techniques such as Stochastic Frontier, DEA approach does not require specification of any functional relationship between inputs and outputs or a priori specification of weights of inputs and outputs. DEA provides gross efficiency scores based on the effect of controllable and uncontrollable factors.

We examine the performance of Indian commercial banks during the period 2005 to 2009. This time period covers before crisis and during crisis time. In particular, the behavior of profitability, cost of intermediation, efficiency, soundness of the banking system, and industry concentration are examined in this paper.

The rest of the paper is organized along the following lines. Section II provides a review of literature that relates to the data envelopment analysis approach. Section III discusses the data envelopment analysis model. Section IV discusses the data and methodology being used in this paper and section V provides an empirical analysis of our results. The paper provides a summary and conclusion of our results in section VI.

LITERATURE REVIEW

Several studies have explored the use of data envelopment analysis models to evaluate the relative efficiency of banks. Kumar and Gulati (2010) evaluate the efficiency, effectiveness, and performance of Indian public sector banks using DEA models. They report that there is not much relationship between efficiency and effectiveness in Indian public sector banks, while performance and efficiency are positively related. Ray and Das (2010) find considerable variation in average levels of profit efficiency across various ownership categories of Indian banks. In general, state owned banks are found to be more efficient than their private counter parts. Staub, Souza, and Tabak (2010) investigate cost, technical and allocative efficiencies for Brazilian banks in the recent period (2000-2007. State-owned banks are significantly more cost efficient than foreign, private domestic and private with foreign participation. Tsolas (2010) provides a framework for evaluating the overall performance of bank branches in terms of profitability efficiency and effectiveness. The study highlights the importance of encouraging increased profitability and efficiency throughout the branch network of the commercial banks. Zhao, Casu, and Ferrari (2008) examine the impact of regulatory reform on the performance of Indian commercial *banks*. Foreign *banks* appear to have acted as technological innovators when competition increased, which added to the competitive pressure in the banking market.

Shanmugam and Das (2004) measure technical efficiency of Indian banks in four different ownership structures during the reform period of 1992-1999. They find that the efficiencies of raising interest margin are time invariant, while the efficiencies of raising other outputs -non-interest income, investments and credits - are time varying. Das and Ghosh (2006) report that during the period 1992-2002, medium-sized public sector banks performed reasonably well and are more likely to operate at higher levels of technical efficiency. They report a close relationship between efficiency and soundness as determined by bank's capital adequacy ratio. Their study also shows technically more efficient banks have, on an average, less non-performing loans. Mostafa (2007) uses DEA to evaluate the relative efficiency of top 50 Gulf Cooperation Council (GCC) banks. The study reports significant room for improvement for the top 50 GCC banks.

Rezvanian, Rao, and Mehdian (2008) used a nonparametric frontier approach to examine the effects of the ownership on the efficiency, efficiency change, technological progress and productivity growth of the Indian banking industry over the period 1998 to 2003. They reported that foreign banks were significantly more efficient when compared to other banks, i.e. the Indian privately-owned and publicly-owned-banks. The findings also provided evidence to indicate that a large number of banks operated below their optimal scale. Kao and Liu (2004) compute efficiency scores based on the data contained in the financial statements of Taiwanese banks. They use this data to make advanced predictions of the performances of 24 commercial banks in Taiwan. Pille and Paradi (2002) analyze the financial performance of Ontario credit unions. They develop models to detect weaknesses in Credit Unions in Ontario, Canada. Halkos and Salamouris (2004) explore the efficiency of Greek banks with the use of a

number of suggested financial efficiency ratios for the time period 1997–1999. They show that DEA can be used as either an alternative or complement to ratio analysis for the evaluation of an organization's performance. The study finds that the higher the size of total assets the higher the efficiency. Neal (2004) investigates X-efficiency and productivity change in Australian banking between 1995 and 1999 using DEA and Malmquist productivity indexes. It differs from earlier studies by examining efficiency by bank type and finds that regional banks are less efficient than other bank types. The study concludes that diseconomies of scale set in very early, and hence are not a sufficient basis on which to allow mergers between large banks to proceed. Paradi and Schaffnit (2004) evaluate the performance of the commercial branches of a large Canadian bank using DEA. Chen et al. (2005) study the efficiency and productivity growth of commercial banks in Taiwan before and after financial holding corporations' establishment. They employ a DEA approach to generate efficiency indices as well as Malmquist productivity growth indices for each bank. Howland and Rowse (2006) assess the efficiency of branches of a major Canadian bank by benchmarking them against the DEA model of US bank branch efficiency. Sufian (2007) uses DEA approach to evaluate trends in the efficiency of the Singapore banking sector. The paper uses DEA approach to distinguish between technical, pure technical and scale efficiencies. Sanjeev (2007) evaluates the efficiency of the public sector banks operating in India for a period of five years (1997–2001) using DEA. The study also investigates if there is any relationship between the efficiency and size of the banks. The results of the study suggest that no conclusive relationship can be established between the efficiency and size

MODEL

The Data Envelopment Analysis Model1
Data Envelopment Analysis (DEA) (Charnes et al., 1978) model uses linear programming to measure the comparative performance of different organizational units. Further, this generalized optimization technique measures the relative performance of different decision-making entities that have multiple objectives (outputs) and multiple inputs structure. In the DEA terminology, entities/organization units under study are called Decision-Making Units (DMUs). In our study, the DMUs are the seven retailers under analysis. DEA measures the efficiency with which a DMU uses the resources available (inputs) to generate a given set of outputs. The DEA methodology assesses the performance of the DMU using the concept of efficiency or productivity, defined as a ratio of total outputs to total inputs. Further, the DEA model estimates relative efficiency, which is with reference to the best performing DMU or DMUs (in case multiple DMUs are most efficient). The DEA allocates an efficiency score of unity or 100 percent to the most efficient unit. The low-performing DMUs' efficiency can vary between 0 and 100 percent in comparison to the best performance.

DATA AND METHODOLOGY

The data for this study was obtained from CNBC's moneycontrol.com website. The sample consists of 20 state owned banks and 15 private banks. Data covers the fiscal year ending March 31st 2005 to March 31st 2009. We used eight financial variables to evaluate the efficiency of the banks in pre- and post-economic crisis period ranging from 2005 to 2009. Given the significance of banking in economic growth, banks are considered private companies with a public purpose. They seek to create value for all

---

[1] The main sources of the DEA Model description are Ramanathan (2003) and Zhu (2003).

the stakeholders and maximize shareholder wealth subject to the various risks and within the constraint of market competition, social constraints, and the legal/regulatory framework. The private nature of banks requires them to make profits to be viable and the public nature of banks emphasizes safety and soundness of the bank's operations. Profitability is important for the survival of a bank, but safety and security is also critical for the survival of the financial system. Every bank makes trade-offs between the profitability level it is striving to achieve and the risks it is willing to take. When evaluating a bank, an analyst should consider both its profitability and its financial condition. Taken alone, short-term profit trends can be misleading. For example, if a bank achieves loan growth by engaging in excessively risky lending, it may be vulnerable to developments that would hurt its earnings or even threaten its survival over time. Therefore, in order to evaluate banks, we consider the following four broad sets of ratios that capture the private-public nature of banking: Profitability Ratios, Cost of Intermediation, Management Efficiency Ratios, and Safety Ratios. The profitability ratios are used to measure how well a bank is performing in terms of profit. The profitability ratios can also be defined as the financial measurement that evaluates the capacity of a business to produce yield against the expenses and costs of business over a particular time period. Profitability of a bank is assessed through:

- Return on Assets: Return on assets is a comprehensive measure of bank profitability and is computed by dividing bank's net income by its total assets.
- Interest income relative to total funds: Interest income relative to total funds also known as yield on earning assets is calculated by dividing interest income on earning assets by the value of these assets during the same period. For the Indian commercial banks, we use interest income relative to total funds as a measure of the yield on earning assets.
- Intermediation costs reflect the cost of transforming deposits from one set of customers into loans for another set of customers. Cost of intermediation can also be used as an indicator of competitiveness in banking industry and can be measured in several different ways:
- Interest Spread: The difference between interest earned and interest expended is one measure of the cost of intermediation. Higher difference will point to a lower cost of intermediation and lower difference will point to a higher cost of intermediation.
- Interest Expense to Interest Earned Ratio: Interest expense represents the cost of obtaining deposits and interest earned the revenue on the loans made. Therefore, this ratio is another way of measuring the cost of intermediation.

Management efficiency ratios for banks mainly focus on the costs other than interest. We use the following two ratios to assess management efficiency of Indian banks.
- Asset Utilization Ratio: The efficiency with which a bank utilizes its assets.
- Efficiency Ratio: This ratio is based on noninterest expenses divided by operating revenue. Noninterest expenses represent all expenses incurred in operations, including such items as personnel and occupancy costs (salaries, technology, building, supplies, and administrative expenses). Operating revenue includes net interest income (interest revenue less interest expense) plus fees income. This ratio is a measure of a bank's productivity based on costs required to generate each dollar of revenue. This reflects the costs involved in maintaining branches and servicing retail accounts. A lower efficiency ratio translates into lower operational costs, which means greater operational efficiency.

A safe and sound banking system is critical for a healthy financial market. Evaluation of a bank must consider the risks that the bank is taking in order to remain profitable. We use two ratios to evaluate riskiness of banks:
- Debt leverage: Banks are usually highly leveraged organizations. Banks borrow money so that

they can expand their capacity to earn more money by either expanding their facilities or by making additional loans for which they do not have sufficient deposits. For the commercial banks in India, we use total debt divided by owners' funds as a measure of debt leverage.

- Capital Adequacy Ratio: Bank capital plays a very important role in the safety and soundness of individual banks and the banking system. Capital adequacy norms ensure that capital should be adequate to absorb unexpected losses or risks involved. If there is higher risk, then it would be necessary to provide back up with capital. Capital Adequacy measures the strength of the bank.

DATA ENVELOPMENT MODEL

Besides the mathematical and computational requirements of the DEA model, there are many other factors that affect the specifications of the DEA model. These factors relate to the choice of the DMUs for a given DEA application, selection of inputs and outputs, choice of DMUs for a given DEA application, selection of inputs and outputs, choice of a particular DEA model (e.g. CRS, VRS, etc.) for a given application, and choice of an appropriate sensitivity analysis procedure (Ramanathan, 2003). Due to DEA's non parametric nature, there is no clear specification search strategy. However, the results of the analysis depend on the inputs/outputs included in the DEA model. There are two main factors that influence the selection of DMUs – homogeneity and the number of DMUs. To successfully apply the DEA methodology, we should consider homogenous units that perform similar tasks, and accomplish similar objectives. In our study, the banks are homogenous as they compete with each other in the same market. Furthermore, the number of DMUs is also an important consideration. The number of DMUs should be reasonable so as to capture high performance units, and sharply identify the relation between inputs and outputs. There are some simple rules of thumb that guide the selection of inputs and outputs, and the number of participating DMUs[2].

To study the performance of these banks, we consider eight factors: return on assets, interest income relative to total funds, interest spread, Interest Expense to Interest Earned Ratio, Asset Utilization Ratio, efficiency ratio, debt leverage, and capital adequacy. Out of these eight factors, we specified efficiency ratio, interest expense to interest earned ratio, and loan to total funds ratio as input, because if a bank has lower efficiency ratio and low debt, it is an indicator of superior performance. All other factors will be considered as output factors as a higher value of these variables improves the efficiency or performance of the bank. Finally, the choice of the DEA model is also an important consideration. We should select the appropriate DEA model with options such as input maximizing or output minimizing, multiplier or

---

[2] The following are the guidelines for DMU model selection:
  a. The number of DMUs is expected to be larger than the product of number of inputs and outputs (Darrat et. Al., 2002; Avkiran, 2001) to discriminate effectively between efficient and inefficient DMUs. The sample size should be at least 2 or 3 times larger than the sum of the number of inputs and outputs (Ramanathan, 2003).
  b. The criteria for selction of inputs and outputs are also quite subjective. A DEA study should start with an exhaustive, mutual list of inputs and outputs that are considered relevant for the study. Screening inputs and outputs can be quite quantitative (e.g. statistical) or qualitative that are simply judgmental, use expert advice, or use methods such as analytical hierarchy process (Saaty, 1980). Typically inputs are the resources utilized by the DMUs or condition affecting the performance of DMUs. On the other hand, outputs are the benefits generated as a result of the operation of the DMUs, and records higher performance in terms of efficiency. Typically, we should restrict the total number of inputs and outputs to a reasonable level. As the number of inputs and outputs to a reasonable level. As the number of inputs and outputs increases, more number of DMUs get an efficiency rate of 1, as they become too specialized to be evaluated with respect to other units (Ramanathan, 2003).

envelopment, and constant or variable returns to scale. DEA applications that involve inflexible inputs or not fully under control inputs should use output-based formulations. On the contrary, an application with outputs that are an outcome of managerial goals, input-based DEA formulations are more appropriate. In addition, for an application that emphasizes inputs and outputs, we should use multiplier version. Similarly, for an application that considers relations among DMUs, envelopment models are more suitable. Furthermore, the characteristics of the application dictate the use of constant or variable returns to scale. If the performance of DMUs depends heavily on the scale of operation, constant returns to scale (CRS) is more applicable, otherwise variable returns to scale is a more appropriate assumption.

In our study, the relationship among these banks is an important consideration as they are all commercial banks. Therefore, we select the envelopment models for our analysis. In addition, inflation is not a very flexible input that cannot be immediately controlled. Therefore, output-based formulation is recommended for our study. Furthermore, the performance of these banks does not depend on the scale of operations, thus variable returns to scale is safe assumption. Also, the structure of the DEA model (in envelopment form) uses an equation and separate calculation for every input and output. Therefore, all the input and output variables can be used simultaneously and measured in their own units.

EMPIRICAL ANALYSIS

Each of the banks is a homogenous unit, and we can apply the DEA methodology to asses a comparative performance of these banks. The study evaluates the progress of the commercial banks by tracking the gains (or losses) made by each of the 35 banks. Using the DEA methodology, we can calculate an efficiency score for the 35 banks on a scale of 1 to 100. We analyzed and computed the efficiency of the banks for the period 2005-2009. Table 2illustrates the efficiency scores and the rankings of the 35 banks from the year 2005-2009. As illustrated in table 2, in the year 2005, 21 banks are 100% efficient as compared to all other banks. The other banks in the order of increasing efficiency in the range of 80%-89% efficiency are: ING Vysya Bank, Development Credit Banks, Dhanalakshmi Bank, Bank of India, and UCO Bank. Dena Bank, Syndicate Bank, Canara Bank , Union Bank of India, Bank of Baroda, State Bank of India, Axis bank, Vijaya Bank, and Indian Bank are in the range of 90%- 99% efficiency. In the year 2006, 20 banks were 100% efficient and on the efficiency frontier. The other banks in the order of increasing efficiency in the range of 80%-89% efficiency are: ING Vysya Bank, UCO Bank, Syndicate Bank, and Dena Bank. Bank of India, Union Bank of India, State Bank of India, Canara Bank, Vijaya Bank, Central Bank of India, Bank of Maharashtra, Punjab National Bank, Jammu and Kashmir Bank, Corporation Bank, and Oriental Bank of are in the range of 90%- 99% efficiency. In 2007, 9 banks are 100% efficient, and the remaining 27 banks are inefficient in the range of 74% to 96%. In the year 2008, 12 banks are 100% efficient, and the remaining 24 banks were inefficient within the range of 85%-98%. In the year 2009, 11 banks are efficient, and 25 banks are inefficient with efficiency levels in the range of 88%-99%. Figure 1 illustrates the efficiency factor on the radial axis for the year 2009. The pareto-efficient banks form the efficiency frontier, and the inefficient banks are below the efficiency frontier.

We present the score in percentage value varying between 0% and 100%. We find that efficiency of Indian Bank, Kotak Mahindra Bank, Yes Bank, Bank of Rajasthan, Federal Bank, HDFC Bank, Jammu and Kashmir Bank, Development Credit Bank, Karnatka Bank, ICICI Bank, and Karur Vysya Bank is 100%. On the other hand, UCO Bank, IDBI Bank, State Bank of Mysore, Oriental Bank of Commerce, Vijaya Bank, Central Bank of India, Canara bank, Andhra Bank, ING Vysya Bank, Allahabad Bank, Syndicate Bank, Bank of Baroda, Dhanalakshmi bank, Bank of Maharashtra, Indian Overseas Bank,

Dena Bank, Union Bank of India, Indusland Bank, Punjab National Bank, Axis Bank, Bank of India, and City Union Bank rank from 12 to 35 in the order of decreasing efficiency of 99%, to 88%. This means that the observed levels of current return on assets, interest income relative to total funds, interest spread, Asset Utilization Ratio, and capital adequacy of Allahabad Bank are .91 times the maximum output level that Allahabad Bank can secure with its return on assets, interest income relative to total funds, interest spread, Asset Utilization Ratio, and capital adequacy . The same rationale applies to all the other inefficient banks. Table 2 illustrates the efficiency scores and the corresponding ranking of the thirty five banks in the year 2009. Figure 2 illustrates the trend in the graphical form for all banks from the year 2005 to 2009. As UCO Bank, IDBI Bank, State Bank of Mysore, Oriental Bank of Commerce, Vijaya Bank, Central Bank of India, Canara bank, Andhra Bank, ING Vysya Bank, Allahabad Bank, Syndicate Bank, Bank of Baroda, Dhanalakshmi bank, Bank of Maharashtra, Indian Overseas Bank, Dena Bank, Union Bank of India, Indusland Bank, Punjab National Bank, Axis Bank, Bank of India, and City Union Bank are inefficient for the year 2009; the next step is to identify the efficient peer group or banks whose operating practices can serve as a benchmark to improve the performance of these banks. Table 3 illustrates the peer group for the inefficient banks.

As shown in the table, Bank of Rajasthan, ICICI Bank, and Kotak Mahindra serve as peer for Allahabad Bank. In addition, Allahabad Bank is more comparable to Kotak Mahindra (weight 37%) and less comparable to its more distant peer Bank of Rajasthan (32%), and far distant peer ICICI Bank (31%). Thus, Allahabad Bank should scale up its efficiency ratio, interest expense to interest earned ratio, and loan to total funds ratio to make them comparable with Kotak Mahindra. Similarly, Andhra Bank has ICICI Bank (53%) as the closest peer that it should emulate and Kotak Mahindra (43%) as the distant peer bank and Bank of Rajasthan (4%) that can also be investigated. Bank of Baroda has Kotak Mahindra (55%) as its immediate peer, and ICICI Bank (44%) as its next distant peer, and Bank of Baroda (1%) as its far distant peer. Similarly, Bank of India has Indian Bank, Federal Bank, ICICI Bank, and Karur Vysya as its peers. Similarly, we can apply the same rationale to Canara Bank, Central Bank of India, Corporation Bank, Dena Bank, IDBI Bank, Indian Overseas Bank, Oriental Bank of Commerce, Punjab National Bank, State Bank of India, State Bank of Mysore, Syndicate Bank, UCO Bank, Union Bank of India, Vijaya Bank, Axis Bank, Bank of Maharashtra, City Union Bank, Dhanalakshmi Bank, Indusland Bank, and ING Vysya Bank. ICICI Bank and Kotak Mahindra Bank are the most immediate or immediate peer for most of the inefficient banks. On the other hand, Bank of Rajasthan is the distant or far distant peer for twenty inefficient countries. Similarly, Jammu and Kashmir Bank is the far distant peer for four of the inefficient banks and an immediate peer for Indian Overseas Bank. HDFC Bank is the far distant peer for three of the inefficient banks and immediate peer for Axis Bank. Therefore, ICICI Bank and Kotak Mahindra Bank are the most efficient bank among all the banks as not only are 100% efficient, but also serves as the role model for all (except three) banks.

Similarly, Bank of Rajasthan is the next most efficient bank among the group of given banks. Jammu and Kashmir Bank serves as the next immediate peer bank for Indian Overseas Bank as the characteristics of Indian Overseas resemble Jammu and Kashmir Bank. Similarly, HDFC Bank is the immediate peer for Axis Bank as the two banks share similar characteristics. Thus, Jammu and Kashmir Bank and Axis Bank are the next most efficient banks among the group of banks under consideration. Karur Vysya Bank is the immediate peer for City Union Bank and far distant peer for Bank of India. This is quite expected as the characteristics of inefficient banks match the peer banks. The efficient peer banks have a similar mix of input-output levels to that of the corresponding inefficient bank, but at more absolute levels. The efficient banks generally have higher output levels relative to the bank in question. The features of efficient peer banks make them very useful as role models that inefficient

banks can emulate to improve their performance. Furthermore, ICICI Bank and Kotak Mahindra Bank are used as an efficient peer to almost all Pareto-inefficient banks, so their frequency of use as an efficient-peer, expressed as a percentage of the number of pareto-inefficient banks, is 88%. Bank of Rajasthan is an efficient peer to seventeen banks with a frequency rate of 71%. Indian Bank is an efficient peer to six banks with net percentage of 25%. In addition, Development Credit Bank and Karur Vysya Bank have the peer efficiency frequencies of 8% each. Thus, we have enhanced confidence that Development Credit Bank and Karur Vysya Bank are truly well performing banks as they outperform all the other banks. Similarly, Karnatka Bank, Federal Bank, and yes Bank serve as a peer to one of the inefficient banks. Furthermore, these banks are more likely to be a better role model for less efficient banks to emulate because their operating practices and environment match more closely those of the bulk of banks. Table 4 displays the benchmarking factor and the hit percentage of efficient bank.

After calculating the efficiency of a bank using DEA, and identifying the efficient peers, the next step in DEA analysis is feasible expansion of the output or contraction of the input levels of the bank within the possible set of input-output levels. The DEA efficiency measure tells us whether or next bank can improve its performance relative to the set of banks to which it is being compared. Therefore, after maximizing the output efficiency, the next stage involves calculating the optimal set of slack values with assurance that output efficiency will not increase at the expense of slack values of the input and output factors. Once efficiency has been maximized, the model does seek the maximum sum of the input and output slacks. If any of these values is positive at the optimal solution to the DEA model that implies that the corresponding output of the bank (DMU) can improve further after its output levels have been raised by the efficiency factor, without the need for additional input. If the efficiency is 100% and the slack variables are zero, then the output levels of a bank cannot be expanded jointly or individually without raising its input level. Further, its input level cannot be lowered given its output levels. Thus, the banks are pareto-efficient with technical output efficiency of 1. If the bank is 100% efficient but one slack value is positive at the optimal solution then the DEA model has identified a point on the efficiency frontier that offers the same level on one of the outputs as bank A in question, but it offers in excess of the bank A on the output corresponding to the positive slack. Thus, bank A is not Pareto-efficient, but with radial efficiency of 1 as its output cannot be expanded jointly. Finally, if the bank A is not efficient (<100%) or the efficiency factor is greater than 1, then the bank in question is not Pareto-efficient and efficiency factor is the maximum factor by which both its observed output levels can be expanded without the need to raise its output. If at the optimal solution, we have not only output efficiency > 1, but also some positive slack, then the output of bank A corresponding to the positive slack can be raised by more than the factor output efficiency, without the need for additional input. The potential additional output at bank A is not reflected in its efficiency measure because the additional output does not apply across all output dimensions.

Table 5 illustrates the slack values identified in the next stage of the DEA analysis. The slack variables for 100% efficient banks as well as less than 100% efficient banks are not zero. Therefore, all the banks used in this study are not Pareto-efficient as the DEA model has been able to identify some feasible production point which can improve on some other input or output level. For example, for Allahabad Bank, besides increasing the output level of interest spread by 5.55 units, there is further scope for increasing interest to total funds by 9.26 (units), asset utilization by .01 (units), capital adequacy by 2.94 (units), and return on assets by 0.09 (units). Allahabad Bank can follow Kotak Mahindra, Rajasthan Bank, and ICICI Bank as its role models and emulate their policies. Similarly, IDBI Bank can reduce its interest expense to interest earned by 4.96 units and increase interest spread by 3.21 units, interest to total funds by 8.84 units, asset utilization by 0.02 units, capital adequacy by 3.51 units, and return on

assets by 0.45 (units) while maintaining efficient levels equivalent to that of its peers—ICICI Bank and Bank of Rajasthan. On the same lines, we can find slack values for Indian Bank. Although Indian Bank is 100% efficient, there is still scope for improvement as Indian Bank can improve its output productivity without additional inputs. Therefore, Indian Bank can increase its output factors, interest spread by 3.47 units and interest to total funds by 8.37 units. Similarly, we can find the slack factors for all the other banks. Table 5 illustrates the slack values of the relevant factors for inefficient banks.

SUMMARY AND CONCLUSIONS

Using data envelopment analysis approach, this study compares the relative performance of thirty five Indian commercial banks against one another with eight performance variables as the benchmark parameters from 2005 to 2009. By studying the time period from 2005 to 2009, we also look at the variations in the performance of these banks over a period of time to assess their progress This study finds that ICICI Bank, Kotak Mahindra Bank, Bank of Rajasthan, HDFC Bank, and Jammu and Kashmir Bank consistently outperform all the other banks with 100% relative efficiency. Indian Bank, Bank of Rajasthan, Development Credit Bank, Federal Bank, HDFC Bank, ICICI Bank, Karur Vysya Bank, and Kotak Mahindra Bank show consistent improvement in industry performance. The study also shows the areas in which inefficient member banks are lagging behind and how they can improve their performance to bring them at par with other participating banks.

The data envelopment analysis is a powerful technique for performance measurement. The major strength of DEA is its objectivity. DEA identifies efficiency ratings based on numeric data as opposed to subjective human judgment and opinion. In addition, DEA can handle multiple input and outputs measured in different units. Also, unlike statistical methods of performance analysis, DEA is non-parametric, and does not assume a functional form relating inputs and outputs. This study finds that there is lack of convergence in the performance of 35 banks and some banks have performed more efficiently in contrast to other banks.

However, as with any other study, this study using DEA has certain limitations (Ramanathan, 2003). The application of DEA involves solving a separate linear program for each DMU. Thus, the use of DEA can be computationally intensive. In addition, as DMU is an extreme point technique, errors in measurement can cause significant problems. DEA efficiencies are very sensitive to even small errors, thus making sensitivity analysis an important component of post-DEA procedure. Also, as DEA is a non-parametric technique, statistical hypothesis tests are difficult to apply. Therefore, further extension of this study would be to perform principal component analysis of the all the DEA model combinations. Furthermore, we can also use logistic regression to test the validity of the results.

TABLES, FIGURES, & REFERENCES

Tables, figures, references, and full paper available upon request from the authors.

AN ONTOLOGICAL APPROACH TO DATA MINING FOR EMERGENCY MEDICINE

Atif Khan, John A. Doucette, Changjiu Jin, Lijie Fu, and Robin Cohen

David R. Cheriton School of Computer Science, University of Waterloo
200 University Avenue West Waterloo, ON, Canada N2L 3G1
`{a78khan,j3doucet,c4jin,l6fu,rcohen}@uwaterloo.ca`

Abstract. We propose a system for assisting emergency personnel in making evidence-based medical decisions in scenarios with constraints on time and knowledge. Our approach is based on the artificial intelligence technique of semantic data mining, focusing on the use of ontology-based knowledge representation to provide the precise definition of entities and their relationships. This is particularly useful in medical applications due to the heterogeneity of medical data and the diversity of health care environments. Leveraging the SNOMED-CT medical terminology encoding system, we are able to extract information by applying context-aware semantic reasoning. Our high level system architecture includes four primary layers: an aggregation of data repositories, an ontology, a reasoning engine, and the application interface. The backbone of the system consists of knowledge representation and a reasoning engine connected to an end user system via a natural language processing mechanism. We also compare our approach with other methodologies for medical decision making (under similar constraints), emphasizing the particular value offered by our ontological approach for these time critical environments.

Keywords: Semantic Knowledge Processing, Information Retrieval, Knowledge Integration, Evidence Based Medacine, Clinical Decision Making

1 INTRODUCTION

With the advance of Information Communication Technologies (ICTs), more medical organizations are utilizing electronic systems to capture, manage and use patient related information. The scope of this information varies from organization wide Electronic Medical Records (EMRs) to Electronic Health Records (EHRs) shared between different organizations.

These systems utilize various formats for persisting and exchanging medical data, some of which are incompatible with one another. Consequently, it can be quite difficult to process large amounts of patient data from various sources and produce relevant information. For time critical scenarios, such as in an emergency room setting, these medical information systems do not provide the desired functionality since an intelligent (human) agent is required to (manually) search through the patient data. Furthermore in other scenarios where the information might be readily available but

access to the information might be restricted by the available communication bandwidth, it might be difficult to bring all pertinent information to medical personnel.

As a motivational example, consider the following scenario. A patient is found unconscious on the sidewalk. The 911 call gets the emergency response team out to bring the patient in. The emergency team's first goal is to stabilize the patient and then to transport him to the nearest hospital. The stabilization protocol might require administration of a drug or use of a procedure that is generally acceptable but might be harmful for this particular patient. We assume that the patient's medical records contain evidence against administration of such a drug or procedure. However, even if the emergency crew could see all patient information, the nature of the situation, the ability of crew members, and the time factor might not allow them to reach the best solution for the patient. Consequently, the patient may suffer adverse reactions which could have been prevented.

Our proposed approach draws from the field of artificial intelligence, utilizing an ontology-based knowledge representation and a semantic reasoner to infer answers to various queries. More specifically, all patient medical data, drug information and medical procedures are represented using well established ontologies (such as SNOMED-CT [16]), while inference can be conducted using established software [29]. For the scenario stated above, our system would create a triple-store (i.e. an ontological structured database) based on (i) data found in patient medical records, (ii) data describing drug interactions, and (iii) data describing various medical procedures.

Once the triple-store has been established, we can apply rules to the knowledge-base to infer new facts. For example, we might apply rules to discover that a diabetic patient who has pancreatic cancer should not be administered insulin to stabilize blood glucose levels. Or that a certain drug should not be given to a pregnant patient. Given these rules and the knowledge-base, the emergency team members can query the system asking questions and receiving answers based specifically on the patient data from the knowledge-base. A semantic reasoner can provide a proof supporting its diagnosis, allowing easy validation by human doctors or other medical personnel.

Our proposed solution is targeted at health care providers with diverse backgrounds and technical abilities, ranging from nurses and emergency response team members to general practitioners and specialists. The user defines the context/scope of use by providing the relevant contextual information. The knowledge-base can then be tailored and refined based on the specified context. The user queries the knowledge-base in an interactive fashion by posing various questions in natural language. The user query is then translated into a triple statement using the terminology from the various ontologies. A referencing engine (Euler[29]) is then invoked to mine the answer from the knowledge-base. The answers can be directly mined or can be inferred from the existing data in the knowledge-base.

The use of an ontology provides two main benefits. First, the ontology provides semantic interoperability so that information (mined from various systems) can be represented for processing and knowledge mining. Second, an ontological model allows us to discover relationships that might not be obvious just from the raw data.

In contrast with other non-semantic systems (with similar objectives), the main advantage of our solution is the ability to respond to ad-hoc queries over dissimilar and distributed data-sets, providing useful information to make an evidence based decision.

2 BACKGROUND

In this section we will briefly review the fundamental primitives and concepts required for structured knowledge representation, communication and processing. We will use these primitives and concepts to lay down the foundation for knowledge-based solutions utilizing ontological representations.

"Knowledge management is concerned with the representation, organization, acquisition, creation, usage, and evolution of knowledge in its many forms. To build effective technologies for knowledge management, we need to further our understanding of how individuals, groups and organizations use knowledge" [25].

Medical information systems utilize ontology based knowledge representations for defining, describing and presenting information. Each specialized medical field may utilize the corresponding specialized ontology best suited for the field. For example, the Systematized Nomenclature of Pathology (SNOP) was originally conceived to specify morphology codes [33] for pathology.

Given the depth and the diversity of the stored information, *"an assortment of techniques for representing and managing codified knowledge has emerged from a number of areas in computer science, notably* **artificial intelligence***, databases, software engineering, and information systems"* [25]. Development of techniques for knowledge representation (in forms that can be exploited by intelligent machine based agents) is the main contribution of the artificial intelligence area and the focus of our particular design.

We begin by reviewing the basic primitives as follows:

2.1 Ontologies in Knowledge-based Systems

Within artificial intelligence, each knowledge base uses three basic primitives: (i) individuals (ii) relations between individuals and (iii) features/attributes of individuals, which represents the current "world" within the confines of a current interpretation. However, in order for multiple individuals to operate and work with this interpretation, they all need to agree on how to represent the various concepts. This is to ensure that knowledge can be:

- shared between various people (semantic interoperability)
- aggregated from various sources
- preserved over time maintaining its original meaning

An ontology is defined as the specification of the meaning of all the symbols in a knowledge-based information system [26]. There are various ways to formalize and represent an ontological model. The most flexible representation of an ontology is the *triple* representation [26]. A triple can be thought of as a simple three-word sentence to represent knowledge. The first word is a **subject** representing an individual. The second word, a **verb**, usually defines an attribute of the individual. The third word is an **object** and represents a value of the attribute. Let us consider a simple example to illustrate this. First we define a prop relationship to represent a property, so that saying prop(X,Y,Z) is the same as saying that X has property Y, with value Z. We can then express knowledge/information in the following example:

- prop (dell_1557, owned_by, bob).
- prop (dell_1557, manufacturer, dell).
- prop (dell_1557, colour, red).

A triple-store is used to capture and store all information in the conceptualized world. Its format is similar to a relational database; however a relational database requires all relationships to be explicitly identifiable. A triple-store has no such restriction. A triple-store can be queried for information in a manner similar to a relational database.

2.2 RDF/Notation 3 Representation

Resource Description Framework (RDF) is a standard model for data interchange on the World Wide Web (WWW). It has features that facilitate data merging even if the underlying schemas differ, and it specifically supports the evolution of schemas over time without requiring all the data consumers to be changed[27]. These properties of RDF make it an ideal candidate for management and exchange of medical information where the data-sets (i) have a particularly large set of consumers (ii) enjoy a long life span and (iii) are constantly evolving.

Notation 3 (N3) is based on RDF standards and represents an analogous syntax to RDF/XML to represent data. Furthermore, N3 adds extra features (like rules and formulas) in order to process data and make inferences from facts in the data. N3 is (i) designed for a human-readability and (ii) comparatively easy to work with, as it is less verbose than RDF/XML[30]. N3 follows the triple representation of information. An N3 triple-store is a collection of these statements where each statement is terminated by a period '.' symbol. Consider the following information store:

N3 Statement (triple)	Translation
:thermo :temp :high.	thermostat temperature is high
:heating :power 0.	heating power is zero

N3 also allows us to represent rules that can govern the relationships between individuals based on their corresponding attributes. For example by using the information above, we can create the following rule in N3 notation which states that if the thermostat temperature is high, then the heating power is zero.

```
:thermo :temp :high ⇒ :heating :power 0.
```

This capability adds considerable deduction and inference capabilities to a simple data-set where the relationships between the individuals might not be explicitly known (in contrast to a relational database information store).

2.3 Querying for Information

Once the knowledge-base has been transformed into a triple-store and the inference rules have been established, we can query for new information by asking questions. Generally a semantic reasoner is required to consume the triple-store and establish the various obvious relationships. The reasoner is also capable of knowledge inference to discover latent relationships between individuals. For example if the knowledge-base defines A equals to B and B equals to C, then the semantic reasoner is capable of making the inference that A equals C.

It is important to realize that the semantic reasoner makes a decision based only on the knowledge-base available to it at the time of the user query. Therefore when the reasoner fails to establish a relationship, it could be that there exists no such relationship (in the world) or that the knowledge-base is incomplete. Let us consider a simple example to further illustrate this. Our triple-store (knowledge-base) is defined as follows.

```
:bob :member :agfa.
:agfa :w3cmember :w3c.
:agfa :subscribed :mailing_list_w3c.
```

We can define a simple access rule as follows to state that any person who is a member of an institution that belongs to w3c and subscribes to the w3c mailing list can have access to the w3c mailing list.

```
{?PERSON :member ?INSTITUTION.
?INSTITUTION :w3cmember :w3c.
?INSTITUTION :subscribed ?MAILINGLIST}
⇒ {?PERSON :authenticated ?MAILINGLIST}.
```

Given the knowledge-base and the access rule, we can now ask the question "Who is authenticated to the w3c mailing list?" as follows:

```
_:WHO :authenticated :mailing_list_w3c.
```

The semantic reasoner will derive all relevant relationships using logical inference, and will produce a list of all individuals satisfying the authentication rule condition.

3   EXISTING DATING MINING TECHNIQUES

In recent years data mining applications in medicine have gained traction, since many organizations and research groups have realized that these techniques are useful for extracting knowledge

from large databases. In the data mining area, some traditional methods are discussed and classified in [4][1][2]. Also, several criteria are proposed to classify these methods, including the kind of patterns discovered (predictive or informative); the representation language (symbolic or sub-symbolic); and the data mining goal (classification or regression).

Methods from various subfields have been applied in several real and artificial domains [1], and their performance has been compared in several studies [2]. The main conclusion is that no method is better than the others in all domains [3], because the performance of a method is related to the application domain. This was subsequently shown to hold across *all possible* data mining techniques by Wolpert and Macready's "No Free Lunch" Theorem [12], which shows that, in the absence of task specific information, the expected performance of all optimization techniques is identical.

We begin by introducing five example methods representing a broad cross section of data mining approaches: (1) Statistics and Pattern Recognition; (2) Machine Learning, (3) Neural Networks, (4) Support Vector Machines, (5) K-Nearest neighbors algorithms.

Statistics and Pattern Recognition: Generally, an explicit underlying probability model is needed for the statistical methods, which provides an estimate of the membership of an example to a class. Before the application of these methods, users must define formal models and hypotheses. Classical methods are based on Fisher's work on linear discrimination [22], which is a method used to find a linear combination of features which characterize or separate two or more classes of objects or events. Three methods are commonly applied, the Linear, Quadratic, and Logistic discriminants.

Machine Learning: Machine Learning is an artificial intelligence research area that studies computational methods for improving performance by mechanizing the acquisition of knowledge from experience [4]. Machine Learning algorithms enable the induction of a symbolic model, decision tree or set of rules, of preferably low complexity, but high transparency and accuracy [5]. Some classic techniques that are still widely used in Machine Learning are discussed in depth below.

Case-Based Reasoning: When a new case is presented to this method, Case-Based Reasoning looks for a previous case (or cases) held in memory which have characteristics similar to this one. These similar previous cases (for which the final outcomes are known) are used to infer the status of the new case. Case-Based Reasoning's performance is very sensitive to the choice of indexing scheme and the similarity metric used.

Decision Trees: Decision trees are used to predict the value of a target variable based on several input variables, by building a tree with nodes that branch based on the values of various data attributes. Decision trees can be generated using algorithms such as Quinlan's ID3 and C4.5 [23][2], or a recursive partitioning method, which is a statistical method for mining large data-sets in order to uncover hidden patterns and find the correct classification (such as in CART [11]). ID3 and its successors are very popular methods and the precursor to implementations with advanced features, like example selection for the training sets; pruning methods; good generalization performance; and automatic conversion to classification rules.

Genetic Algorithms: The origin of Genetic Algorithms is the study of cellular automata [10]. They simulate the natural selection mechanism by starting with a population of candidates, also called organisms, and growing successive generations through applying crossover and mutations to the organisms. When evaluation functions use "elitism" (i.e. allow only the best candidates to survive), proofs of convergence to the optimal solution exist [32]. Genetic Algorithms are most

often used to extract complex concepts from small databases. This is because the technique makes several scans over a training database to predict accurate rules, making it more I/O intensive and inappropriate for mining extremely large data-sets. However, more recent research has provided practical solutions to this problem[34].

Neural Networks: Neural Networks are based on models of the human brain. They represent knowledge as a network of units, or neurons, distributed in one or several layers that transmit activation values from the input nodes to the output nodes. Neural Networks are used for classification, clustering, prediction and regression. Some configuration rules are defined for the weighting of the transmission, the activation of the nodes and the connection pattern of the network. Different settings of these configuration rules produce different types of Neural Networks.

Support Vector Machines: A Support Vector Machine (SVM) is a discrimination technique, which consists in separating several sets of points by a hyperplane [24]. It is based on the use of kernels, which allow an optimal separation of the points of the plane in various categories. SVMs are based on the structural risk minimization principle, while traditional neural networks utilize the empirical risk minimization principle[35]. SVMs are used for classification [8] and regression [9].

K-Nearest Neighbors Algorithms: The K-Nearest Neighbors (KNN) algorithm is intended for classification tasks, but can be extended to estimation tasks as well. KNN is a case-based reasoning algorithm, based on the idea of making decisions using similar previously solved cases in memory. CBR also matches new cases to find similar past cases, however CBR does not use Euclidean distance measures. Generally, the functioning of a KNN algorithm is carried out in two steps: (1) the distance between a new sample and each training sample is computed using Euclidean distance measures. (2) the K samples which are the closest to the new sample are identified and the new sample is labeled with the most frequent result among them. KNN algorithms are used for classification [6][7].

3.1 Semantic Approaches

As discussed above, the "No Free Lunch" (NFL) theorem for search explicitly proves the expected equivalence of all conventional data mining techniques when used in the absence of task specific information [12]. More specifically, NFL shows that a particular data mining technique can only perform better than average by $X\%$ on a particular task if it also performs exactly $X\%$ worse in total on some set of other problems. Naturally, this motivates incorporation of domain specific knowledge into data mining algorithms and the medical domain, with its substantial diversity of data mining tasks, has proven to be a fertile starting point for many researchers. This is likely due to the widespread availability of computerized domain knowledge, like the SNOMED-CT and Unified Medical Language System (UMLS) ontologies [16,17], which index an enormous number of terms and the relationships between them.

One area which has received particular attention is the improvement of query results for doctors seeking advice in the practice of evidence based medicine. In this domain, doctors attempt to obtain the most relevant and current information on the appropriate treatment for patients. Studies of doctors' search queries indicate that common questions regard the ideal dose of a specific medication or the causes of a specific symptom [14,13], suggesting the introduction of domain specific knowledge by way of question interpretation. Mendonca et al. present a system for automatically

formulating appropriate search queries from typical physician question formats [15]. Additionally, the system is capable of automatically adjusting queries based on relevant patient information, so that queries about a patient with "congestive heart disease" can be automatically structured in terms of that specific condition, even if formulated in terms of a more general condition (e.g. "heart disease"). This is accomplished using the subsumption relationships in UMLS.

Similarly, Fushman and Lin develop, implement, and benchmark a system allowing physicians to augment their search of the PubMed database [18] using ontological relationships between document meta-data concepts [19]. The system requires use of a specialized query structure which necessitates converting the natural language query into a meta-data like format. For example, the authors show how the query "In children with an acute febrile illness, what is the efficacy of single-medication therapy with acetaminophen or ibuprofen in reducing fever?" is required to be formatted as:

- Search Task: therapy selection
- Problem/Population: acute febrile illness/in children
- Intervention: acetaminophen
- Comparison: ibuprofen
- Outcome: reducing fever

The similarity of meta-data in the search results returned by the natural language search terms can then be compared with the formatted query using conventional classification techniques to improve relevancy. The use of explicit query formatting addresses previous work which showed that a major contributor to doctor's dissatisfaction with search results is poorly formatted or ambiguous queries [20]. Fushman and Lin's system is empirically validated, and shown to outperform the NIH's own query system [19].

Jacquemart and Zweigenbaum also propose the use of ontologies in a medical question answering system. The UMLS ontology [17] is used convert medical students' questions into a list of keywords for a conventional search engine [21].

Although all the approaches above utilize ontologies for data mining, they are primarily concerned with first-order relationships between terms. Indeed, the question posed in our motivating example ("Can I administer insulin?") is strikingly similar to one used by Jacquemart and Zweigenbaum as an example of a challenging problem for their system, "My patient is diabetic; can I use vasoconstrictors?" [21]. This query is challenging for their system because it involves higher order relationships between terms. There is unlikely to be an explicit relationship in a medical ontology indicating that mixing vasoconstrictors and diabetes is problematic, but such a relationship may still be present implicitly, requiring a reasoning engine to discover it.

After presenting our proposed solution in the section that follows, we return to it contrast with the various competing methods outlined in this section.

4   PROPOSED SOLUTION

4.1   System Architecture

Kararia and Juric suggested an ontological layering model to share e-Health information across heterogeneous data sources [31]. We adopt a modularized architecture loosely based upon this

Examination of the State of Digital Forensics Education Programs in the United States

Doug White, John McQuilkin
Gabelli School of Business
Roger Williams University

ABSTRACT

This paper reports on data collection efforts to document the status of digital forensics [DF] programs currently in the United States. As Digital Forensics continues to grow as a discipline, more programs will develop at both the undergraduate and graduate level. This paper represents an initial attempt to begin documentation of those programs, which degrees they offer, where they are housed administratively, and what they are called. This will begin the process of developing a published resource for reference to these programs.

Fig. 1. System Architecture.

system to provide a foundation capable of incorporating a board array of medical data sources into our model's reasoning framework (see Figure 1). The architecture is composed of five modules. The *data repository module* contains medical information stored in a variety of heterogeneous data stores (ranging from relational database systems to unstructured documents (e.g. doctors' notes). In addition, they could reside in the WWW or local hospital information systems.

The *ontological module* is responsible for providing semantics to the diverse medical information from the data repository module. It includes the ontologies used to define and classify different types of health care data. For the purpose of illustration, we assume that all medical data can be represented by an ontology such as SNOMED-CT. In order to facilitate the aggregated semantic representation, we utilize constraints provided by the *requirement module*.

The *reasoning module* is where the semantic reasoning take place. It takes the output a natural language processing (NLP) agent and utilizes semantic reasoners to fulfill the user query. The inner workings of this module are discussed in detail in sections 4.2 and 4.3.

The *application module* provides an abstraction of the user interface for our system. We do not enforce any particular requirements on an actual implementation. Consequently our system is capable of supporting multiple user-interfaces corresponding to various user characteristics and medical devices. For example, a user interface used by a physician could allow for more precise and diverse terminology than the interface used by an emergency response team member.

The *requirement module* cooperates with the reasoning layer. It decides which data sources and ontologies are needed in order to retrieve the correct information for this query. This is accomplished by extracting the keywords from the query, and generating constraints on the data sources and ontologies to narrow the searching range.

4.2   Details of Natural Language Processing Module

Our proposed solution takes advantage of the comparatively easy natural language processing mechanisms provided by use of ontologies in question answering systems. Like earlier authors [21,15], we propose to incorporate matching of user queries to commonly used formats, to facilitate automatic conversion to logically structured queries. We limit ourselves to three possible formats in this work, while noting the possibility of extensions:

1. "Is there any reason not to administer treatment Y to patient X?"
2. "What is an alternative treatment to Y for patient X, given condition Z?"
3. "What is the nature of the adverse reaction if patient X is given treatment Y?"

Obviously there exists a nearly infinite number of variants on these basic forms, but previous research [21,15] suggests that patten matching is quite viable if our system limits itself to answering a set of question types. For example, the question "Can I give John Doe insulin?" has two nouns, "John Doe" and "insulin". Our ontology will readily confirm that "John Doe" is a patient, and that "insulin" is a treatment. Utilizing synonyms in conjunction with other ontological resources, such a query can be found to be most similar to question 1 above.

 The use of a finite set of canonical question forms also provides us with the ability to use "canned" queries formats, so that, continuing with our example above, the given natural language is converted into a a logical expression whose form is equivalent to the natural language in question 1 above. For example: $\exists I, s.t. I = <JohnDoe, negative\_interaction, Insulin>?$.

4.3   Motivational Scenario

In this section we consider a simple scenario to highlight the various components of our proposed solution and their interactions with each other. We use a simplistic model of the world as depicted in Figure 2. The main entities (players) defined in our model are patients, health care providers, drugs, diseases and various medical conditions. There are various relationships defined between these entities, the main relationship being the IS_A relationship (sometimes called "subsumption"). For example a doctor is a health care provider which is a person. Similarly Insulin is an allopathic drug which is a drug. Besides the IS_A relationship, we can find various other relationships between the entities defined. For example the disease ulcer has a condition called bleeding, the drug Viagra has contra indication to the drug nitroglycerin. Please refer to Figure 2 for details.

 Using the triple notation (discussed earlier) we can capture the semantic model in a triple-store (see Appendix). Furthermore, we can add additional information based on observing actual objects from our conceptualized world (these objects have been omitted from Figure 2 to reduce the visual complexity). The following N3 statements represent some of the entities defined along with their observed attributes and properties.

```
:John a :Patient;
  :name "John";
  :hasDisease :Hypoglycemia,:PancreaticCancer.
:James a :Patient;
  :name "James";
  :hasDisease :Ulcer.
:Mary a :Person;
  :name "Mary";
  :condition :Pregnancy.
:Dr.Smith a :Doctor.
:ERT.Jane a :EmergencyResponseTeam.
```

We will pose a series of questions to the semantic reasoner. Some of the answers can be derived directly from the information stored in the knowledge-base, while others will require making inferences based on the available knowledge and the data-mining rules (see Appendix).
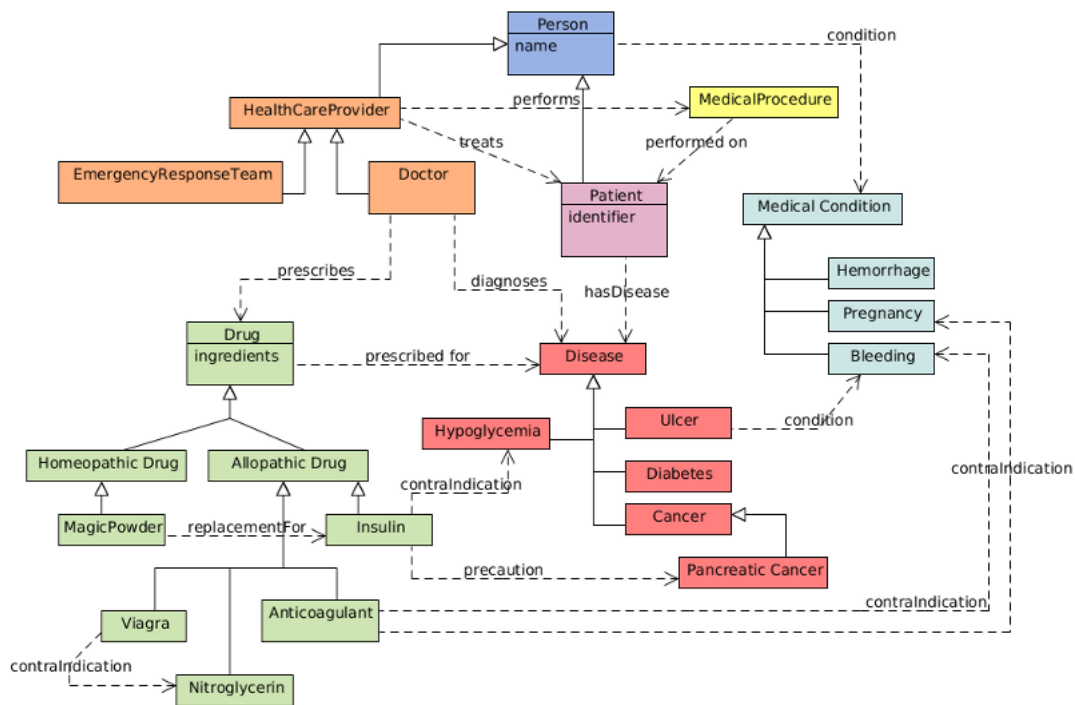
Fig. 2. Simple semantic model representing main entities & their relationships.

Consider the following scenarios:

Scenario1 : Assuming that an emergency response team member is interested in administering an insulin injection to patient John. The patient's medical record indicates that John has hypoglycemia. Furthermore from the knowledge-base we know that insulin is a contra indicator to hypoglycemia, and therefore should not be administered tox patients with hypoglycemia. In order to make the best possible decision for the patient, the emergency response team member poses the following queries to our system.

Can John be administered insulin? The system replies back with evidence that John can not be given insulin, since he is suffering from Hypoglycemia.

```
:John :hasDisease :Hypoglycemia e:evidence file:knowledge-base#68>
:Insulin :contraIndication :Hypoglycemia e:evidence <file:knowledge-base#33> implies
:John :canNotBeGiven :Insulin e:evidence <file:rules#15>
```

The reasoning engine found the evidence that John has disease hypoglycemia and insulin is a contra indicator for hypoglycemia from the knowledge-base. It then used the following rule from the inference rule-set to decide that John should not be given insulin.

```
{ ?PATIENT :hasDisease? DISEASE.
?DRUG :contraIndication? DISEASE.
} ⇒ {?PATIENT :canNotBeGiven ?DRUG}.
```

What can be used as a replacement of insulin for John? From Figure 2 we can see that a homeopathic drug "Magic Powder" can be used as a replacement for insulin.

Furthermore there exists no contra indication relationship between hypoglycemia and the drug magic powder. A query to the system finds no evidence as to why John should not be administered the magic powder.

Scenario2 : We can also pose more generic queries to the system, such as asking for a list of everyone who an anticoagulant drug should not be administered to.

```
_:WHO :canNotBeGiven :Anticoagulant.
```

The reasoner searches the whole knowledge-base and finds that an anticoagulant drug should not be administered to the following people: Mary, since she has a condition pregnancy; James, since he has an ulcer; anyone with an ulcer. The relevant part of the knowledge base is displayed below.

```
:Mary :condition :Pregnancy} e:evidence <file:knowledge-base#76>.
:Anticoagulant :contraIndication :Pregnancy} e:evidence <file:knowledge-base#28>} implies
:Mary :canNotBeGiven :Anticoagulant} e:evidence <file:rules#20>}.

:James :hasDisease :Ulcer} e:evidence <file:knowledge-base#72>.
:Ulcer :condition :Bleeding} e:evidence <file:knowledge-base#51>.
:Anticoagulant :contraIndication :Bleeding} e:evidence <file:knowledge-base#28>} implies
:James :canNotBeGiven :Anticoagulant} e:evidence <file:rules#26>}.

:Ulcer :condition :Bleeding e:evidence <file:knowledge-base#51>.
:Anticoagulant :contraIndication :Bleeding} e:evidence<file:knowledge-base#28> implies
:Ulcer :canNotBeGiven :Anticoagulant} e:evidence <file:rules#20>}.
```

The following rules are utilized in the above inference.

```
{?ANY :condition ?CONDITION.
?DRUG :contraIndication ?CONDITION.
} ⇒ {?ANY :canNotBeGiven ?DRUG}.

{?PATIENT :hasDisease ?DISEASE.
?DISEASE :condition ?CONDITION.
?DRUG :contraIndication ?CONDITION.
} ⇒ {?PATIENT :canNotBeGiven ?DRUG}.
```

5   COMPARING WITH EXISTING NON-SEMANTIC STRATEGIES

5.1   Intrinsic Disadvantages of Non-Semantic Data Mining Strategies in Medicine

There are numerous drawbacks intrinsic in avoiding the use of semantics in data mining, both in Medicine and in other fields. Our approach attempts to remedy many of these shortcomings, and to allow application of data mining techniques in areas which were previously infeasible.

No Free Lunch: As discussed briefly in the existing work section above, the most readily apparent drawback of non-semantic approaches is the "No Free Lunch" (NFL) theorem for optimization [12]. The theorem is so named because of the way in which it is often described: in terms of restaurant menus. Given an arbitrary set of menu options, M, and a set of prices P, such that $|P| = |M|$, we can imagine a set of restaurants R, such that every possible combination of menu items and prices is represented as some member of R. That is, each row of *PXM* is a member of R. Then, a consumer who does not know which restaurant they will be dining in has the same expected price for their meal (namely, the average of the elements in P), *independent* of their menu item preferences. Thus a vegetarian (who will be more likely to pick certain dishes than others) will have exactly the same expected price for their meal as a person who is equally likely to pick any menu item.

This analogy may seem fairly far removed from the problem of optimization, but the essential similarity lies in the fact that in optimization, in the absence of a priori knowledge about the properties of the function to be optimized, all possible strategies for optimization must perform precisely the same on average. The menu items are replaced by possible "next steps" in optimization. The prices are replaced by the improvement that will result from making a particular next step, and the dining preferences are replaced by strategies for picking the next step in optimization.

Since answering ad-hoc queries with conventional data mining techniques necessitates the construction of a new regression or classification model, and that construction requires optimization of some criteria (e.g. minimizing empirical error), all commonly used semantic-free approaches are subject to NFL. However, NFL applies only in the absence of domain specific information (if you know which restaurant you're going to, you can make a better estimate about how much you will pay). The addition of domain specific knowledge is readily accomplished using ontologies, as we have shown here, but can also be accomplished using non-semantic approaches, in principle.

For example, instead of using ontological processes to find relevant information, we could construct a new support vector machine to answer each patient query (i.e. by formulating the query as a classification problem, training the SVM using a relevant subset of a patient database, and using the resulting model to make predictions about the patient of interest). For example, we could phrase the query "Can I give Joe insulin?" in terms of building a classification model which predicts whether a patient should be given insulin, and train such a model using data from people who were given insulin in the past. The use of domain specific knowledge here would be equivalent to selecting only those attributes relevant to answering the question of interest, a process often called feature selection. In the absence of semantic knowledge, feature selection techniques are essentially limited to information-based methods, whose efficacy will depend on the quality of the data, and which may uncover relationships which are incidental (e.g. diabetes $\rightarrow$ heart disease), rather than causal (diabetes $\leftarrow$ overweight $\rightarrow$ heart disease). Although the relationships may exist in practice, the lack of a causal basis for the resulting model could be highly problematic when supporting clinical reasoning.

Opacity: One of the principal concerns with using certain modern data mining techniques is the opacity of the resulting models. While techniques like support vector machines have the ability to discover patterns involving non-linear interactions between dozens or hundreds of features, clinicians may be hesitant to trust such patterns, especially when there is no guarantee of causal relationships. While these effects can be mitigated by choice of model and judicious pruning of rule sets, the inherent opacity of these models precludes their use by computational laypersons in emergency situations, where the results of an ad-hoc query cannot necessarily scrutinized in detail before application.

Speed: Utilizing a conventional model for time critical ad-hoc queries may be infeasible, since obtaining high quality results typically requires an abundance of data, and a substantial number of features may be required in the absence of domain specific knowledge. Training time for conventional data mining techniques typically scales no better than linearly in the number of data points and the number of features used.

5.2   The Advantages of Our Solution

Optimization Independent: Our system relies on a static (or nearly static) ontology which is assumed to define the sum total of empirically validated human medical knowledge. As such, answering an ad-hoc query with our system involves reasoning about the given relationships and facts (of the semantic entities), rather than the construction of a new data model. While conventional approaches must automatically construct a new model from the data in response to each query, our model is static and so does not do so. This removes the need to engage in optimization when answering a new query, and so is not subject to the No Free Lunch theorem. This is an important point, since it ensures that it is possible for our system to have consistently reliable performance.

Transparent: Compared with the opacity of the conventional data mining techniques, our semantic approach is much more transparent. The user can define the semantic rules clearly, in advance of the data mining process, according to prior knowledge and experience. Consequently, the results can be better controlled and are more readily accepted by the user. In addition, our approach can provide its chain of reasoning directly to the end user for verification, allowing the user to review the thought process behind an unexpected result.

Fast Response: Again, the fact that our system relies on a static model provides a solution to a problem with conventional methods. While the requisite theorem proving may still take some time, most queries can be expected to terminate quickly. Also, because our system uses a static implementation, it may be possible to produce highly efficient system-specific optimizations more readily than with conventional approaches.

6   FUTURE WORK

Future work on our proposed system would focus on implementation and extensions. It appears that implementation of most system components could be accomplished using a suite of open source software (The Euler Reasoning Engine, SNOMED-CT Ontology), and natural language processing techniques outlined in publicly available literature (e.g [21]). Hardware and user interface could be deployed using existing infrastructure via smart phones with a proprietary application. Arguably, the most difficult component to implement would be to integrate patient records into the data records portion of the system, a process which relies on the prior presence SNOMED-CT tags. Fortunately, as there are other good reasons to want such tags in patient records (e.g. for billing purposes), eventual widespread adoption of the technology seems plausible.

In terms of extensions, we intend to prioritize the addition of one additional question into the system: "What is the approximate risk of administering treatment Y to patient X?". This question might be important for emergency responders if a patient is identified as being unsuitable for treatment with available resources due to the possibility of adverse effects. If no other treatment is available, and the patient's condition is sufficiently urgent (i.e. life threatening), then the relative risk of treating vs. not treating the patient becomes pertinent. Thus, a patient who will certainly die before reaching hospital, but only has a 5% chance of adverse reaction to the proposed treatment, might be treated in spite of the possibility of adverse reaction. In contrast, a patient whose faces an 80% chance of a serious adverse reaction might not be treated if there is any chance of reaching the hospital alive. We propose the attachment of supporting evidence to relationships in the SNOMED-CT ontology as a possible strategy for answering this query. This would consist of locating a document from the medical literature which describes the relationship in question, possibly including a non-technical document summary or an extracted scalar valued for the strength of the relationship. For example, we might attach a study showing the relationship between diabetes and obesity to the obesity →diabetes link, along with a simple summary (e.g. "Obesity increases the likelihood of diabetes by a factor of 5"), and a scalar value (e.g. "5"). Although SNOMED-CT contains an enormous number of inter-concept relationships (approximately 1.5 million), evidence supporting each link can, in principle, be gathered automatically using specialized information retrieval techniques [19]. Concise summaries of the resulting documents could be provided as a response to queries about explicit relationships in the ontology.

Implicit relationships require a somewhat more complex approach. While in principle it might be possible to extract a scalar value for the "strength" of each link from the corresponding supporting evidence, it is not clear that doing so makes sense in all cases. Even if a scalar corresponding to the magnitude of the effect implied by a particular relationship could be identified for all possible interpretations present in SNOMED-CT, it is not clear how to interpret the series of such scalars that would be returned by a query about an implicit relationship in the ontology. For example,

suppose we are told that we cannot administer a clotting agent based on the following chain of reasoning about patient John Doe (note: the given numbers do not correspond to reality):

1. John Doe is overweight, but has not been diagnosed with hypertension (100% chance)
2. Overweight people often have undiagnosed hypertension (25% chance)
3. People with hypertension sometimes react badly when given procoagulants (12% chance)

There are several possible ways we could interpret these results, but all of them are problematic. First, we could assume that being overweight does *not* increase the risk of complications when administering procoagulants to a patient with hypertension (i.e. items 2 and 3 are *uncorrelated*). In this case, we anticipate a 3% chance of adverse reaction if we administer the drug ($0.12 * 0.25 = 0.03$). However, we could just as easily assume that the two events are correlated, in which case the risk of complications can range from 25% if the events are fully positively correlated (i.e. if $2 \rightarrow 3$), to 0% if they are fully negatively correlated (i.e. if $2 \nrightarrow 3$). This issue arises because SNOMED-CT (or for that matter, medical science as a whole!) cannot possibly quantify the nth-order relationships between all possible combinations of terms. As a stop gap measure, our system could report the range of possible values, or perhaps just the maximum value. Continuing with the above example, we could return {0,0.25} as the range of values, though this is not necessarily very helpful. In the long term, we could consider incorporation of Bayesian models for patient data in to the system to provide better estimates. Under this system, a large database of patient records could be used to estimate the strength of the correlations between various relationships based exclusively on previous patient data, though this introduces other problems.

7   CONCLUSIONS

In this paper, we propose a system for semantic data mining to aid in medical decision making. Semantic data mining is enabled by ontology-based knowledge representation, which provides the precise definition of entities and their relationships. A high level system architecture is proposed, including data repository, ontology, reasoning, and application/interface layers. We also compare our approach with the other methodologies, in particular statistical methods such as decision tree, support vector machine, and so on. In contrast with other non-semantic systems, our solution responds to ad-hoc queries over dissimilar and distributed data-sets, providing useful information to an agent to make an evidence based decision.

References

[1] V. Maojo, and J. Sanandr's. *A Survey of Data Mining Techniques*, ISMDA 2000.

[2] Weiss, S and Indurkhya. *Predictive Data Mining. A Practical Guide.*, Morgan Kaufmann.San Frncisco, CA. 1998.

[3] Carla E. Brodley. *Recursive automatic bias selection for classifier construction.*, Machine Learning, 20:63-94, 1995.

[4] Pat Langley and H.A. Simon. *Applications of machine learning and rule induction.*, Communications of the ACM, 38(11):55-64, Nov 1995.

[5] Nada Lavrac, Elpida Keravnou, and Blaz Zupan. *Intelligent Data Analysis in Medicine and Pharmacology: An Overview*, pages 1-13. Kluwer, 1997.

[6] T. M. Cover, P. E. Hart, *Nearest neighbor pattern classification*, IEEE Transactions on Information Theory J., Vol. 13, p21-27: 1967.

[7] E. H. Han, *Text Categorization Using Weight Adjusted k-Nearest Neighbor Classification*, PhD thesis, University of Minnesota, October 1999.

[8] N. Cristianini, J. Shawe-Taylor *An Introduction to Support Vector Machines*, Cambridge University Press, Cambridge, UK, 2000.

[9] T. Y. Kwok *Support Vector Mixturefor Classification And Regression Problems*, in Proc. 14th International Conference on Pattern Recognition, 1998.

[10] John H. Holland, J. Holyoak K, R. E. Nisbett, and P. R. Thagard. *Induction: Processes of Inference, Learning, and Discovery.*, MIT Press, Cambridge, MA, 1987.

[11] L. Breiman, J. Friedman, R.Olshen, and Stone C. *Classification and Regression Trees.*, Wadsworth International Group, 1984.

[12] Wolpert, D.H., and Macready, W.G. *No Free Lunch Theorems for Optimization*, IEEE Transactions on Evolutionary Computation, 1997

[13] Ely JW, Osheroff JA, Ebell MH, et al. *Analysis of questions asked by family doctors regarding patient care*, British Medical Journal, 1999; 319(7206):358–61

[14] Covell DG, Uman GC, Manning PR. *Information needs in office practice: are they being met?*, Annals of Internal Medicine, 1985; 103(4):596–9.

[15] Eneida A. Mendonca, James J. Cimino, Stephen B. Johnson, and Yoon-Ho Seol *Accessing Heterogeneous Sources of Evidence to Answer Clinical Questions*, Journal of Biomedical Informatics 34, 85–98 (2001)

[16] International Health Terminology Standards Development Organisation (IHTSDO) *SNOMED-CT*, http://www.ihtsdo.org/snomed-ct/

[17] United States National Institutes of Health *Unified Medical Language System*, http://www.nlm.nih.gov/research/umls/

[18] United States National Institutes of Health *PubMed*, http://www.pubmed.nl/

[19] Dina Demner-Fushman, and Jimmy Lin *Answering Clinical Questions with Knowledge-Based and Statistical Techniques*, Computational Linguistics, 2007

[20] Ely, John W., Jerome A. Osheroff, M. Lee Chambliss, Mark H. Ebell, and Marcy E. Rosenbaum. *Answering physicians' clinical questions: Obstacles and potential solutions.*, Journal of the American Medical Informatics Association, 12(2):217–224.

[21] Pierre Jacquemart, and Pierre Zweigenbaum *Knowledge and reasoning for medical question-answering*, Proceedings of the 2009 Workshop on Knowledge and Reasoning for Answering Questions

[22] RA Fisher *The use of multiple measurements in taxonomic problems*, Annals of eugenics, 1936

[23] JR Quinlan *C4. 5: programs for machine learning*, Morgan Kaufmann Publishers, 1993

[24] Corinna Cortes and Vladimir Vapnik *Support-vector networks* , Machine Learning, Volume 20, Number 3, 273-297, 1995

[25] Igor Jurisica, John Mylopoulos and Eric Yu *Using Ontologies for Knowledge Management: An Information Systems Perspective*, In Proceedings of the 62nd Annual Meeting of the American Society for Information Science, 482–496, Oct. 31 - Nov, 1999.

[26] L., Poole David and K., Mackworth Alan *Artificial Intelligence: Foundations of Computa-*

tional Agents, Cambridge University Press, New York, NY, USA, 2010

[27] *Resource Description Framework (RDF)*, http://www.w3.org/RDF/

[28] *Notation 3 (N3): A readable RDF syntax*, http://www.w3.org/TeamSubmission/n3/

[29] *Euler Proof Mechanism*, http://eulersharp.sourceforge.net/

[30] *Notation 3 Cideas about web architecture*, http://www.w3.org/DesignIssues/Notation3.html

[31] Kataria, Pavandeep and Juric, Radmila. *Sharing e-Health Information through Ontological Layering*, IProceedings of the 2010 43rd Hawaii International Conference on System Sciences, 2010.

[32] G. Rudolph *Convergence analysis of canonical genetic algorithms* IEEE Transactions on Neural Networks, Jan 1994, 96 - 101

[33] http://neamh.cns.uni.edu/MedInfo/snop.html

[34] P. Vivekanandan and R. Nedunchezhian *A Scalable Genetic Algorithm for Rule Discovery in Large Datasets.* International Journal of Soft Computing, 98-103, 2010

[35] S.Sujatha and N.Pappa *GMC Design Based On SVM Model of Batch Reactor.* International Journal of Computer Applications, 2010

APPENDIX

Semantic Knowledge Store

Knowledge (Triple) Store:
All information is represented as triples
prefix rdfs: <http://www.w3.org/2000/01/rdf-schema# >.
prefix : < # >.
Model Representation
:Person a rdfs:Class.
:Patient rdfs:subClassOf :Person.
:HealthCareProvider a :Person.
:Doctor rdfs:subClassOf :HealthCareProvider.
:EmergencyResponseTeam rdfs:subClassOf :HealthCareProvider.
:MedicalProcedure a rdfs:Class.
:Drug a rdfs:Class.
:HomeopathicDrug rdfs:subClassOf :Drug.
:AllopathicDrug rdfs:subClassOf :Drug.
:Anticoagulant rdfs:subClassOf :AllopathicDrug;
  :name "Anticoagulant";
  :contraIndication :Bleeding,:Pregnancy.
:Insulin rdfs:subClassOf :AllopathicDrug;
  :name "Insulin";
  :precaution :PancreaticCancer;
  :contraIndication :Hypoglycemia;
  :contraIndication :Pregnancy.
:Viagra rdfs:subClassOf :AllopathicDrug;
  :name "Viagra"; :contraIndication :Nitroglycerin.
:Nitroglycerin rdfs:subClassOf :AllopathicDrug.
:MagicPowder rdfs:subClassOf :HomeopathicDrug;
  :name "Magic Powder"; :replacementFor :Insulin.
:Disease a rdfs:Class.
:Diabetes rdfs:subClassOf :Disease.
:Hypoglycemia rdfs:subClassOf :Disease.
:Ulcer rdfs:subClassOf :Disease;
  :name "Ulcer"; :condition :Bleeding.
:Cancer rdfs:subClassOf :Disease.
:PancreaticCancer rdfs:subClassOf :Cancer.
:MedicalCondition a rdfs:Class.
:Hemorrhage rdfs:subClassOf :MedicalCondition.
:Pregnancy rdfs:subClassOf :MedicalCondition.
:Bleeding rdfs:subClassOf :MedicalCondition.

Observed and Factual Knowledge
:John a :Patient;
:name "John";
:hasDisease :Hypoglycemia,:PancreaticCancer.
:James a :Patient;
:name "James";
:hasDisease :Ulcer.
:Mary a :Person;
:name "Mary";
:condition :Pregnancy.
:Dr.Smith a :Doctor.
:ERT.Jane a :EmergencyResponseTeam.

Inference Rules

replacementfor
{?DRUG1 :replacementFor ?DRUG2
} ⇒ {?DRUG1 :exchanged ?DRUG2}.
canNotBeGiven
{?DRUG1 :contraIndication ?DRUG2.
} ⇒ {?DRUG1 :canNotBeGiven ?DRUG2}.
{ ?PATIENT :hasDisease ?DISEASE.
?DRUG :contraIndication ?DISEASE.
} ⇒ {?PATIENT :canNotBeGiven ?DRUG}.
{ ?ANY :condition ?CONDITION.
?DRUG :contraIndication ?CONDITION.
} ⇒ {?ANY :canNotBeGiven ?DRUG}.
{ ?PATIENT :hasDisease ?DISEASE.
?DISEASE :condition ?CONDITION.
?DRUG :contraIndication ?CONDITION.
} ⇒ {?PATIENT :canNotBeGiven ?DRUG}.

EXHAUSTIVE AND HEURISTIC APPROACHES FOR MINIMIZING DIMENSIONALITY AND MISCLASSIFICATION COST

Parag C. Pendharkar
Penn State Harrisburg
email: pxp19@psu.edu
phone: (717)-948-6028

ABSTRACT

We consider a special type of dimensionality reduction classification problem where the decision-making objective is to minimize misclassification cost and attributes (MMCA). We propose a two-stage solution approach for solving the MMCA problem. Using simulated data sets and different misclassification cost matrices, we test our two-stage exhaustive and simulated annealing procedures. Our results indicate that the simulated annealing heuristic approach provides competitive performance on test datasets and is computationally efficient.

1. INTRODUCTION

Supervised classification problem solving involves learning a classification function using a multi-dimensional input vector of decision-making attributes and an output target decision variable. Traditionally, the input decision-making attributes are either selected by hand or by an expert, and the classification algorithm uses all the input decision-making attributes to learn a classification function that predicts the target output decision variable [22].

For datasets containing a large number of input decision-making attributes, an automatic approach for identifying a low cardinality subset of relevant input decision-making attributes is required [22]. The selection of low cardinality relevant inputs reduce the organizational cost of

information acquisition of recording and monitoring information on irrelevant attributes that have low or no impact on decision-making. The procedure of identification of lower cardinality subset of relevant input decision-making attributes is called attribute *dimensionality reduction* or *variable elimination* or *feature selection*. Several real-world classification problems require dimensionality reduction [6][10] and research suggest that dimensionality reduction can not only reduce the organizational overhead costs, but also improve the performance of a classifier because irrelevant attributes often degrade the performance of a learning algorithm [22].

There are a variety of dimensionality reduction approaches proposed in the literature. Stracuzzi and Utgoff [22] categorize these approaches into two categories: *filter* methods and *wrapper* methods. The primary difference between filter and wrapper methods is that the filter methods select input decision-making attributes independent of the learning algorithm, whereas the wrapper methods make learner-dependent selection of input decision-making attributes.

The filter methods use statistical or information theoretical measures to reduce dimensionality. Among the techniques used in filter methods are cross entropy, correlations, principal component analysis, χ^2 test and distance metrics [12][7][5][14][8]. Wrapper methods, on the other hand, use search based methods for selecting input decision-making attributes. These search based methods include greedy search [4], backward elimination beam search [1], nearest neighborhood search [13], backward best first search [11], randomized search [22], and heuristic genetic algorithm search [23] algorithms. The search methods evaluate each candidate set by executing the learning algorithm on each candidate set to evaluate its merit. A general consensus in the literature is that wrapper methods outperform filter methods [22].

Given that wrapper methods perform well for dimensionality reduction, we propose two search based wrapper methods for our research. The first wrapper method uses exhaustive

backtracking search and considers all possible subsets of cardinality greater than or equal to two input decision-making attributes. The exhaustive backtracking search method always finds the optimal sub-set of input decision-making attributes that best predict the output classification objective value. The major drawback of the exhaustive backtracking search is that it is computationally intensive and is unrealistic for large number of input decision-making attributes. Our second wrapper method uses simulated annealing heuristic to identify an input decision-making attributes subset of cardinality greater than one. The simulated annealing heuristic procedure does not guarantee an optimal solution, but uses realistic computational time to provide a heuristic solution.

For learning a classification function, we use an objective that minimizes misclassification cost as opposed to learning target output decision variable (correct classification). Learning to maximize correct classification is a special case of minimizing misclassification cost where costs of all the errors are equal. In designing our classification algorithm, we consider some principles of support vector machines (SVM) where a classification function that minimizes misclassification cost also minimizes an upper bound on generalization error [3]. While there are several classification algorithms that allow decision-makers to minimize misclassification costs, Pendharkar and Nanda [18] showed that genetic algorithm based misclassification cost minimizing classifiers generally perform best and are more flexible. Thus, we use a genetic algorithm (GA) based linear classifier that minimizes the misclassification cost and provides a separating plane that maximizes the margin of separation between classes. We call our classifier max-margin genetic algorithm (MMGA).

The rest of the paper is organized as follows. In section two, we formally present the problem of minimizing misclassification cost and attributes (MMCA); and propose a framework

of two stage solution approach. In section three, we describe the MMGA classifier for classification problems. In section four, using the proposed framework of two stage solution approach, we describe two hybrid approaches, backtracking-MMGA (BT-MMGA) and simulated annealing MMGA (SA-MMGA), for solving the MMCA. In section five, we describe our simulated data, experiments and results. In section six, we conclude our paper with a summary.

2. THE MMCA CLASSIFICATION PROBLEM AND A TWO STAGE SOLUTION APPROACH

In order to mathematically describe the MMCA classification function learning problem, we assume that $X = \{x_1, x_2, \ldots, x_n\}$ is a vector n input attributes that are available for use by a machine learning classification algorithm that minimizes misclassification cost. Let $U= \{u_1, u_2, \ldots, u_n\}$, $u_i \in \{0,1\}$, $\forall i \in \{1,..,n\}$ be a binary vector that indicates the input attributes considered by a classification algorithm.   Further assume a vector $Z \subseteq X$ defined as $Z=\{x_i \mid \forall i, u_i \neq 0\}$. If $D = \{d\}$ is an output decision, where $d \in \{0,1\}$ represents whether an example belongs to an accept class or not, then the MMCA problem can be represented as $f(Z) \rightarrow D$. The problem consists of finding a combination of misclassification cost minimizing classification function $f()$ and a vector Z such that $f(Z)$ has the lowest misclassification cost among all possible values of vector Z.   Since Z depends on U, which is a binary vector of cardinality n, there are a total of 2^n-1 non-null unique values for vector Z that have to be considered for identifying a classification function and attribute set with lowest misclassification cost. To avoid solutions with only one input

attribute[1], we impose a constraint $|Z| > 1$, where $|Z|$ denotes cardinality of set Z. Our

constraint leads to a total of $2^n - (n+1)$ possible solution search space for finding a set of input

attributes that provide the lowest misclassification cost. Since from the Binomial theorem, we

have $2^n = \sum_{g=0}^{n} \binom{n}{g}$, the search space complexity of the MMCA problem is exponential.

In our research, we use a two-stage solution procedure to solve the MMCA. Our solution

procedure consists of systematically identifying all values of U, learning $f(Z)$ for each value of

U using a misclassification cost minimizing MMGA algorithm (described in next section), and

selecting the optimal value U^* that provides the lowest misclassification cost (optimal solution).

Given the search space complexity is factorial, for small values of n, all possible values of U can

be identified and $f(Z)$ can be solved for all values of U. However, for large values of n, it is not

practical to solve $f(Z)$ for all possible values of U. A heuristic approach can be used for a large

value of n, and a sub-optimal heuristic solution can be obtained for MMCA problem in

reasonable time.

Figure 1 illustrates a two-stage general framework for solving the MMCA problem. In

the first stage at the top of the figure, we identify different values of U and related Z using either

an exhaustive search or a heuristic technique. In the second stage, we solve a classification

problem using the attributes identified in the first stage and a MMGA procedure. The first and

second stages work iteratively and sequentially until all the values of U and related Z are

identified or certain termination criterion is met. Optimal or heuristic solution(s) is (are)

returned as an output at the end of a run.

---

[1] We impose this constraint to create a classifier with more than one variable. We assume that the classification problem is sufficiently complex to require the use of two or more variables.
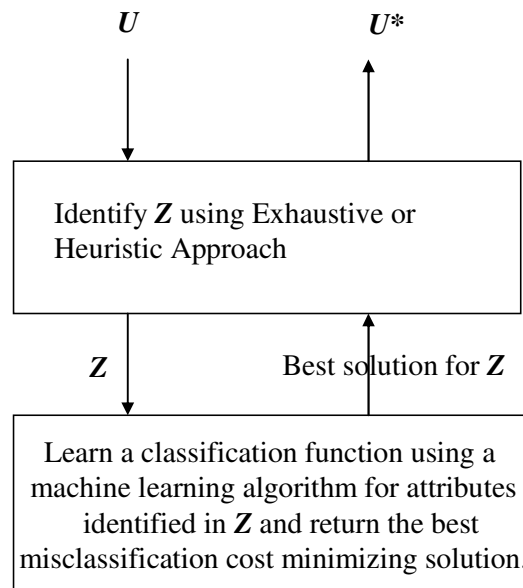
Figure 1: A general framework for solving the MMCA Problem

3. The MAX-MARGIN GENETIC ALGORITHM (MMGA) CLASSIFIER

There are several linear classification models that can be used for classification. Popular among these models are Fisher's linear discriminant analysis model, mathematical programming and linear genetic algorithm model [2]. Nanda and Pendharkar [15] show that when it is desired to minimize misclassification costs, linear GA based models are the most computational friendly, accurate, flexible and do not make any data distribution assumptions. We use Nanda and Pendharkar's [15] approach to build a misclassification cost minimizing linear classification model. Unlike Nanda and Pendharkar [15], where the linear classification model is used to minimize misclassification cost, our model is designed to first minimize misclassification cost and then minimize the distances between the misclassified examples from the separating classification decision plane.

When a GA is used for a binary classification problem, each population member in a GA is a potential solution. For example, assume a data set of ten attributes represented by an attributes vector $\boldsymbol{x}$. The population member (POP_1) used to learn the coefficients for a linear discriminant function will consists of a coefficient vector ($\boldsymbol{w}$) and an intercept (q), which can be represented as follows:

$$POP_1 = \langle w_1, w_2, \ldots, w_{10}, q \rangle,$$

$$where, \quad w_1, w_2, \ldots, w_{10}, q \in \Re.$$

Each population member represents a discriminant function of the following form:

$$w_1 x_1 + w_2 x_2 + \ldots + w_{10} x_{10} + q = 0.$$

The classification heuristic can be represented as:

IF $w_1 x_1 + w_2 x_2 + \ldots + w_{10} x_{10} + q \geq 0$ Then Class=G_1

ELSE Class = G_2.

The individual elements of the vector $\boldsymbol{w}$ and q are called genes (coefficients & intercept) of a given population member POP_1.

Since GA uses *survival of the fittest* strategy to evolve its population, a key design issue for GA is design of fitness function. For minimizing the misclassification cost, following simple fitness function can be used to evolve populations of higher fitness:

$$\frac{K \times (\text{Correctly Classified Cases})}{1 + (\text{Cost of Type I Error} \times \text{Total Type I Errors}) + (\text{Cost of Type II Error} \times \text{Total Type II Errors})}$$

The Type I error is an error of misclassifying an example into class 1 when it should be correctly classified into class 2. Type II error and misclassification costs are defined similarly. The fitness function represented above is always a positive number greater than or equal to 0, and the

denominator of the fitness function can never be zero for positive error costs. K can be any fixed non-zero positive number and may be equal to the total number of examples in the sample. The aforementioned fitness function has several desirable properties. For example, the fitness of the population member can be increased by either increasing the total number of correct classifications or by decreasing the total number of misclassifications. In case of asymmetric misclassification costs, a reduction of total misclassification cost would result in higher fitness.

For our MMGA, in addition to minimizing total misclassification cost, we also minimize the distance between the classification decision surface and the misclassified examples. To illustrate the concept of how we consider the distances between the misclassified examples and classification decision surface, we assume a two-dimensional problem with a classification decision line shown in Figure 2. In Figure 2, it can be seen that only two examples are misclassified and the distances related to these two misclassified examples are s_1 and s_2. Using simple coordinate geometry, we can compute the value of s_1 as follows.

$$s_1 = \frac{w_1 x_1 + w_2 x_2 + q}{\sqrt{w_1^2 + w_2^2}}.$$

The value of s_2 can be computed similarly. In order to include the distances of misclassified examples in the fitness function of a genetic algorithm, we take the sum of absolute values of the distances of misclassified examples and normalize them by sum of the absolute values of distances of correctly classified examples. Using Figure 2, we define a term called normalized total distance of misclassified cases from the classification surface, η, as follows.

$$\eta = \frac{\sum_{i=1}^{2} |s_i|}{\sum_{j=1}^{6} |d_j|}.$$
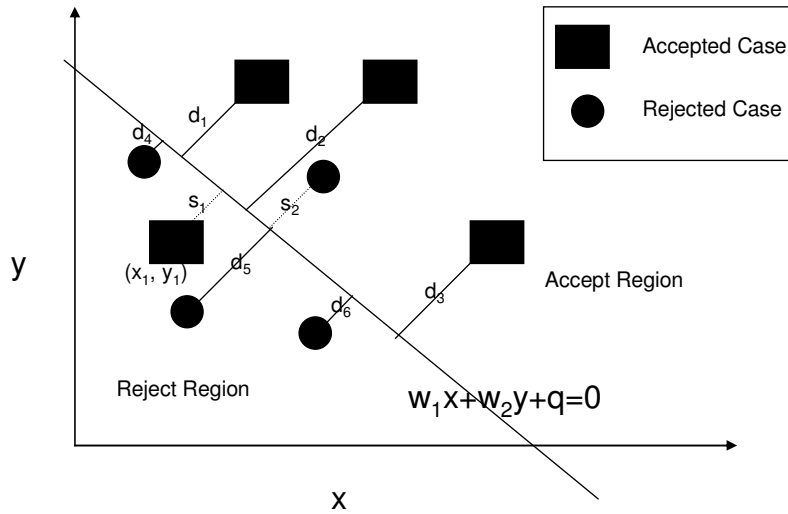
Figure 2: A linear classification decision surface for binary classification

The normalization of distances of misclassified examples ensures that η takes a value which is usually less than one. A fitness function for MMGA that first minimizes the total misclassification cost and then minimizes the distances of misclassified examples from the classification decision surface can be written as follows.

$$\frac{K*(\textit{Correctly Classified Cases})}{1+(\textit{Cost of Type I Error * Total Type I Errors})+ (\textit{Cost of Type II Error * Total Type II Errors})+\eta}$$

The small values of η guarantees that minimization of misclassification cost takes a higher precedence.

4. THE BT-MMGA AND SA-MMGA APPROACHES FOR SOLVING MMCA PROBLEM

In this section we propose two hybrid approaches for solving the MMCA problem. The first approach uses hybrid backtracking exhaustive search (in first stage of Figure 1) and the

MMGA (in the second stage of Figure 1) for solving the MMCA problem. The second approach uses simulated annealing as a heuristic search algorithm for the first stage of Figure 1 and a MMGA in the second stage.

Using backtracking exhaustive search [21], we obtain the elements of set U by generating all 2^n possible solution vectors and then eliminating solution vectors where the cardinality of the solution vector is one or zero. For each feasible solution of U we compute the value of Z and solve a classification problem using the MMGA in second stage.

For the heuristic SA algorithm [9], a random feasible solution is selected and, beginning with this solution in search space, a random move is made in the neighborhood of the initial solution. If the move introduces a better solution (positive move), it is accepted. If the move introduces a worse solution (negative move), it is accepted with a decreasing probability over time.

We define U as a neighborhood of a binary vector $K = \{k_1,\ldots,k_n\}$-- values of which are initially generated at random. The vector U is neighborhood of K if following condition is satisfied:

$N(K)= \{ U \in \{0,1\}^n\colon \|K,U\|=1\}$.

In other words, the neighborhood of K consists of all binary n-tuples in which exactly one entry of K has been changed. The elements of vector U in the neighborhood vector K can be obtained by generating a random integer j such that $1{\leq}j{\leq}n$, and then defining

$$u_j = \begin{cases} k_i & if\ i \neq j \\ 1-k_i & if\ i = j. \end{cases}$$

Figure 3 illustrates pseudocode for SA algorithm used to identify feasible values of U. The values T_0, z_{max}, and α are the initial temperature, maximum iterations and step values

respectively. These values are user defined and are kept constant. The function *MMGA( )* in pseudocode returns the best misclassification cost minimizing classification for a data vector Z associated with the binary values of input parameter vector of U.

We use the MMGA as a misclassification cost minimizing classifier in second stage for both hybrid approaches. We implement two hybrid BT-MMGA and SA-MMGA approaches using C++ programming language in the Visual Studio. Our system was tested rigorously for its accuracy and performance. We experimented with GA parameters of population size, crossover rate, mutation and number of iterations; and found best results were obtained when population size=100, crossover rate=0.3, mutation rate=0.1, and termination iterations=500 respectively. For our SA implementation, we obtained the best results for z_{max}=10, T_0=1000, and α=0.799.

```
z ← 0
T ← T_0
K ← {k_1,…,k_n} = rand{0,1} // random binary vector of cardinality greater than one
U_Best ← K
BestClassification ← MMGA(U_Best)
U ← K
while z ≤ z_max
        let j ← Rand(1,n) // a random integer between 1 and n
        u_j ← 1-k_j
        if (|U| < 2 OR |K| < 2) then U ← Fail
        else U ← Pass
        if U ≠ Fail

do

        then

                diff ← MMGA(U) – MMGA(K)

                if (diff ≥ 0 && MMGA(U) > BestClassification)

                then    U_Best ← U
                        BestClassification ← MMGA(U)

                else    r ← Random (0,1)
                        if r < e (diff)/T

                        then    K ← U


        z ← z+1
        T ← αT

        return (U_best)
```

Figure 3: Pseudocode for SA algorithm for identifying feasible values of *U*

5. SIMULATED DATA, EXPERIMENTS AND RESULTS

We compare the performance of BT-MMGA and SA-MMGA using simulated data that

has been used in the literature [16]. The use of simulated data allowed us to test the robustness

and performance of the two approaches across different experimental conditions [16]. There are

several studies in the literature that have used simulated datasets for testing performance of

classification algorithms [2][17][19][20].   Using the methodology described in Vale and

Maurelli [24], we generate 20 random data samples for a normal distribution. Each of the 20

random data samples for each data distribution consisted of 300 examples of ten independent attributes $(x_1, x_2, \ldots, x_{10})$ equally split between two groups (50:50 ratio). The mean for group 1 was approximately set equal to zero, and the mean of group 2 was approximately set equal to 0.5. The standard deviation for all the attributes was approximately set equal to one and skewness was set equal to zero.

In order to test the impact of misclassification cost asymmetries on our two techniques, we use three different cost matrices for our experiments. These three matrices represent equal misclassification costs (1:1), mildly unequal misclassification costs (2:1) and unequal misclassification costs (4:1). Figures 4a-4c illustrate these three different misclassification cost matrices.

	Reject	Accept
Predicted Reject	0	1
Predicted Accept	1	0

Figure 4a: The first "equal" misclassification cost matrix

	Reject	Accept
Predicted Reject	0	2
Predicted Accept	1	0

Figure 4b: The second "mildly unequal" misclassification cost matrix

	Reject	Accept
Predicted Reject	0	4
Predicted Accept	1	0

Figure 4c: The third "unequal" misclassification cost matrix

Using twenty datasets and three misclassification cost matrices, we perform a total of 120 (20 experiments per misclassification cost matrix) training and testing experiments for each of the two techniques (BT-MMGA and SA-MMGA). Figures 5 & 6 illustrate the results of 60 training and 60 test data sample misclassification costs. Figure 7 illustrate the number of input attributes

used by the techniques to obtain best misclassification cost on the training data. The SA-MMGA used a mean of 6.83 attributes (max= 8; min=6) attributes for 60 training data experiments, whereas the BT-MMGA used a mean of 8.63 attributes (max=10; min=6) attributes.
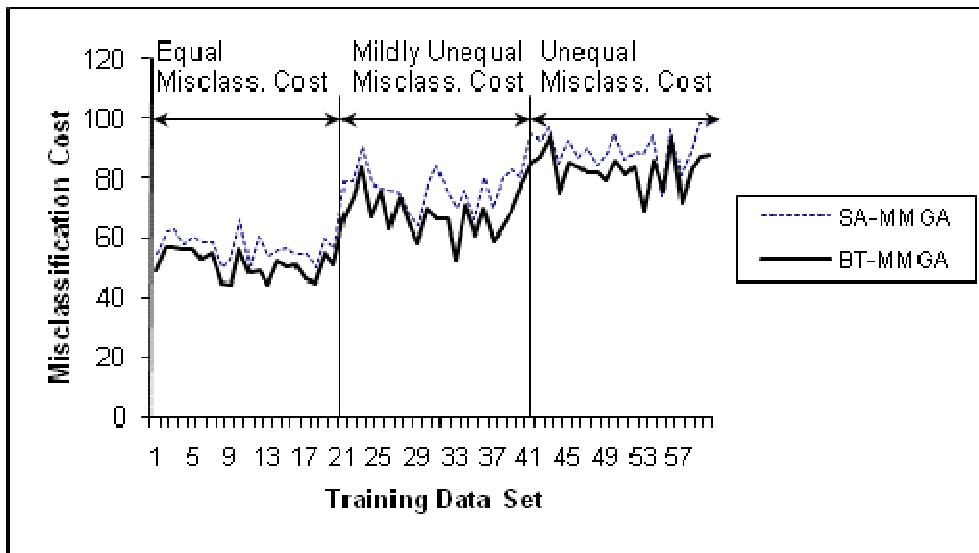
Figure 5: The comparison of misclassification costs for two techniques on training data
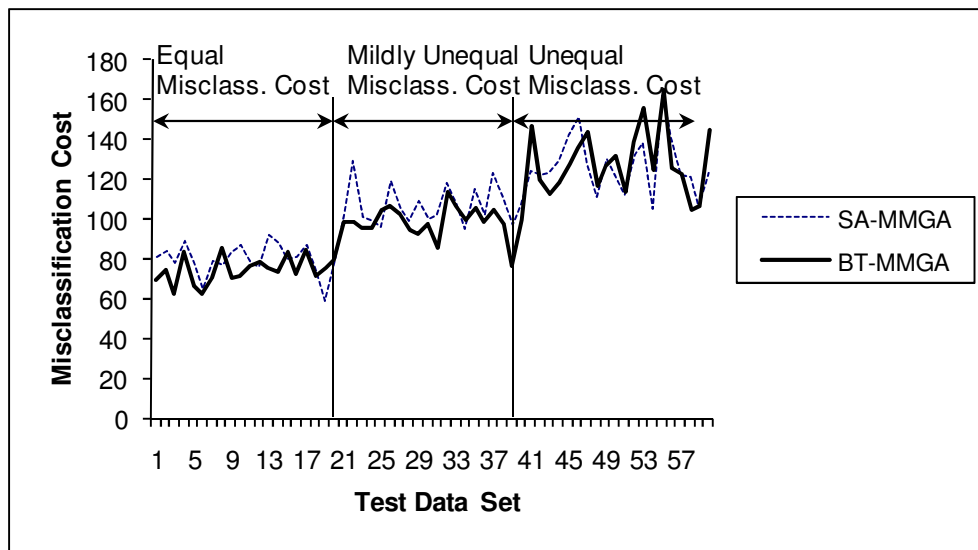
Figure 6: The comparison of misclassification costs for two techniques on test data
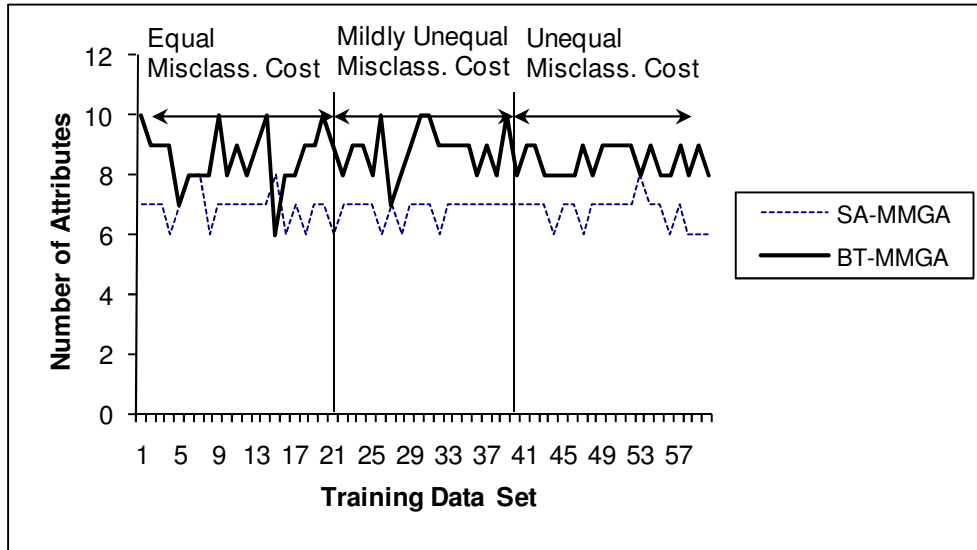
Figure 7: The comparison of the number of attributes retained by two techniques on training data

The average misclassification costs for each of two techniques by three misclassification cost asymmetries on training and test data are shown in tables 1 and 2. The results indicate that, for training data, the two techniques have about 10% difference in their performances. However, for test data the performance gap between the two techniques appears small.

Table 1: The misclassification cost descriptive statistics for training data experiments

Technique	Cost	Mean	Std. Deviation	Sample Size
BT-MMGA	Equal	50.95	4.57	20
	Mildly Unequal	82.90	6.32	20
	Unequal	67.60	7.48	20
	Total	67.15	14.52	60
SA-MMGA	Equal	57.10	4.38	20
	Mildly Unequal	90.05	6.27	20
	Unequal	76.50	6.44	20
	Total	74.55	14.77	60

Table 2: The misclassification cost descriptive statistics for test data experiments

Technique	Cost	Mean	Std. Deviation	Sample Size
BT-MMGA	Equal	73.90	6.74	20
	Mildly Unequal	129.20	16.35	20
	Unequal	98.20	8.10	20
	Total	100.43	25.35	60
SA-MMGA	Equal	79.65	7.85	20
	Mildly Unequal	127.30	14.66	20
	Unequal	106.90	9.54	20
	Total	104.62	23.95	60

We use two-way analysis of variance (ANOVA) to test the impact of technique (SA-MMGA and BT-MMGA), cost asymmetries (equal, mildly unequal and unequal) and the interaction between technique and cost asymmetry on the classification misclassification cost for training and test data. Tables 3 and 4 illustrate the results of our ANOVAs. The results indicate that both factors --technique and cost-- are significant factors in explaining variance in misclassification cost. Further, the interaction between two factors appears to be insignificant factor in explaining variance in both training and test data misclassification costs.

Table 3: The ANOVA summary table for training data experiments

Source	Sum of Sq.	Df	Mean Sq.	F Value	Significance
Corrected Model	22828.00	5	4565.60	126.41	0.000**
Intercept	602366.70	1	602366.70	16678.36	0.000**
Technique	1642.80	1	1642.80	45.49	0.000**
Cost	21146.45	2	10573.22	292.75	0.000**
Technique*Cost	38.75	2	19.38	0.54	0.586
Error	4117.30	114			
Total	629312.00	120			
Corrected Total	26945.30	119			

R-Squared= 0.847 (Adjusted R Squared = 0.840); ** significant at 99%

Table 4: The ANOVA summary table for test data experiments

Source	Sum of Sq.	Df	Mean Sq.	F Value	Significance
Corrected Model	54117.18	5	10823.44	87.08	0.000**
Intercept	1261365.08	1	1261365.08	10148.79	0.000**
Technique	525.01	1	525.01	4.22	0.042*
Cost	52993.55	2	26496.78	213.19	0.000**
Technique*Cost	598.62	2	299.31	2.41	0.095
Error	14168.75	114	124.29		
Total	1329651.00	120			
Corrected Total	68285.93	119			

R-Squared= 0.847 (Adjusted R Squared = 0.840); ** significant at 99%; * significant at 95%

It is important to note that the technique factor is a significant in explaining variance in misclassification cost for test data only when the level of significance is 95%. Since test misclassification cost is a better performance metric than training misclassification cost, it appears that the heuristic SA-MMGA provides a very competitive performance when compared with the exhaustive BT-MMGA. On average the differences of performance between BT-MMGA and SA-MMGA test misclassification cost were less than 8% of the BT-MMGA performance. The SA-MMGA provides this close performance with, on average, two fewer input attributes than BT-MMGA and 85% quicker than the BT-MMGA computational running time.

6. SUMMARY AND CONCLUSIONS

We proposed an MMCA dimensionality reduction classification problem and provided two, exhaustive and heuristic, procedures to solve the MMCA problem. Using simulated datasets, three different misclassification cost matrices and proposed solution procedures, we solved the MMCA problem. Our experiments indicate that the simulated annealing heuristic

approach provides computationally efficient competitive performance when compared to optimal solution provided by the exhaustive search approach.

Our two stage framework is a general framework which is suitable to include other misclassification cost minimizing classifiers such as probabilistic neural networks. Future research may focus on comparing the performance of MMGA and other misclassification cost minimizing classifiers for solving the MMCA problem. For large scale problems containing over 10 attributes, it may be desirable to develop procedures to establish lower bounds on overall misclassification cost. These lower bounds can provide a good benchmark for performance comparisons of the heuristic techniques.

References

1. D. W. Aha, and R. L. Bankert. Feature selection for case-based classification of cloud types: An empirical comparison, *Working Notes of the AAAI-94 Workshop on Case-Based Reasoning*, 106-112, 1994.

2. S. Bhattacharyya and P.C. Pendharkar. Inductive evolutionary and neural techniques for discrimination. *Decision Sciences* 29, 871-899, 1998.

3. P. S. Bradley, U. M. Fayyad, and O. L. Mangasarian. "Mathematical programming for data mining: Formulations and challenges, *INFORMS Journal on Computing* 11 (3), 217-238, 1999.

4. R. Caruana, and D. Freitag. Greedy attribute selection, Machine Learning: *Proceedings of the Eleventh International Conference*, New Brunswick, NJ, 1994.

5. G. H. Dunteman. *Principal component analysis*, Sage Publications, Newbury Park CA, 1989.

6. S. E. Fahlman, and C. Lebiere. The cascade-correlation learning architecture, *Advances in Neural Information Processing Systems*, 2, 524-532, 1990.

7. M. A. Hall. Correlation based feature selection for machine learning. Ph. D thesis, Department of Computer Science, University of Waikato, Hamilton, New Zealand, 1999.

8. K. Kira, and L. Rendell. A practical approach to feature selection, Machine Learning: *Proceedings of the Ninth International Conference*, Morgan Kaufmann, San Francisco, CA, 1992.

9. S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by simulated annealing, *Science*, 220(4578), 671-679, 1983.

10. J. Kivinen, and M. K. Warmuth. Additive versus exponentiated gradient updates for linear prediction, *Information and Computation*, 132, 1-64, 1997.

11. R. Kohavi, and G. H. John. Wrappers for feature subset selection, *Artificial Intelligence* 97 (1-2), 273-324, 1997.

12. D. Koller, and M. Sahami. Toward optimal feature selection, Machine Learning: *Proceedings of the Fourteenth International Conference*, 284-292, 1996.

13. P. Langley, and S. Sage. Oblivious decision trees and abstract cases, *Working Notes of the AAAI-94 Workshop on Case-Based Reasoning*, AAAI Press, 1994.

14. H. Liu, and R. Setino. Feature selection via discretization, *IEEE Transactions on Knowledge and Data Engineering* 9 (4), 642-645, 1997.

15. S. Nanda, and P.C. Pendharkar. Linear models for minimizing misclassification costs in bankruptcy prediction, *International Journal of Intelligent Systems in Accounting, Finance and Management*, 10, 155-168, 2001.

16. P. C. Pendharkar. Hybrid approaches for classifications under information acquisition cost constraint, *Decision Support Systems*, 41, 228-241, 2005.

17. P. C. Pendharkar, and J. A. Rodger. An empirical study of impact of crossover operators on the performance of non-binary genetic algorithm based neural approaches for classification, *Computers & Operations Research*, 31, 481-498, 2004.

18. P. C. Pendharkar, and S. Nanda.A misclassification cost minimizing evolutionary-neural classification approach, *Naval Research Logistics*, 53(5), 432-447, 2006.

19. P.C. Pendharkar. A computational study on the performance of artificial neural networks under changing structural design and data distribution, *European Journal of Operational Research*, 138, 155-177, 2002.

20. P. C. Pendharkar. An empirical study of design and testing of hybrid evolutionary-neural approach for classification, *Omega: An International Journal of Management Science*, 29, 361-374, 2001.

21. E. Rich, and K. Knight. *Artificial Intelligence*, McGraw Hill Publishing, New York, 1991.

22. D. J. Stracuzzi, and P. E. Utgoff. Randomized variable elimination, *Journal of Machine Learning Research*, 5, 1331-1362, 2004.

23. H. Vafaie, and K. D. Jong. Genetic algorithms as a tool for restructuring feature space representations, *Proceedings of the International Conference on Tools with AI*. IEEE Computer Society Press, 1995.

24. C. D. Vale, and V. A. Maurelli. Simulating multivariate non-normal distributions, *Psychometrika* 48, 465-471, 1983.

Extended Abstract
How can an organization develop and retain competent IT workforce: the role of IT leadership

This study aims to provide a systematic framework with which business organizations learn to develop competent IT professionals and retain them. Derived from both IT academics and practices, the proposed model is based on a comprehensive set of critical knowledge and skills currently required on the job and uses IT leadership as a key factor playing an important role. Specifically, the study focuses on examining *how* leadership behaviors on the part of IT managers/executives influence their IT personnel to develop and enhance their knowledge and skills sets. This will, in turn, increase IT personnel's needs satisfaction, quality-of-work-life (QWL) to improve the chance of retaining those IT personnel (with high in their needs satisfaction). In so doing, the study expects to provide IT practitioners with practical insights in how to develop currently required IT skills and retain the qualified IT workforces for long-run to maintain the level of the performance of the IT unit.

Introduction

As organizations are facing increasing challenges in managing global operations as they become more interlinked through technology and networks. Accordingly, an organization's Information Technology (IT) unit assumes more strategic responsibilities and executes activities that develop, operate, and manage IT, provides functional needs, and contributes to clients meeting their strategic goals (e.g., IT development and management)(Chan, Huff, Barclay, & Copeland, 1997). As a result, organizational expectation towards the IT unit has certainly changed IT professionals' job definitions and related roles (Fondas & Stewart, 1994). This in turn leads to the changes in critical knowledge and skills set required. Current IT workers are expected to not only have in-depth knowledge in technology, but also possess softer, non-tech skills to get the job done and provide organizations with solutions to their business problems (e.g., problem-solving, ethics and tolerance, oral and written communication, collaboration and team, business analysis, and functional area knowledge). Overall, IT professionals are needed to have their core competencies in operations and IT management, supported by a well-rounded business foundation.

However, as baby-boomers retire and a lack of interest in IT careers, it is becoming a great challenge to attract (and/or develop) qualified and competent IT professionals with the '*right'* skills set to meet organizations' such IT needs (e.g., IT professionals who speak both business and IT languages and can make a good sense of IT in the business context) and retain them for a long run (Luftman, 2007a). In fact, this has been one of the big concerns for many incumbent IT leaders including Chief Information Officers (CIO). This study aims to provide a systematic framework with which business organizations learn to develop competent IT professionals and retain them. Derived from both IT academics and practices, the proposed model is based on a comprehensive set of critical knowledge and skills currently required on the job and uses IT leadership as a key factor playing an important role. Specifically, the study focuses on examining *how* leadership behaviors on the part of IT managers/executives influence their IT personnel to develop and enhance their knowledge and skills sets. This will, in turn, increase IT personnel's needs satisfaction, quality-of-work-life (QWL) to improve the chance of retaining those IT personnel (with high in their needs satisfaction).

IT Leadership

IT leadership is defined as a set of consistent behaviors displayed by an organization's IT managers/executive to influence their personnel to attain the IT unit's goal (Bass, 1985). IT leadership is conceptualized in terms of two behaviors: transactional and transformational leadership behaviors. Even though, both behaviors are considered to be important for effective leadership (Bass, 1985), the study focuses on transformational leadership as it is likely to be more effective in influencing IT personnel who tend to possess unique characteristics valuing intrinsic nature of the work (e.g., highly-educated, intellectually-driven, professional, artistic, and favorable of independent and autonomous work environment).

Transformational IT leaders motivate followers by transcending the followers' self-concepts (self-efficacy, esteem, development, intrinsic value of work, etc.). It has four types of behaviors: idealized influence (IIF), inspirational motivation (IM), intellectual stimulation (IST), and individualized consideration (IC). IIF involves being a role model by displaying exceptional capabilities and strong conviction towards the vision. IM focuses on articulating a compelling vision, providing meaning and challenge to their work, and inspiring by expressing high expectations and confidence. IST involves in encouraging thinking outside of the box and approaching old situations with innovative and creative ideas. IC focuses on paying attention to IS personnel's individual needs for achievement and development by acting as a mentor (Bass, 1985).

Elements of aforementioned transformational IT leadership behaviors are expected to impact IT personnel's unique characteristics in ways that they are motivated to develop and/or improve their knowledge and skills (Wynekoop & Walz, 1998). For instance, IT personnel (e.g., programmers, systems analysts, network administrators, database administrators and designers, etc.) are generally a highly-educated, intellectually immersed, and trained work-force, who tend to possess a high degree of professionalism, value autonomy at work and artistry of their accomplishments (Brancheau & Hoffman, 1987). Thus IT professionals are likely to be more ambitious, self-confident, and creative workers. They also have tendency to pursue opportunities for personal growth, cultivate horizontal relationships with external referents, and give high credence to peer-review processes (Conger, 1999; Locke, 1968).

IT Personnel's Critical Knowledge and Skills

To champion organizational strategy effectively, IT professionals should have accurate specification of clients' IT needs and business priorities so that an organization's IT resources will be implemented in line with their strategic directions and objections. That means that IT personnel should be multi-faceted and multi-talented professionals who understand both IT and business functions: this requires skills in diverse areas in addition to technical skills such as communication (verbal and written), teamwork (works well with others), interpersonal (relates well to others), problem-solving (with reasoning and analytical critical thinking), creativity, selling, leadership/management (especially in project management) to name the few. This is in line with what is required for the next generation of IT professionals: to possess skills set including technical knowhow and softer, non-tech skills to get the job done (Luftman, 2007a). These skills set generally includes the following (Luftman, 2007a, , 2007b):

1. Problem-solving: IT is to reconcile business problems,
2. Communication (both oral and written): since, regardless of technical skills, IT professionals must be able to effectively present how IT can contribute and enhance different/idiosyncratic business operation/process,

3. Collaboration (including team and/or project management): IT professionals must be able to work with business partners, other IT professionals, and vendors,
4. Business analysis: IT professionals must understand their business and industry to successfully leverage technology to help their clients to compete,
5. Functional area/domain knowledge: in addition to technical skills, IT professionals must possess specialized skills in various business operations to effectively contribute to their organizations.

Quality of Work-Life (QWL)

QWL is found to be positively related to various organizational affective variables such as job satisfaction (Danna & Griffin, 1999) and organizational commitment (Mowday, Porter, & Steers, 1982). QWL is defined as employee satisfaction with a variety of needs through resources, activities, and outcomes stemming from participation in the workplace and measured as employees' psychological results of evaluations of the products of organizational work: discrepancy between outcome (e.g., economic rewards, promotion opportunities, challenges, co-worker relations, etc.) and standard and the weight of each outcome (e.g., expectations, values, motives, wants, social comparisons, etc.) (Efraty & Sirgy, 1990). The differences between outcomes and standards are weighed differently by the personal value of each outcome. When desired work outcomes such as performance and satisfaction are achieved, individuals are likely to experience three psychological states of experienced meaningfulness, experienced responsibility, knowledge of results (Hackman & Oldham, 1976).

Substantive autonomy, clear role descriptions, team work, involvement in the solutions of work problems, and learning opportunities are known to positively affect QWL (Nandan & Nandan, 1995), while as excessive workloads, forced overtime, and ambiguous or conflicting role demands cause emotional distress among employees and lower QWL (Menaghan & Merves, 1984).

IT Leadership developing/retaining IT Personnel through their QWL

Transformational IT leadership behaviors are expected to enhance IT personnel's knowledge and skills set. A transformational IT leader is likely to influence IT personnel by articulating the significance of their work for the organizational strategic goals and transcending their values and beliefs tying their vision with organizational vision (IIF). A transformational IT leader transcends IS personnel's perception of users as valued customers on whom IT personnel's livelihood depends. Such an IT leader can transform IT personnel's view on their roles by showing strong conviction to his IT vision, articulating and reinforcing the overall IT role (e.g., "IT as a strategic weapon") till the message was transpired to the entire IT unit. Such a leader encourage IT personnel to take customers' insights and suggestions for improvements be a liaison to properly educate IT personnel on how to serve their in and outside clients (e.g., how to communicate with customers, how to ask and understand whether IT can offer product/service they wanted, which IT product/service offering they would like to see, etc.).

IT personnel come to realize the value in developing those traits such as communication, personal traits, organization and negotiation. This will transform the IT unit to be more service-oriented and IS personnel to focus more on people-skills (Veal III, 2000). Thus, IT personnel will become personable, approachable and responsive to users in ways to concern more with their business problems and IT needs.

Efraty, D., & Sirgy, M. J. (1990). The effects of quality of working life (QWL) on employee behavioral responses. *Social Indicators Research, 22*(1), 31-47.

Fondas, N., & Stewart, R. (1994). Enactment in managerial jobs: A role analysis. *Journal of Management Studies, 31*(1), 83-103.

Hackman, J. R., & Oldham, G. R. (1976). Motivation through the design of work: Test of a theory. *Organizational Behavior and Human Decision Processes, 16*, 250-279.

Heller, M. (2000, June, 15). How are you improving IS-user relations? *CIO Magazine*.

Locke, E. A. (1968). Toward a theory of task motivation and incentives. *Organizational Behavior and Human Performance, 3*, 157-189.

Luftman, J. N. (2007a). 6 key skills CIOs seek in entry-level workers. *CIO Insight* Retrieved October 23, 2007,
from http://www.cioinsight.com/slideshow/0,1206,pg=0&s=304&a=217671,00.asp

Luftman, J. N. (2007b). Top 10 management concerns of CIOs. Retrieved October 23,
from http://www.cioinsight.com/slideshow/0,1206,pg=0&s=300&a=217341,00.asp

Menaghan, E. G., & Merves, E. S. (1984). Coping with occupational problems: The limits of individual efforts. *Journal of Health and Social Behavior, 25*, 406-423.

Mowday, R. T., Porter, L. W., & Steers, R. M. (1982). *Employee-organization linkages: The psychology of commitment, absenteeism, and turnover*. New York, NY: New York: Academic Press.

Nandan, S., & Nandan, M. (1995). Improving quality of care and quality of work life through interdisciplinary health care teams. In *Developments in Quality-of-Life Studies in Marketing* (Vol. 6, pp. 80-86): Academy of Marketing Science and the International Society for Quality-of-Life Studies.

Veal III, R. (2000, March 15). Ending the user-IS cold war. *CIO Magazine*.

Wynekoop, J. L., & Walz, D. B. (1998). Revisiting the perennial question: Are IS people different? *Database for Advances in Information Systems, 29*(2), 62-71.

Social Media as a Moderating Factor of Hostility
Alexander Pelaez - *Hofstra University*
acsazp@hofstra.edu

Introduction

Social media has received a significant amount of attention in the media for both its positive and negative effects on society. Today, social media accounts for 22% of all time spent online in the United States according to a report from Nielsen in June 2010. With such a large population engaging in exchanging messages, sharing pictures and videos, and playing games within the context of social groups, it is important to understand the effect of social media technologies on behavior of individuals within these groups.

Social media technologies have enabled behaviors such as Cyberstalking, and Cyberbullying (Smart and Salinas 2010), which have led researchers to attempt to understand how technologies might alter behaviors. Researchers in past have shown how electronic communication, or computer mediated communication (CMC) attenuates social cues and thus leads to self-centered, and unregulated behavior (Sproull & Kiesler, 1986) due to the decreased anonymity (Zimbardo, 1969; Sproull & Kiesler, 1986; Lea 1991). If forms of CMC lead to more unregulated behavior can technologies such as social media amplify the level of a participant's hostility?

This paper proposes an experiment to understand the effect of participant's measures of hostility after being ostracized in a social media setting. Leveraging prior work of experiments conducted on behaviors of ostracized individuals, established through an exclusion condition of a ball tossing game (Williams & Sommer, 1997), the experiment will seek to understand the moderating the effect of social media on hostile attitudes in a similar exclusion condition. Based on literature review, it is presumed that participants will demonstrate higher feelings of hostility in a social media environment than in a physical environment. An additional condition to be tested includes participant's beliefs of a direct exclusion condition versus a condition where the participant doesn't know if they are being excluded or if a technical failure has occurred. Electronic mediums may experience technical failures and the uncertainty of a breakdown in communication may lead to a feeling of exclusion based on expectations of a response (Taylor and Harper 2003). Thus end result of the experiment will be a 2x3 between subjects design to explore the effect of direct or unknown ostracism within the social media and compare the levels of hostility to the same measures after the exclusion condition is conducted in a physical environment.

Hostility

Measuring aggression, hostility and anger is complicated by the difficulty in establishing clear differences between definitions (Siegman & Smith, 1994). Hostility represents a belief in the devaluation of the worth of others, and that others are a source of wrong doing, whereas, aggression is an overt behavior including destructive and harmful actions, and anger is an emotion as a result of some behavior or action (Siegman & Smith, 1994). Measuring an item such as hostility is further made difficult because hostility contains three components including cognitive, affective and behavioral, and thus any meaningful measure of hostility must contain elements of these components.

Research in aggression or hostility through technology is not unique and has been conducted through varying communication mediums, e.g. video games (Arriaga, et. al. 2006; Barthalow, et. al 2005), movies (Anderson 1997) and CMC (Keisler, et. al. 1984; Sproull & Keisler, 1986); however, there are different ways to measure aggressive behavior including hostility. Anderson and Dill (2000) proposed a General Aggression Model (GAM) adapted from prior models, which explains the process that heightens interpersonal aggression (Anderson & Dill 2000; Arriaga, et. al. 2006). The model focuses on the influences of aggressive behavior through the cognitive, affective, and arousal route (Anderson, Dill 2000; Arriaga, et. al. 2006). Inputs, i.e. situational factors such as pain, frustration or attacks, may operate between any one of the routes (Arriaga, et. al. 2006), and a sufficiently strong activation may increase affect factors in the other routes (Anderson 1997). The overall increase of measures along the affective route, such as hostility, may increase aggressive behavior. However, by definition, it would be important

1

for any study to have a destructive or harmful action in order to fully measure aggression, and therefore certain measures of aggression may threaten the validity of the construct. Limitations measuring aggression were discussed in detail by Tedeschi and Quigley (1996), as they sought to explain the threats to construct validity in measuring aggression; however, Giancola and Chernmak (1998) argued against the view that questions the validity of constructs stating that Tedeschi and Quigley "grossly" underestimated the generalizability of the constructs used to test aggression in laboratory settings.

Therefore, although it would be preferable to measure aggressiveness for our experiment, it is important to look for a different construct that can be more easily measured. Since hostility is a belief, it should be easier to utilize a simple instrument that can capture these feelings adequately. A review of the literature yields an instrument recently developed and used called the State Hostility Scale (SHS) (Andersen & Morrow 1995). While there are two types of hostility, state and trait, state hostility is the one most relevant since it deals with hostile behavior as a result of some form of provocation. The SHS scale has been used effectively to measure increases in hostility scores under various conditions such as pain, provocation, violent movie clips, and temperature differences (Anderson 1997; Anderson et. al. 1995). Therefore the SHS appears to give us a good measure of hostility that has been empirically validated and is easy to replicate.

Ostracism

In order to generate feelings of hostility, some form of provocation must be conducted on a participant. Ostracism has been shown as a method used by many different cultures over the centuries, within organizations, and at an individual level (Williams & Sommer 1997). Williams and Sommer (1997) specifically researched the effect of ostracism on social loafing by a participant. Social ostracism occurs where a participant has the perception of being ignored by others in one's presence, which can lead to serious negative reactions that can be used to be manipulate others (Willams & Sommer 1997). An ostracized individual is deprived of four fundamental needs: belonging, self-esteem, control, meaningful existence (Williams & Sommer, 1997).

The ostracism experiment created a condition whereby a participant along with two confederates formed a group; the true nature of the experiment was not disclosed to the participant. While waiting for the experiment to begin, one of the confederates would pick up a ball from a toy box and begin a ball tossing game between the other confederate and the participant. After a few tosses, in the exclusion, or ostracized condition, the participant would no longer be thrown the ball, thus ostracizing the participant in a nonverbal manner (Williams & Sommer 1997). This experiment has been conducted numerous times including through the use of a computer (Williams, Cheung, Choi 2000) to measure cyberostracism. In that experiment the participant did not know other players were not involved and that condition was merely a computer simulation. The study found that participants still reported negative attitudes even when not in the presence of others.

Other experiments have been conducted on ostracism including chat rooms (Williams, et. al., 2002) and using cellphone or SMS messages (Smith & Williams, 2004). In the latter experiment, participants, after an initial set of messages, were either included in the continuing conversation or received no messages (Smith Williams, 2004). These participants reported more negative feelings, worse mood, and wrote more provocative messages in response. These provoking responses are consistent with the expectations that an ostracized individual will be aggressive at getting back in the group (Williams & Sommer, 1997).

There are two ways in which the ostracism condition could be created. First is by overt exclusion, a condition in which the target knows they are being excluded; however, what about the condition in which the target doesn't know that they are being excluded? This condition is very likely possible in CMC since messages might not be delivered due to communication failure. For example, email messages may be sent to a spam folder, dropped by the Internet Service Provider (ISP), or the sender's computer may have crashed before a message was sent. In any of these cases, the communication between the target and sender has been severed and may lead the recipient to an incorrect conclusion about the situation. This situation has been explored, somewhat, within SMS messages and it was found that when a recipient didn't receive an expected response an underlying agreement was broken and left the individual feeling excluded (Taylor and Harper 2003).

2

The ball tossing experiment has been extensively used to study the effect of ostracism; however, to date, a review of the literature has not yielded an experiment conducting a similar experiment in a social media environment. Social media isn't much different from other CMC technologies; however, it is important to test the effect of social media since its primary use is to create a network of people who share common values or interests. Studies of this particular form of CMC can potentially identify factors of the technology that could moderate the effect.

Social Media and Electronic Communication

Electronic Communication can mean a variety of technologies such as email, chat rooms, instant messaging and social media. While the Information Systems field is relatively young, there is a large body of research on different aspects of IT artifacts. Various forms of electronic communication have permeated every aspect of our lives and therefore it is virtually impossible today find people whose sole method of communication is face to face. Understanding the impact of these technologies is critical in understanding their effects on behavior.

Technologies are impersonal in nature and they have been shown to break down barriers of traditional communication. Research has shown that electronic communication is likely to attenuate social context cues. (Sproull & Keisler, 1986). Social context cues are important in framing behaviors during conversations and communications. For example, when office colleagues are communicating, static cues, such as office space, titles, etc. are significant inhibitors for communication certain messages (Sporull & Keisler, 1986). Dynamic cues such as appearance, tonality, and body language are also a significant factor. Sproull and Keisler (1986) show that dynamic cues are completely eliminated in email conversations, and static cues are minimized. The result demonstrated that individuals were more likely to engage in conversation, increase conflict and disagree than they were in person. Thus, electronic communication enabled and empowered individuals regardless of social context.

However, the results of the study went beyond just providing more opinions and simple disagreement. Sproull and Keisler (1986) further demonstrated that people behaved more irresponsibly on electronic communication than face-to-face. This irresponsibility was directly attributed to the absence of social cues which act as boundaries for conversation. Overall, absence of social cues can lead to **more uninhibited behavior** such as verbal aggression, blunt disclosure, and non-conforming behavior (Whitty & Gavin 2001).

The behavior demonstrated by Sproull and Keisler wasn't limited to just individuals. Many electronic communication mediums have numerous conversations between multiple parties. Many times individuals with in the groups can behave irresponsibly as well; overall research has shown that groups that communicate electronically exhibit **more anti-social behavior** (Siegel, et. al 1983).

It is critical to understand why the anti-social, uninhibited and unregulated behavior occurs. Once again attention is focused back on the social context and the cues of the conversation. Information exchange between parties is influenced through perception, cognitive interpretation and communication behavior (Sproull & Kiesler, 1986). Perception, cognitive interpretation and communication behavior are factors of the dynamic and static cues, and therefore these cues influence the social context of information exchange. When the social context cues are removed or weakened, the result will be an increase in perceived anonymity, which produces self-centered, unregulated behavior (Sproull & Kiesler, 1986). Classical work in behavior has shown that anonymity increases aggressive behavior, known as deindividuation theory (Zimbardo, 1969).

Therefore, the literature review demonstrates that electronic communication significantly reduces social context cues. The reduction in social context cues leads to an increase in anonymity and thus leads to unregulated, uninhibited and more anti-social behavior at both a group and individual level.

How does social media more specifically play into the effect of electronic communication or CMC? One of the aspects of social media is the purpose behind which users are engaging in activities through the IT artifact. Social media users are more connected to each other through some set of values of interests. The

proliferation of social media is directly attributed to the externalities of the communication medium, i.e. they are positive benefits to others in the network (Kraut, et. al. 1998). Users generally benefit from these types of systems if others who use it are important to them (Kraut, et. al. 1998). Forms of social media technologies only serve a use when others are using it, but more importantly when those who use it share some interest.

Social media differs from other electronic communications technologies in appearance and functionality. Overall, it provides a richer user interface than email, chat rooms and instant messaging tools, even though social media has incorporated these features. Social media adds features such as photos, video sharing, audio, tagging, and posting, all of which enhance the user experience. However, what impact does this have on communication? Media Richness Theory attempts to describe the effects of different media. It states that the effectiveness of the communication will improve based on the medium (Daft & Lengel, 1984). Richer mediums such as video conferencing are able to reproduce social cues such as body language, but email cannot. As referenced above, attenuated social cues result in anti-social and uninhibited behavior, therefore, the richer the medium, one would expect to find more controlled behavior and therefore less aggressiveness or hostility. The theory further states that the richness of a particular format is derived from capacity for immediate feedback and the degree to which intent is focused on the recipient (Daft & Lengel, 1984). Since social media's functions are normally directed and feedback is more immediate than email, it can be inferred that social media is richer than email. Therefore, utilizing the definition of Media Richness Theory social cues should not be attenuated. It is believed that even though immediacy and directed intent are inherent in the technology, this richness has no effect on enforcing social cues. Dennis (1998) found no support for the central idea of Media Richness supporting earlier work that reached the same conclusion (Kinney & Watson, 1992; Valacich, et. al, 1994).

An argument against CMC leading to deindividuation lies in Social Identity Theory (Richer, et. al 1995; Spears et al. 2002). It has been argued that the CMC technologies do not lead to deindividuation but rather to depersonalization, making group level identities more important (Spears, et. al 2002), and thus the content and group norms mediate the positive or negative effects (Postmes, et. al , 1998). This would mean that the technology itself could not be a moderating factor, but rather any change in behavior would be attributable to the group and the situation presented. Unfortunately, this is contrary in some ways to the original work by Keisler and Sproull (1986). Proponents of Social Identity model of Deindividuation Effects (SIDE) argue that anonymity by itself is not enough, but the effects of anonymity are only valid through the social context (Postmes, et. al. 1998).

Even though a review of the literature provides some evidence to the effect that technology may not affect behavior as much as the context of their usage, there is enough evidence to suggest that social media technologies may attenuate social cues and therefore lead to more aggressive behavior. The competing ideas in the literature serve to emphasize the need for more research of the effect of technologies on behavior.

Therefore, this experiment, based on the work in psychology and information systems, attempts to explore the behavioral change when exposed to social media. More specifically, under a provoked situation, a participant's level of hostility should be higher in a social media condition, than under a face to face condition, and thus the experiment will test two hypotheses:

H1: Ostracism, compared to inclusion, will lead to higher levels of hostility, and the effect will be amplified when using social media technology as compared to an in-person condition.

H1a: Ostracism, compared to inclusion, will lead to higher levels of hostility, and the effect will be amplified when using social media technology when the ostracism is obvious, compared to exclusion that is not obvious.
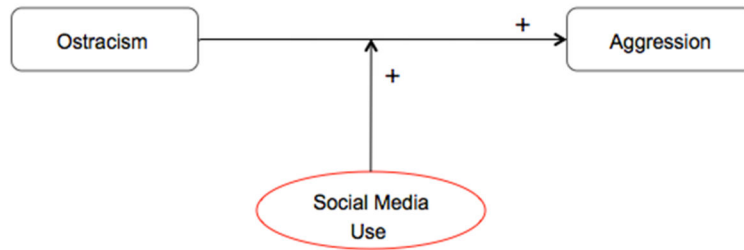
Method

Participants

The study will solicit volunteers from a pool of undergraduate students of a major northeastern university. Since three experiments will be conducted using a 2x3 between subjects design, it is desired to have at least 120 participants in the study, yielding at least 20 subjects within each group. The participants will be randomly assigned to the groups and will be tested for their knowledge of social media. Each participant should have some familiarity (usage of social media at least once a week) with sites such as Facebook, Twitter or MySpace.

Procedure

Each participant, once they arrive at the laboratory setting will be taken to a room with two confederates. Preferably the room will be set up like a lounge (e.g. couch, chairs, etc.) and will be a comfortable setting where the two confederates and the participant can converse. The goal of this initial setting is to create a social-like connection, whereby the participant can socialize. The purpose is to make the participant more likely to establish a connection with the confederates and ultimately be willing to engage in conversation in the social media experiment.

After 5 minutes of casual conversation, the participant and the confederates will be told they will be participating in an experiment to test new social media features for product development. Once the experiment is discussed the participant will be brought to a room for the real purpose of the experiment

Inclusion / Exclusion in Face to Face Setting

Participants will be brought to a room where the original exclusion/inclusion condition, originally designed by Williams and Sommer (1997), will be conducted. The participants will be seated in a room along with the two confederates. After the experimenter leaves the room for a few minutes, one confederate will reach into a box labeled children's toys, pull out a ball and begin tossing the ball to the other confederate and participant.

For one group of participants, the confederates will continue to throw the ball an equal number of times until the experimenter returns the room five minutes later. This condition will be the face-to-face inclusion condition. For the other participants, after 1 minute, the confederates will cease throwing the ball to the participant and not make eye contact or speak to the participant, exactly as conducted in the Williams and Sommer experiment (1997). This condition is considered the face-to-face exclusion condition.

5

Social Media Instrument

In order to create an effective means of testing the social media condition, a simple yet familiar social media instrument must be created. The instrument will have a similar look and feel to Facebook but will only contain three main widgets on the screen.

Widget 1

First, the central widget will be the cyberball tossing game where the participant will see a triangle with themselves at one point and the two confederates on the other two points. The widget will contain two buttons labeled "Throw to P1" and "Throw to P2", enabling the participant to the cyberball when they have possession, identified as a circular graphic ball next to the participants' name.

Widget 2

The second widget will be a text chat area where the participant can communicate to the other participants using a very familiar text area and where any messages that have been typed in by either participants or confederates can be seen.

Widget 3

The third widget will be a sliding scale that will measure the mood of the participant on a scale of feeling positive and feeling negative. This widget could be useful as participants attempt to convey their feelings in a graphical manner. The information collected will be a bipolar scale from -10 to 10, not visible to the participant.

Inclusion / Exclusion in Social Media

Similar to the face-to-face experiment, the participant will be seated at a computer workstation in a room separate from the confederates who will be seated at two workstations in a different room. Prior to leaving the room, the experimenter will ask the participant to test the chat area by typing "hello" in the text box. One confederate will reciprocate; this is to demonstrate the use of the chat widget, and verify operability. In addition, the experimenter will ask the participant to slide the mood scale all the way to the top (feeling positive) and all the way to the bottom (feeling negative) and then return the slider to the neutral position, again to demonstrate the use of the widget and verify operability. After the experimenter leaves the room, one confederate will "throw" the cyberball to the other confederate. Since the participant can see the ball on the screen being thrown, the participant will know the confederates are playing the game. The confederates will be given a script to type messages into the chat box upon each throw such as "Nice Throw" and "Throw it back". The confederates, similar to the face-to-face condition, will throw the ball to the participant to get the participant engaged in the game.

For one group of participants, the confederates will continue to throw the ball an equal number of times until the experimenter returns the room five minutes later. This condition will be the social media inclusion condition. For the other participants, after 1 minute, the confederates will cease throwing the ball to the participant and cease any text messages in the text area. This condition is considered the social media exclusion condition.

Inclusion / Exclusion in Social Media (not visible)

This condition is conducted in the exact same manner as the previous condition; the only difference is the main widget, or the ball-throwing widget. This widget will be altered to show only when the participant has the ball. It will not show the ball being thrown between confederates. In this case, the experiment is looking to test whether knowing if the ball is being thrown, or not, results in the same exclusion condition. This is adapted from the qualitative research regarding SMS messages (Taylor and Harper 2003). All aspects of this test including the throwing of the ball and the scripted text messages from the confederates are the same.

Measures

Once the experiment has been completed, the participant will be removed from the room and asked to fill out a brief questionnaire. This questionnaire will conduct a manipulation check and test the level of hostility using the State Hostility Scale.

Manipulation Check

A simple manipulation check will be included once the experiment has concluded. The participant will be asked a single item from 0 *(not at all)* to 9 *(very much)* on these two items: "I was excluded" and "I was ignored" (Smith and Williams 2004). This check will verify whether or not the exclusion condition was successful.

Hostility

State hostility will be measured using the State Hostility Scale (Anderson and Morrow 1995). The State Hostility Scale is a 35-item 5-point Likert Scale ranging from 1 (strongly disagree) to 5 (strongly agree). The State Hostility Scale has shown to have an internal reliability above .90, even though three items have shown weakness, vexed, tender, and willful (Anderson 1997). The SHS scale has 4 subscales (Unsociable, Mean, Lack of Positiveness and Aggravation), which will be used to provide more granular analysis.

Analysis

Upon completion of the experiment, the data will be collected from the data sets, hostility measures, mood scale, and the messages. Each will provide valuable insight into the results of the experiment.

The State Hostility Scale data will be collected and analyzed across the four subscales. For each of the subscales, Chronbach's alpha will be calculated for reliability and F-value will be used to examine the effect of the manipulation across each of the six conditions.

The single data point can be used to test validate mood measures from the State Hostility Scale, however, of even more value, might be the recorded time of change in mood. It is possible to record the time the participant changes the mood scale, and as such examine the frequency of changes in comparison to the exclusion condition. This can provide insight into how long the participant waits before they are affected by the exclusion.

The text messages can provide a rich qualitative set of data that can further explain the feelings of the participant. Independent coders will be given the data to attempt the classification of messages looking for the best categories for identifying the messages. The classifications might correspond with the measures identified by Williams and Sommer (1997) for the conditions that ostracism threatens (Belonging, Self-Esteem, Control and Existence).

Expectations

It is expected upon conclusion of the experiment that social media will amplify the effect of ostracism, and thus, the level of hostility measured by ostracized participants will be greater for those who used social media than those who did not. In addition, because CMC may exhibit technical problems, i.e. the user doesn't know that an exclusion condition is present, it is expected that participants who are knowingly excluded in social media will exhibit higher levels of hostility than those who cannot see that they are being excluded. Those participants who cannot see they are being excluded will question whether or not there is a problem with the technology before concluding that they are being excluded. Furthermore, it is expected, at the very least, that the realization of the exclusion condition will occur later when the participant cannot see they are being excluded. This should be evident from the mood scale and the text messages.

A transformational IT leader is also likely to encourage IT personnel to approach the traditional ways with innovative methods in solving business problems and incorporating priorities and IT needs for users (IST), devote considerable time and effort to understand users' learning curve and anxieties they encounter with IT (IM; IC). Such a leader can encourage IT personnel to use their creativity and critical thinking abilities gave autonomy to IT personnel in making key timely decisions in solving business problem (e.g., changing the line of product/services customized to customers' needs). An IT leader can also encourage IT personnel to be innovative and creative, think outside the box in finding ways for IT to serve businesses better and champion their strategies. In so doing, the IT cultivates the environment in which IT personnel's efforts are publicly recognized in front of their peers at the regular meetings, retreats, etc., which in turn encourages IT personnel to be more productive and innovative in what they do. This will direct IT personnel to develop mentors' skills and traits such as better communicator, superior speaking and domain knowledge, great degree of empathy, tolerance and patience, and such, which will be critical in handling users as protégé (Heller, 2000).

IT professionals' skills are expected to be positively related to their evaluation of needs satisfaction at work. As their competencies are developed or improved in line with what is expected and/or required, IT professionals' self-efficacy and self-regulation of motivation to work will be increased (Bandura, 1986; Menaghan & Merves, 1984). This will lead IT professionals to feel more confident, be more active, and seek innovative, persuasive ways to work with others. Accordingly, the enhanced skills and confidence facilitate the achievement of desired outcomes in terms of need satisfaction such as IT professionals' social, esteem and actualization needs (e.g., IS personnel feel good about themselves by getting recognition for their work and enhancing their professional skills, competencies, and potential).

Concluding Remarks

This study expects to provide IT practitioners with practical insights in how to develop currently required IT skills and retain the qualified IT workforces for long-run to maintain the level of the performance of the IT unit as a whole, and eventually enhance the overall organizational performance. Currently the study is work-in-progress and stands at the early stage of data collection using web-based survey.

References

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall.

Bass, B. M. (1985). *Leadership and performance beyond expectations*. New York: Free Press.

Brancheau, J. C., & Hoffman, T. C. (1987). Managing information systems for effectiveness and humanity: Applying research on organizational behavior. *Information & Management, 13*(5), 233-243.

Chan, Y. E., Huff, S. L., Barclay, D. W., & Copeland, D. G. (1997). Business strategic orientation, information systems strategic orientation, and strategic alignment. *Information Systems Research, 8*(2), 125-150.

Conger, J. (1999). Charismatic and transformational leadership in organizations: an insider's perspective on these developing streams of research. *The Leadership Quarterly, 10*(2), 145-179.

Danna, K., & Griffin, R. W. (1999). Health and well-being in the workplace: A review and synthesis of the literature. *Journal of Management, 25*(3), 357-384.

Limitations

There are some significant limitations to this experiment. First, measuring hostility, anger and aggression is difficult (Siegman & Smith, 1994). Using measures such as the State Hostility Scale are an attempt to measure the attitudes given some form of provocation. While the current literature states the validity of the State Hostility Survey, this may not be the best approach to measure hostility after provocation. Furthermore, the implication of this experiment is that hostility is a proxy for aggressive behavior. A measure of aggression is more difficult from an operational and ethical perspective. To adequately measure aggression the experiment would need to have an overt intention to cause physical or emotional harm (Siegman & Smith, 1994), and this may not be practical due to ethical considerations. Furthermore, it is difficult to conduct an actual experiment that has aggression using the definition stated, although an experiment was conducted measuring the amount of hot sauce a participant put in a confederates cup after an ostracism condition (Warburton & Williams, 2006).

Another potential issue is the social media instrument with the three widgets may be too complex for this particular experiment. While each widget has a distinct purpose, the main ball tossing widget is the primary component. The other two widgets are for more data collection and may confuse the participant, thereby leading to some inaccurate or inconclusive results. More specifically, the text chat component needs to be clearly defined to ensure that it is simple to use and the participants understand how to use it and that only the two confederates are available. Finally, the mood widget will need to be matched to other mood scales that have been found empirically valid. Further research into various mood scales is required to ensure that the mood scale widget is appropriate and clearly understood by the participant.

From a technology perspective, the instrument developed is a simple attempt to reconstruct a popular social media tool. While its simplicity makes the experiment more practical, the lack of a broader network reach and lack of more widgets may threaten the external validity of the experiment; however, this seems to be a reasonable tradeoff for experimental purposes. Finally, while the instrument is simple, it must be developed using a qualified software developer and thoroughly tested to ensure that no technical difficulties arise. This could require funding or resources, e.g. computers, server and network, that may not be available.

Another concern is the replication of the 'implicit' ostracism condition, whereby the participant cannot see that they are being ostracized. While not proposed in this experiment, it may be possible to replicate this specific "online" condition in the physical world. It is recommended that future researchers explore the possibility of separating the participants and confederates into different rooms and instead of ball tossing experiment, pass notes through an "intermediary", another researcher. However, instead of writing notes the note would be an innocent game of "tag" similar to the online condition, found in Facebook, of "poking". This could be a useful instrument for future research.

Implications for Future Research

It is expected that the results of this experiment could lead to other research in the social media space. Based on the results, it would be important researchers to explore the real causes of anonymity online to identify specific factors leading to higher levels of aggression in an online setting. The result could be an exploration of factors within social media that might lessen the conditions of hostility or aggression. One possible explanation for how this could be done falls under Media Richness Theory. For example, if social cues are attenuated by social media, then an exploration into how social cues be reintroduced may prove important, for example, the effect of video sessions and audio sessions within social media. Work has been done to test the effect of video conferencing, but these were done in an organizational setting (Rice 1992), where video conferencing was more available, but also where social cues exist more prominently.

The mood scale could also be interesting for research into social media. Mood scales are widely used and serve different purposes; however, a more acute use in the social media space may lead to some compelling results. Future uses of this mood scale should be tied to an empirically validated mood scale.

If findings are confirmed, researchers could focus more attention on aggression from a social media space. Use of available constructs may prove valuable in the study of aggression online as opposed to the study of

hostility. A nice study could include both an online aggression component, such as turning off someone's computer remotely, compared to a physical aggression model, such as the hot sauce experiment conducted by Warburton and Williams (2006).

On a different note, researchers could use this type of result to study the amplifying effect of social media on positive behaviors such as giving online. With the number of charitable institutions seeking donations online, through social media and cell phones, it could be shown that social media amplifies all behaviors both positive and negative, which could be significant.

Discussion

Social media's explosion has surprised many technology experts. The negative and positive effects of social media continue to generate a significant amount of press in the mainstream media. Social media serves a deeper purpose for many users that goes beyond the information hunting use of Internet in the late 1990's and early 2000's. End users today are looking for more social interaction online as a means to connect or reconnect with individuals. These technologies may be generally the same as older technologies such as email or video conferencing, but their use has become more personal and as such behavioral patterns will change. It is therefore important to understand how these positive and, possibly more important, negative behaviors may or may not change through the use of social media technologies.

Appendix 1: Social Media Instrument #1

This instrument shows the position of the ball between the three players.
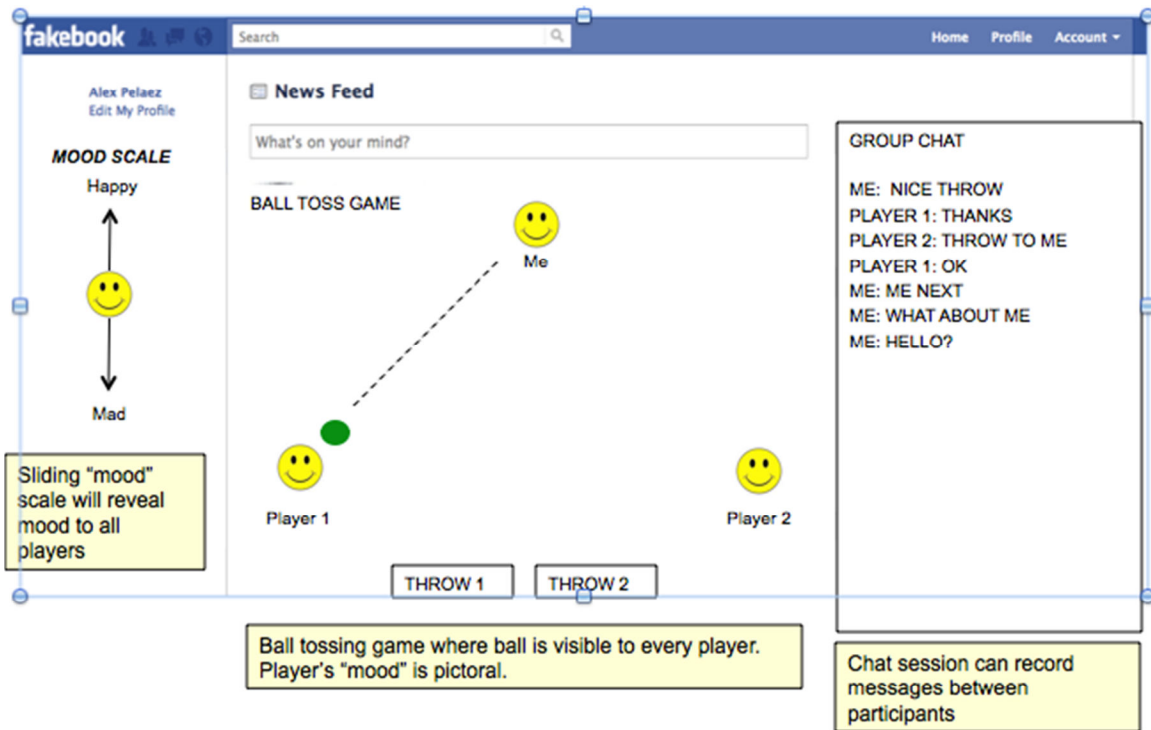
Appendix 2: Social Media Instrument #2

This instrument doesn't show the participant who possess the ball or even if it's being thrown.

References

Anderson, C. (1997). Effects of violent movies and trait hostility on hostile feelings and aggressive thoughts. *Aggressive Behavior, 23*, 161.

Anderson, C. A., & Dill, K. E. (2000). Video games and aggressive thoughts, feelings, and behavior in the laboratory and in life *Journal of Personality and Social Psychology, 78*(4), 772.

Anderson, C. A., Deuser, W. E., & DeNeve, K. M. (1995). Hot temperatures, hostile affect, hostile cognition, and arousal: Tests of a general model of affective aggression. *Personality and Social Psychology Bulletin, 21*(5), 434-448. doi:10.1177/0146167295215002

Anderson, C. A., & Morrow, M. (1995). Competitive aggression without interaction: Effects of competitive versus cooperative instructions on aggressive behavior in video games. *Personality and Social Psychology Bulletin, 21*(10), 1020-1030. doi:10.1177/01461672952110003

Arriaga, P., Esteves, F., Carneiro, P., & Monteir, M. B. (2006). Violent computer games and their effects on state hostility and physiological arousal. *Aggressive Behavior, 32*(2), 146-158. doi:10.1002/ab.20111

Bartholow, B. D., Sestir, M. A., & Davis, E. B. (2005). Correlates and consequences of exposure to video game violence: Hostile personality, empathy, and aggressive behavior. *Personality and Social Psychology Bulletin, 31*(11), 1573-1586. doi:10.1177/0146167205277205

Daft, R. L., & Lengel, R. H. (1986). Organizational information requirements, media richness and structural design. *Management Science, 32*(5, Organization Design), pp. 554-571.

Dennis, A. R., & Kinney, S. T. (1998). Testing media richness theory in the new media: The effects of cues, feedback, and task equivocality. *Information Systems Research, 9*(3), 256-274.

Dill, K. E., Anderson, C. A., Anderson, K. B., & Deuser, W. E. (1997). Effects of aggressive personality on social expectations and social perceptions. *Journal of Research in Personality, 31*, 272-292(21).

Giancola, P. R., & Chermack, S. T. (1998). Construct validity of laboratory aggression paradigms: A response to tedeschi and quigley (1996). *Aggression and Violent Behavior, 3*(3), 237-253. doi:DOI: 10.1016/S1359-1789(97)00004-9

Kiesler, S., Siegel, J., & McGuire, T. W. (1984). Social psychological aspects of computer-mediated communication. *American Psychologist, 39*(10), 1123-1134. doi:DOI: 10.1037/0003-066X.39.10.1123

Kinney, S. T., & Watson, R. T. (1992). *The effect of medium and task on dyadic communication*. Dallas, Texas, United States: University of Minnesota.

Kipling D. Williams, & Kristin L. Sommer. (1997). Social ostracism by coworkers: Does rejection lead to loafing or compensation? *Pers Soc Psychol Bull, 23*, 693.

Kraut, R., Mukhopadhyay, T., Szczypula, J., Kiesler, S., & Scherlis, W. (1998). Communication and information: Alternative uses of the internet in households. *In Proceedings of the CHI 98,* , 368.

Lea, M., & Spears, R. (1991). Computer-mediated communication, de-individuation and group decision-making. *International Journal of Man-Machine Studies, 34*(2), 283-301. doi:DOI: 10.1016/0020-7373(91)90045-9

Postmes, T., Spears, R., & Lea, M. (1998). Breaching or building social boundaries? *Communication Research, 25*(6), 689-715. doi:10.1177/009365098025006006

Rice, R. (1992). Task analyzability, use of new media, and effectiveness: A multi-site exploration of media richness. *Organization Science, 3*(4), 475.

Siegel, J., Dubrovsky, V., Kiesler, S., & McGuire, T. W. (1983). Group processes in computer-mediated communication. *Org. Behavior and Human Processes, 37*, 157.

Siegman, A., & Smith, T. (1994). *Anger, hostility and the heart* Psychology Press.

Smart, L., & Salinas, G. (2010). *Social networking breeds new degree of stalking.* Retrieved 11/20, 2010, from

Smith, A., & Williams, K. D. (2004). R U there? ostracism by cell phone text messages. *Group Dynamics: Theory, Research, and Practice, 8*(4), 291-301. doi:10.1037/1089-2699.8.4.291

Social Networks/Blogs now account for one in every four and a half minutes online. (2010). Retrieved 11/20, 2010, from http://blog.nielsen.com/nielsenwire/global/social-media-accounts-for-22-percent-of-time-online/

Sproull, L., & Kiesler S. (1986). Reducing social context cues: Electronic mail in organizational communication. *Management Science, 32*(11), 1492.

Taylor, A. S., & Harper, R. (2003). The gift of the gab? : A design oriented sociology of young people's use of mobiles. *Computer Supported Cooperative Work: The Journal of Collaborative Computing, 12*(3), 267-296.

Tedeschi, J. T., & Quigley, B. M. (1996). Limitations of laboratory paradigms for studying aggression. *Aggression and Violent Behavior, 1*(2), 163-177. doi:DOI: 10.1016/1359-1789(95)00014-3

Valacich, J. S., Mennecke, B. E., Watcher, R. M., & Wheeler, B. C. (1994). Extensions to media richness theory: A test of the task-media fit hypothesis. 11.

Warburton, W. A., Williams, K. D., & Cairns, D. R. (2006). When ostracism leads to aggression: The moderating effects of control deprivation. *Journal of Experimental Social Psychology, 42*(2), 213-220. doi:DOI: 10.1016/j.jesp.2005.03.005

Whitty, M., & Gavin, J. (2001). Age/Sex/Location: Uncovering the social cues in the development of online relationships. *CyberPsychology & Behavior, 4*(5), 623-630. doi:10.1089/109493101753235223

Williams, K., D. Govan, C. L., Croker, V., Tynan, D., Cruickshank, M., & Lam, A. (2002). Investigations into differences between social- and cyberostracism. *Group Dynamics, 6*(1), 65.

Williams, K. D., Cheung, C. K. T., & Choi, W. (2000). Cyberostracism: Effects of being ignored over the internet. *Journal of Personality & Social Psychology, 79*(5), 748-762. doi:10.1037//0022-3514.79.5.748

Zimbardo, P. G. (1969). The human choice: Individuation, reason, and order versus deindividuation, impulse, and chaos. *Nebraska Symposium on Motivation, *, 237.

THE TURING TEST, SEARLE'S CHINESE ROOM, AND E.T.'S CHINESE ROOM

Jinchang Wang
School of Business, Richard Stockton College of New Jersey, Pomona, NJ 08240, USA
Jinchang.wang@stockton.edu , 609-652-4628

ABSTRACT

The Turing Test is an empirical approach and a conceptual idea for testing whether a machine's intelligence level is on a par with humans based on the machine's behaviors and performances. Searle's Chinese Room argument challenges the Turing Test by showing that a machine following syntactical programs does not know semantics of the programs and does not understand the meaning of what it is doing. He claimed that such a dumb machine might pass the Turing Test but it was not intelligent. We develop an argument of E.T.'s Chinese Room to refute Searle's points. E.T.'s Chinese Room shows that using the behaviors due to intelligence, instead of the internal process of generating intelligence, to judge intelligent level of a machine is consistent with the purpose of AI researches, which is to develop intelligent machines with no restrictions on the approaches of generating intelligence.

Key Words: artificial intelligence, Turing test, computer intelligence

The Turing Test was proposed by Alan Turing in 1950 to test machine intelligence through reviewing the machine's behaviors. The idea of Turing Test remains relevant after sixty years and it has been a subject of intense debate among artificial intelligence (AI) scientists and philosophers. John Searle used his Chinese Room argument to show that a machine would not be intelligent if it did not understand what it was doing albeit it looked intelligent. We in this article challenge Searle's Chinese Room argument with an E.T.'s Chinese Room argument, showing that the top priority of AI researchers is to develop machines that act and perform as intelligently as humans do, rather than machines with same internal mechanism as human brains. The idea of the Turing Test serves that purpose well. People may disagree on whether a machine that is able to pass the Turing Test is really intelligent, but they would not have disagreement that a machine that is able to pass the Turing Test would behave intelligently and do intelligent jobs as humans do.

In the first section, we review briefly the Turing Test, its features, its role in AI, and debate on it. In Section 2, we discuss Searle's Chinese Room argument and Searle's "proof" of the dumbness of computers. In Section 3, we present our E.T.'s Chinese Room argument that refutes Searle's Chinese Room argument, and elaborate our position on the issue.

1. THE TURING TEST

At a conference at Dartmouth College in 1956, artificial intelligence (AI) as a subject of computer sciences was formally established. Six years before that, Alan Turing, a British

mathematician and a founder of computer sciences, foreseeing the potential of computer intelligence, put forward a method of testing machine intelligence. The method and the idea behind it are still timely at present, albeit it has been a subject of intense debate in the field of AI and philosophy. In his paper "Computing Machinery and Intelligence" (1950), Turing suggested that, instead of asking whether machines can think, we should ask whether machines can pass a behavioral intelligence test, which has come to be called the Turing Test: - *A machine and a human counterpart are placed in two rooms. An interrogator in the third room, who cannot see the machine and human counterpart, talks to the machine/human through the keyboard and screen. If the interrogator is unable to distinguish the machine from the human, then the machine is assumed to be as intelligent as a human.*

The Turing Test is both a conceptual idea and an empirical method of testing machine intelligence. It is not supposed to test "Can machines think", "How well machines think", "Whether machines think in the way same as humans". Instead, it is supposed to test "How smart machines are". It has three important features, as summarized by Luger & Stubblefield (1989):
(1) It gives us an objective notion of intelligence. That is, the behavior in response to a particular set of questions. This provides a standard for determining intelligence which avoids the inevitable debates over the true nature of intelligence .
(2) It prevents us from being sidetracked by currently unanswerable questions and equivocal conceptions such as internal process of intelligence, whether or not machine understands what it is doing, whether machine is conscious.
(3) It eliminates any bias in favor of living organisms over machine intelligence by focusing on solely the content of the answers to questions.

The Turing Test represents an answer to a fundamental issue in AI: - How to test the quality of the product of AI. If we say that the ultimate task of AI researchers is to make machines with intelligence on a par with human, then developing the method to check the quality of the machine is an imperative task. The Turing Test provides an implementable approach and a conception for that issue. It has been cited and discussed widely in AI literatures. It was in significant early papers of AI collected by Webber and Nilsson (1981) and by Luger (1995), in The Encyclopedia of AI (Shapiro, 1992) which contains survey articles on almost every topic in AI; and in Wikipedia. It was in Haugeland book (1985) which gives an account of the philosophical and practical problems of AI, and in Nillson's book (2009) which provides an insightful and comprehensive history of AI.

Alan Turing conjectured in 1950 that, by the year 2000 a computer could be programmed well enough to pass the test. He was a little too optimistic about the development of machine intelligence – programs have yet to fool a sophisticated human judge. On the other hand, many people are now being fooled when they don't know they might be chatting with a computer. The ELAZA program and Internet chatbots such as MGONZ (Humphrys, 2008) and NATACHATA have fooled their correspondents repeatedly, and the chtbot CYBERLOVER has attracted the attention of law enforcement because of its penchant for tricking fellow chatters into divulging enough personal information that their identity can be stolen. The Loebner Prize competition, held annually since 1991, is the longest-running Turing Test-like contest. (Russell and Norvig 2010)

There have been equivocalities in the field of AI, such as "what is thinking", "what is a machine", "what is intelligence", and "what is mind". Instead of joining in the debate of those ambiguities, Turing proposed the empirical standard, behavior, for machine intelligence, which is more clearly defined and easily to be implemented. He bypassed the equivocalities, but did not resolve them, which left rooms for a number of criticisms. One is aimed at its bias toward purely symbolic problem-solving tasks. It does not test abilities that require perceptual skill or manual dexterity, even though these are important components of human intelligence. Conversely, the Turing Test needlessly constrains machine intelligence to fit a human mold. Do we really wish the machine to do mathematics as slowly and inaccurately as the human does it? Shouldn't an intelligent machine capitalize on its own assets, such as an infallible memory, rather than trying to emulate human cognition? (Luger and Stubblefield 1989). "Art intelligence" and "sport intelligence" are not tested in Turing Test. Shieber (1994) severely criticized the usefulness of the instantiation of Turing Test in the Loebner Prize competition. Ford and Hayes (1995) argued that the test itself was not helpful for AI. Bringsjord (2008) gave an advice for a Turing Test judge. Shieber (2004) and Epstein et.al. (2008) collected a number of debating essays on the Turing Test. Searle's Chinese Room argument (1997) challenged that a machine that passed the Turing Test might not be intelligent because it did not have semantics, which would be discussed in Section 2 and rest of this article.

The idea of Turing Test has been extended to testing full range of intelligence. The so-called *Total Turing Test* includes a video signal so that the interrogator can test the subject's perceptual abilities. To pass the total Turing Test, the computer will need capability of vision to perceive objects, and robotics to manipulate objects and move about. (Russell and Norvig 2010)

Two philosophical concepts related to Turing Test are weak AI and strong AI. As Russell and Norvig defined (2010), the *weak AI hypothesis* asserts that machines could act as if they were intelligent. The *strong AI hypothesis* asserts that machines that look intelligent are actually thinking (not just simulating thinking). According to Searle (1997), *weak AI* advocates that computers are useful tools in doing simulations of the mind; while *strong AI* claims that implementing the right program in any hardware at all is constitutive of mental states. That is, strong AI claims that the implemented program, by itself, is constitutive of having a mind, and the implemented program, by itself, guarantees mental life. Many AI researchers take the weak AI hypothesis for granted, and don't care about the strong AI hypothesis – as long as their program works, they don't care whether you call it a simulation of intelligence or real intelligence.

2. SEARLE'S CHINESE ROOM ARGUMENT

The argument for the Turing Test is that a machine is intelligent if it *looks* intelligent enough. That is, the standard of machine intelligence is its behavior or performance, rather than the internal mechanism and process that produce intelligence. Many have disagreed at this point. They argued that intelligence is a much more sophisticated and superior matter than it "looks", and that a machine looks intelligent might be cheating or be following unintelligent "mechanical" rules. They acknowledge the facts that computers are inexorably and increasingly becoming smarter and smarter, but they sensed something delicate in human spirit and mind that

was missed in the idea of the Turing Test. John Searle, a philosopher at University of California, Berkley, challenged the idea of the Turing Test with his Chinese Room argument (Searle 1997). "They (computers) are immensely useful devices for simulating brain process. But the simulation of mental states is no more a mental state than the simulation of an explosion is itself not an explosion."

Searle's Chinese Room works in this way: - *A person, like Searle himself as he put, who does not know Chinese at all, is locked in a room and is given a rule book. He receives questions in Chinese, looks up the rule book, does operations according to the rules, and give back bunches of symbols of Chinese as the answers. From the views of the people outside the room, he understands Chinese perfectly. But he does not actually understand a single word of Chinese!*

Searle argues by using his "Chinese Room" that computers looks capable of doing intelligent work such as answering questions in Chinese, but they do not understand what they are doing. A machine having passed the Turing Test looks as smart as a human, but it is actually not intelligent since it simply following the rules to do mechanical operations without knowing what it is doing for.

John Searle rejects Strong AI's claim that "the mind is just a computer program". He once tried to "prove" computer programs can't be mind in three steps: (1) programs are entirely syntactical; (2) minds have semantics; (3) syntax is not same as, or sufficient for, semantics. So, computer programs are not minds. (Searle 1997) The "proof" has serious flaws. His central argument is, "programs can't be mind because syntax is not semantics". But he did not define "semantics" and he did not prove that syntax will never lead to semantic in any case. He was assuming what he was proving: - He assumes that "programs are entirely syntactical" and proved that "computer programs are entirely syntactical therefore not mind." Searle seemed not serious on his proof. He receded and gave up the intention of proving it in 2002, "Someone is bound to ask, can you prove that the computer is not conscious? The answer to this questions is: Of course not. I cannot prove that the computer is not conscious, any more than I can prove that the chair I am sitting on is not conscious." (Searle 2002).

With his Chinese Room argument, Searle tried to show that a computer may *look* intelligent, understanding, touching, and emotional, but it achieved them by following symbol processing rules, which are the senseless *simulations* of intelligence, understanding, touching, and emotions. Even though such a machine successfully passed Turing Test, it would not be assumed to be intelligent since it would have no understanding or no mind.

3. OUR E.T.'S CHINESE ROOM ARGUMENT

Searle argues by using his Chinese Room that the person in the room has no semantics about Chinese, therefore not intelligent, by simply following the syntactical rule book, even though his answers make perfect sense. Our questions to Searle's Chinese Room argument are:
    If the person in the room answered every question in Chinese perfectly, then why don't we think that he "understand" Chinese?
    Do we really care how a person does it in his brain when he *shows* that he knows Chinese?

We have developed an *E.T.'s Chinese Room* scenario to address the above questions: *E.T., an extraterrestrial stranded on Earth (a figure in movie "E.T.", directed by Steven Spielberg in 1982), instead of Searle, is locked in the Chinese Room. It answers the questions in Chinese passed to it in some way, following the rule book in Chinese Room or using some functions inside its brain. It answers the questions perfectly in the eyes of the people outside. No one knows how E.T. does it.*

Now what would people say about E.T.'s capability on Chinese? We don't think people would say that "we don't know whether E.T. understands Chinese or not, because we do not know how it does that." We don't think people would insist looking into E.T.'s brain to investigate how it does it in order to judge whether it was really doing it or was simulating. We don't think people would have problems admitting the fact that E.T. understands Chinese even they do not know how it does it. People would say, "Look, E.T. knows Chinese!" "E.T. is smart and learns fast!" There are geniuses or savants in this world who can quickly tell what day any given date is quickly. No one knows how they do it. But it does not prevent us to call their works intelligent.

Consider how to test the intelligence of E.T. It seems there is no method better than the Turing Test, as far as we do not want to kill and dissect it (it's likely that even we dissect E.T., we cannot understand what happens inside to generate E.T.'s intelligent behaviors, even though we may realize that E.T.'s intelligence processing mechanism is different from our biological mechanism). Why is the Turing Test good for testing a "robot" E.T. from somewhere deep in the universe but no good for a robot made by ourselves? The stumbling block in our minds is: we tend to take something mysterious and unknown as "intelligent"; once the unknown becomes known it is no longer intelligent. Thirty years ago, people thought that a machine that could do spelling checking and grammar check would be "intelligent". Now, millions of machines can do them, and people do not take them as intelligent functions, because people know that the spelling checking in the computer is nothing but searching on a huge database. We tend to use double standards. For mysterious E.T., we would like to say it is intelligent without insisting to know how it reaches that level of intelligence. But for a robot we make, we would demand its internal process must also be "intelligent" in addition to its behaviors.

With the above arguments on E.T.'s Chinese Room, we have to admit that the idea of the Turing Test is objective and reasonable, and it is an empirical and feasible method to test machine's intelligence, although the details of the test can arguably changed to adapt to different scenarios.

Among the issues concerning about the intelligence of a machine, such as "whether it acts smartly", "whether the way it thinks is smart", and "whether it thinks in the way same as us", "whether it acts smartly" is most important and interested by the people who are making intelligent machines. Airplanes made human dreams come to true to "fly" in the air like birds. Submarines made human dreams come to true to stay and "swim" deep in the ocean like fish. We have created calculators that calculate at a rate and accuracy no one ever dreamed of. In developing these machines, we did not demand the way machines did to be the same as birds, fish, and human brain. Our concerns were on the machines' "behaviors": moving in the air which is called "flying", moving in the water which is called "swimming", or finding results of a mathematical problem which is called "solving". As far as the machine can fly, swim, or compute, the people who made it would not be tied down to certain mechanism to realize the

designated behavior. If people had restricted themselves on developing machines that worked same way as birds, fish, and human brains, we would not have had airplanes, submarines, and calculators.

AI people are working on creating machines whose level of intelligence and consciousness is on a par with humans. The top priority is to make a machine acting as intelligently as a human. How the machine does it is not the top concern. No one has proved that the way our biological brains work is the only way to have intelligence. We thus should not take a particular intelligence process, such as our brain's process, as the standard for intelligence. There are currently alternative ways for automated reasoning with uncertainties, such as rule-based system, frames, neural network, Bayesian network, none of them are the way that exactly occurred in our brains when we are doing reasoning. In fact, we do not know yet what exactly occurs in our brains when we are thinking. For those who insist that machines are not intelligent unless they are "thinking" exactly in the way our biological brains do, they themselves do not know what they are talking about because no one is aware yet of what is the "way of our biological brains." Put it in another way, just because people are unaware of what exactly happens in our brain to come up with intelligence, they insist that only the unknown mysterious process is intelligent. People imitated birds to fly because they knew little about "flying" and thought the way of birds is the only way of flying. They eventually gave up the way of mimicking birds to fly after they knew more about the "flying" and aerodynamics: - There turned out to be alternative ways of flying!

The benchmark of intelligence should be set up on intelligence itself, rather than the way of generating intelligence. Some may dispute that intelligence is both a "process" and the "outcome" of the process. We do not have disagreement on it. But between the outcome and process, our first priority is on the outcome. That is, at this stage, we are more concerned to have machines whose "intelligent outcomes" are comparable to human "intelligent outcomes".   The Wright brothers might have been thinking in the same way when they were making their first plane hundred years ago: - Let's make a machine that could "fly" as the outcome, regardless the way of flying and the definition of "flying" (some insisted that flying is a process in which two wing flap up and down). Using machine's internal process to judge its intelligent level would empirically cause problems since we do not yet know much about human "internal process" of generating intelligence. It would be less ambiguous if judging by its external behaviors and actions.

4. CONCLUSION

For AI people, the top interest is in making a machine with its intelligence and consciousness level on a par with humans, with whatever approach to generate the intelligence, albeit the research on how our brain works is a topic for scientists in cognitive science, computer sciences including AI, biology, physics, and philosophy.   For the top concern of AI, the Turing Test is a proper method by using the external behaviors, rather than internal process, as the measurement to assess the achievements of AI. After people have better understandings of semantics, intelligence, consciousness, mind, and spirit, as well as the process of generating them, an approach of testing machine's mentality better than the Turing Test might emerge.

REFERENCES

[1] Bringsjord, S. (2008). If I were Judge. In Epstein, R., Roberts, G., and Beber, G. (Ed.). Parsing the Turing Test. Springer.

[2] Epstein, R., Roberts, G., and Beber, G. (Ed.). (2008). Parsing the Turing Test. Springer.

[3] Ford, K. M. and Hayes, P. J. (1995). Turing Test Considered Harmful. Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI-95). pp.972-977

[4] Haugeland, J. (Ed.) (1985). Artificial Intelligence: The Very Idea. MIT Press.

[5] Humphrys, M. (2008). How my program passed the Turing Test. In Epstein, R., Roberts, G., and Beber, G. (ed.). Parsing the Turing Test. Springer

[6] Luger, G. F. (Ed.) (1995). Computation and Intelligence: Collected Readings. AAAI Press.

[7] Luger. G. and Stubblefield, W. (1989). Artificial Intelligence and Design of Expert Systems. Benjamin / Cummings.

[8] Nilson, N. J. (2009). The Quest for Artificial Intelligence: A History of Ideas and Achievements. Cambridge University Press

[9] Russell, S. and Norvig, P. (2010). Artificial Intelligence, A Modern Approach. Prentice Hall, Third Ed.

[10] Searle, John R. (1997). The Mystery of Consciousness. The New York Review of Books, New York, NY

[11] Searle, J. R. (2002). I Married a Computer. In Richard J. (Ed.). Are we spiritual machine? Discovery Institute Press

[12] Shapiro, S. C. (Ed.) (1992). Encyclopedia of Artificial Intelligence (second edition). Wiley

[13] Shieber, S. (1994). Lessons from a Restricted Turing Test. Communications of the Association for Computing Machinery (CACM), 37, 70-78

[14] Shieber, S.(2004). The Turing Test. MIT Press

[15] Turing, A. (1950). Computing Machinery and Intelligence. Mind, Vol. 59, pp.433-466

COMMUNICATIONS TECHNOLOGY IN THE 2008 ELECTION AND BEYOND: AN ANALYSIS OF OBAMA EMAILS

Janet Prichard, Bryant University, (401) 232-6269, prichard@bryant.edu
Richard Holtzman, Bryant University, (401) 232-6097, rholtzma@bryant.edu

ABSTRACT

No political candidate has adopted and employed communications technology to connect with potential voters more effectively than Barack Obama in the 2008 presidential race. Yet, scholarly literature has neglected systematic studies of perhaps the most significant piece of Obama's communications revolution: the use of millions of targeted emails. No in-depth analyses of these emails have been undertaken. Our research aims to fill this important gap by systematically examining the targeted emails themselves in an effort to understand this phenomenon and its larger significance. We do so by charting their frequency and identifying distribution trends, as well as exploring their content.

Barack Obama; Campaigns; Communications Technology; Email Analysis; 2008 Election

INTRODUCTION

The state-of-the-art in campaign communications has developed rapidly over the past decade. More efficient and effective means of establishing and maintaining lines of communication with potential voters, such as robo-calls and candidate websites, have largely come to replace traditional, labor-intensive methods of contact, such as door-to-door campaigning. Reflecting this trend, scholarly research on the development, utility, and significance of this recent transformation of the campaign communications environment has emerged as a veritable cottage industry in political science.

No campaign has adopted and employed communications technology in its efforts to connect with potential voters more effectively than that of Barack Obama in the 2008 presidential race. His campaign's use of multiple websites, YouTube, Facebook, and Twitter changed the rules of the game for presidential candidates. And yet the political science literature has largely neglected systematic studies of what was perhaps the most significant piece of Obama's campaign communications revolution: the use of targeted emails. Throughout the primary and general campaigns, Obama's organization compiled a list of millions of email addresses and consistently bombarded recipients with information, pleas for financial contributions, links to pertinent websites and videos, opportunities to volunteer and organize, and reminders to vote. While the larger contours of this development have been broadly noted in recent research, no in-depth analyses of these emails themselves have been undertaken.

This research project aims to fill this important gap by exploring the Obama campaign's use of targeted emails by systematically examining the emails themselves—over 250 of them. To address the methodological challenges raised by such a study, which perhaps accounts for the dearth of research on this topic, we employ the qualitative data analysis tool MAXQDA. Our

purpose in doing so is to make sense of this targeted email phenomenon both at the macro level, by charting their frequency and identifying distribution trends, and at the micro level, by exploring their qualitative content and interpreting its significance. This paper represents the beginning of this research project and therefore its focus is largely limited to the macro analysis. It is our first shallow dip into a very rich, deep data set consisting of hundreds of emails distributed by the Obama campaign and his administration over a two year period.

We engage this material by first reviewing relevant studies of recent developments in campaign communications technology and identifying the contributions, as well as the gaps, in this literature. Second, drawing upon national media coverage, we discuss the Obama campaign's use of targeted emails and their effectiveness. This is followed by an introduction of our methodological considerations, including the use of the MAXQDA qualitative data analysis tool. We then offer our analysis of a significant sample set of the emails distributed by the campaign, as well as during the presidential transition and first 19 months of the Obama administration. The paper concludes with a discussion of the significance of this phenomenon, both in terms of Obama's election and its larger significance for the state of modern campaign communications.

THE CHANGING LANDSCAPE OF CAMPAIGN COMMUNICATIONS

Communicating with potential voters is, of course, the focal point of any political campaign. In an effort to disseminate a presidential candidate's message effectively, campaigns have become increasingly professionalized in recent years. Image consultants, pollsters, focus groups, and marketing have become the centerpiece of the modern campaign. With the aid of massive databases of information on potential voters, it has become possible to strategically craft and target messages to particular audiences through telephone solicitations, direct mail, advertisements on radio, local and cable television, and more recently, the Internet [20].

As with the developments in communications technology that preceded it, the emergence of the Internet as a means of targeting voters has fundamentally changed the landscape of political campaigns. In her essay on the ethics of contemporary campaigning, Carol A. Whitney observes that "opportunities for delivering campaign messages, coordinating staff and volunteer activities, and raising funds have ballooned through the increasingly sophisticated use of the Internet" [21]. And this has occurred during a relatively short period of time. Senator Dianne Feinstein (D-CA) was the first candidate to build a campaign website in 1994, when less than 14 percent of Americans were Internet users [10]. And yet, by the 1996 election, all major presidential candidates had developed a website, as did a third of all congressional candidates. By the 2000 election, nearly every major candidate had a website, as did political parties, interest groups, and a wide variety of political organizations [6].

During the 2004 Democratic presidential primary race, Howard Dean's website represented the centerpiece of his campaign strategy, with blogs that encouraged posts from visitors, a "meet up" tool that facilitated the informal gathering of supporters, and a simple way to donate money online. After John Kerry emerged as the Democratic nominee, his campaign adopted Dean's tools for his own website. George W. Bush's reelection campaign likewise used a website to organize grassroots momentum by providing supporters with talking points and information on how to contact their neighbors and the local media [6]. Four years later, Obama's campaign

demonstrated that it had learned the lessons of its predecessors and, in turn, was able to capitalize on online communication technology in an astounding way. As Dean's former campaign manager, Joe Trippi, put it: The Dean campaign "turned out to be nothing but the Wright brothers: We proved that you could fly." By contrast, "Obama really is landing a man on the moon" (Quoted in Brownstein [2]).

Obama's strategists recognized that, beyond using a campaign website, online communications technology offered a wide variety of other means to reach a large number of potential voters. According to a report produced by Aaron Smith for the Pew Internet & American Life Project, 55 percent of the voting-age population used the Internet to connect to the electoral process during the 2008 cycle, which represents three out of every four Internet users [18]. Along with blogs, "meet up" tools, and simplified means of fundraising and recruiting volunteers, the Obama campaign sought to reach Americans through "viral marketing" efforts that inundated social media sites such as YouTube with videos, established an unavoidable presence on social networking sites such as Facebook and Twitter, and embraced the use of email to an extent that had been unthinkable just four years prior.

Understandably, the new campaign communications landscape precipitated by the development and increasingly sophisticated use of the Internet has become an area of focus in the study of modern campaigns and elections [1] [3] [5] [6] [15] [20]. And yet, few contributions to the large body of literature examining campaigns and elections investigate the use of emails as their primary object of study (for examples, see Doherty [4]; Williams and Trammell [19]). As Joseph Graf states in his discussion of new media and its impact on campaign communications, emails have "become part of every political campaign, especially because it may be the cheapest means of reaching potential voters,….[yet] little is known about the effects of political e-mail on its recipients" [6, p. 56]. We argue that before the effects on potential voters can be understood, the phenomenon of Obama's targeted emailing campaign must be explored. This paper is a first step in that exploration and an effort to better understand the larger significance of this phenomenon in modern political campaigns. In this initial analysis, we examine the emails themselves as our primary object of study, with an aim toward identifying and conceptualizing the basic contours of this phenomenon.

THE USE OF TARGETED EMAILS BY THE OBAMA CAMPAIGN

Writing in the *National Journal*, Ronald Brownstein argued that ability of Obama's efforts to communicate "directly with previously unimaginable numbers of voters" represented the "first true 21st-century campaign" [2]. He quotes Peter Leyden, director of the New Politics Institute, a Democratic group that studies campaign tactics and technology: "If I had to boil down what has really happened in the election cycle, it is [that] you are finally seeing the real fruition of the full power of...the Internet on politics" (quoted in Brownstein [2]). The Obama campaign's opportunity to maximize their advantage in this quickly-developing communications technology arena was provided by the strong embrace of this technology by the candidate's supporters. For example, while more supporters of Republican John McCain's candidacy identified themselves as Internet users (83 percent), compared to 76 percent of Obama supporters, the latter was clearly the more engaged group of the two. According to the 2009 Pew Report [18], 18 percent of Obama supporters signed-up online for election alerts through the candidate's website, compared

to 9 percent of McCain supporters. A similar lopsidedness was reported in regards to how many supporters went online to contribute money (15 percent for Obama to McCain's 6 percent), volunteer (11 percent for Obama, 4 percent for McCain), set-up custom news alerts (12 percent Obama, 8 percent McCain), and share multimedia content with others (21 percent Obama, 16 percent McCain).

However, the communication medium in which the Obama campaign truly outpaced that of his Republican rival was the targeted email. It was not the first organization to utilize this communications technology on a mass scale—Voter Contact Services and Voter Emailing Company, two private firms financed by the Republican Party, built a database of 24.8 million email addresses during the 2004 election cycle. This database was primarily employed in the party's "get-out-the-vote" effort in support of Bush's reelection. This number is in fact significantly larger than the estimated 13 million email addresses collected by the Obama campaign (the actual number has not been disclosed). However, a significant difference that set Obama's 2008 effort apart from those of the Republican organizations four years prior was that the contact information gathered by the Obama campaign was not purchased from marketing database firms—it was voluntarily provided by individuals who visited the campaign website, as shown in Figure 1.

FIGURE 1 First page of Obama Campaign website

To paraphrase an old adage in the advertising business, half of all money spent on political campaigns is wasted, but no one knows which half. And while the amount of money spent on campaigns has exploded to record numbers, financing is still finite and strategists must still make cost/benefit judgments. Emails are a particularly attractive means to communicate with potential voters because they are incredibly cheap compared to other targeting methods, such as television advertisements. While they risk being automatically filtered as spam, accidentally or purposefully deleted by recipients, or simply left unread, the cost/benefit analysis still heavily works in their favor. Emails are also versatile, which is considered value-added. As developed by the Obama campaign, they offered a simple way to solicit donations, provide voters with bits of timely information on a daily basis, recruit volunteers, link to videos and other websites, and

help supporters navigate the voting process. As a result, according to Michael Scherer, three million Americans made financial contributions to the campaign, often in small increments; two million created profiles on Obama's social-networking site, MyBarackObama.com; and more than 1.2 million volunteered for the campaign [16].

The Obama campaign regularly inundated supporters with emails throughout the primary and general election races. According to the 2009 Pew Report, 48 percent of all Obama voters received regular emails directly from the Democratic Party and/or the campaign; compared with 38 percent for McCain. Additionally, the content of these emails was often targeted to the particular region, state, and city of the recipient, which allowed the campaign to not only provide information on policy issues that were locally-relevant, but to also inform potential voters of campaign events occurring in their area. However, the true effectiveness of these emails was not necessarily located in the information provided, but rather in the way that these messages spread virally from their original recipients. Easily forwarded to others, a survey of Obama voters indicates that these emails were more typically used for sharing and communication rather than information gathering. An incredible 59 percent of polled email users reported exchanging emails with others concerning the campaign; including 17 percent who exchanged emails about the race daily! As Brownstein [2] observed, "the expanding capacity of the supporters themselves to communicate with large numbers of like-minded people and to pursue independent work for their candidate, undirected by a campaign," which has been likened to "the chain letters we received as kids" [8], constitutes a new and significant trend in campaign communications. The genius of this email strategy is that it represents "mass communication without the mass media" [2]. According to Micah L. Sifry and Andrew Rasiej, of the Personal Democracy Forum, which studies and reports on the effect of technological innovation in politics, this represents a paradigm shift from "sound-bite" politics controlled by mass media, to "sound-blast" politics that can be shaped by the campaigns themselves [17].

DATA COLLECTION AND ANALYSIS

The collection of data for this project began as an accident of sorts. In late June 2008, one of the authors submitted a question to the Obama campaign through its website. In order to submit this question, further information was requested including a name, an email address, and home address. In the text accompanying the question, the questioner asked that no automated response be sent and that the email address provided not be placed on a mailing list. Almost immediately, a general response email indicating receipt of the request was received and two days later the barrage of campaign messages began. During the month of July 2008, fifteen messages were received, and twenty-two more were received in August 2008. When the messages began coming in, a conscious decision to save all of the emails began. At the time of this writing, 342 emails have been collected.

These emails were first forwarded to a gmail account and then, in order to extract them for further analysis, they were all labeled as ObamaEmails, which allowed them to be copied to an ObamaEmail folder on a local computer using Mozilla Thunderbird. Thunderbird uses the mbox format for folders, storing all of the emails in a single file. Using a free application called IMAPSize, the command mbox2eml converts the single mbox file into individual files know as

eml files. For the textual analysis of emails using MAXQDA, the email files were converted to rtf files using an application called EML to PST.

For this study, we decided to look at emails over a two-year period, beginning on September 1, 2008, immediately following Obama's official nomination as the Democratic Party's presidential candidate, through August 31, 2010. This constitutes a data set of 258 messages. An initial investigation of these emails began with an analysis of basic identifying components—sender, dates, subject, and message. The results were as follows.

Sender
There were a total of twenty-nine different sender names on the 258 messages. In some instances, a single person may have sent out mail with two different sender names. For example, over the two year period, messages from Barack Obama were generally sent out as "From: Barack Obama." But from February 2, 2009 to March 21, 2010, messages were from "From: President Barack Obama." Other senders whose sender names were consolidated include Vice President Joe Biden, Michelle Obama, and campaign manager David Plouffe. Also, there were thirteen senders that only sent out a single email; this list includes the late Senator Ted Kennedy, former Vice President Al Gore, Senator John Kerry, and state level representatives (from Rhode Island). For the purposes of this analysis, these thirteen one-time senders are categorized as "other." Figure 2 represents the distribution of emails from the various senders.
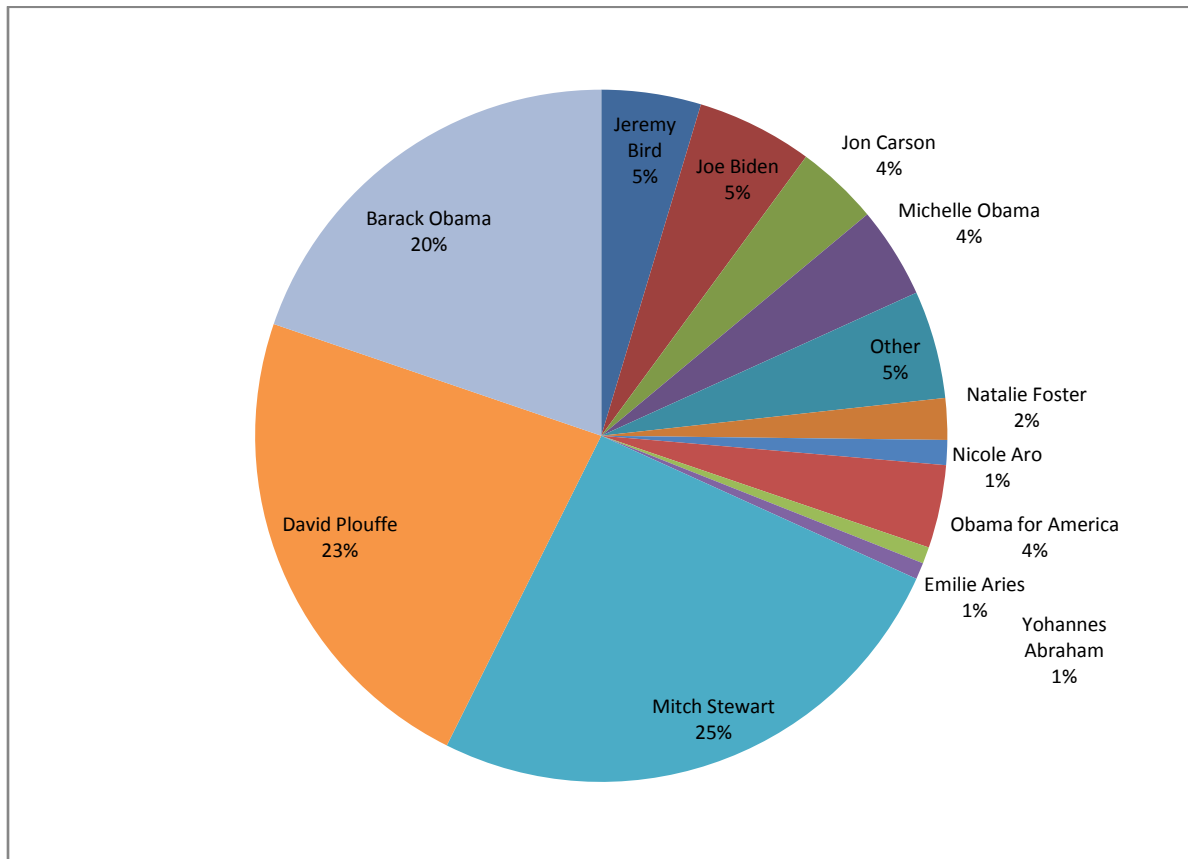
FIGURE 2 Frequency of emails by sender

Dates

The frequency of the emails month-by-month was then examined with the results shown in Figure 3.
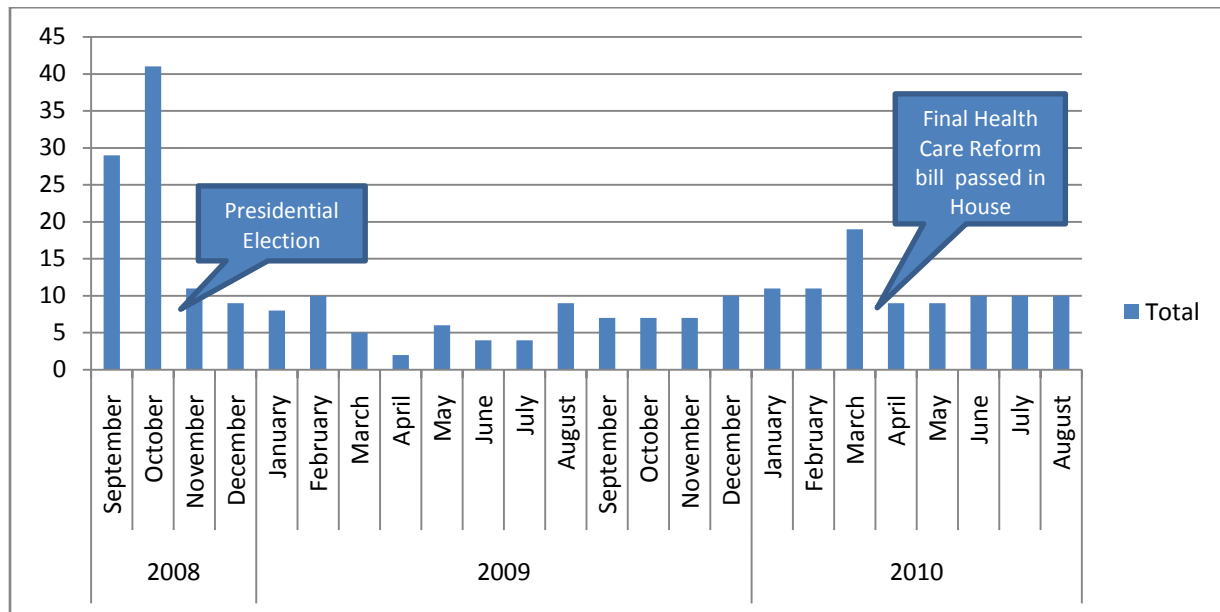
FIGURE 3 Frequency of emails by month

Subject

The subject lines of the 258 messages were examined for keyword frequency. The more interesting results are summarized in Figure 4:

Location Sensitive	
Rhode Island	10
RI	1
Barrington, RI	5
Providence, RI	3
Sen. Whitehouse	4
Rep. Langevin	1
Massachusetts	1
Warwick	1
Rep. Kennedy	4
Total	30

Mention Obama	
Barack	14
Obama	5
President Obama	5
Barack Obama	1
President	9
Total	34

Health Care	
health reform	6
health care	3
health care reform	1
Total	10

Miscellaneous	
video	14
Person's name	12
debate	9
vote	8
help	7

Time Sensitive	
midnight deadline	6
other deadlines	4
urgent	5
Total	15

FIGURE 4 Keyword frequency in subject lines

Note that thirty of the messages (11.6%) are targeted by location—through the mention of the state (Rhode Island), city of residence (Barrington), state capital (Providence), or Rhode Island's U.S. Senators or Representatives (Whitehouse, Langevin, and Kennedy). Also, twelve of the messages used the recipient's first name (Janet), further personalizing the message.

Messages

Using MAXQDA, some preliminary analysis was done on topics contained in the email messages themselves. Some of the topics searched were based on phrases such as "Health care" and "Wall Street," while others were grouped by related terms, "War related" includes "war," "Iraq," "Afghanistan," and "troops." Both the number of hits on each topic and the number of documents involved are included in the graph shown in Figure 5.
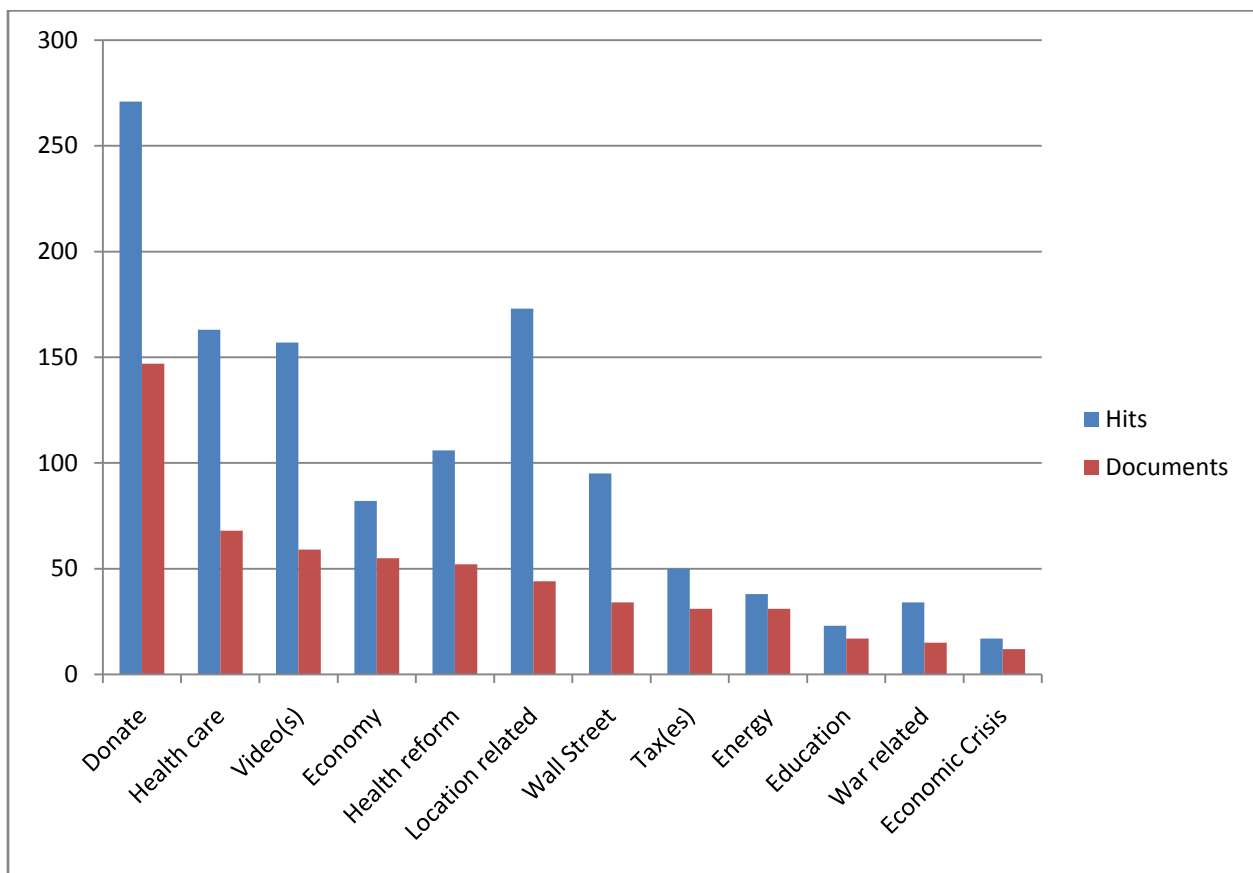
FIGURE 5 Number of emails by topic

Topics that had two hits or less included "environment," "poverty," "immigration," "Social Security," and "stem cell." Terms that had no hits located in this data set included "Iran," "Israel," "abortion," "gun(s)," "Cuba," and "free trade."

CONCLUSIONS

The massive employment of targeted emails by the Obama campaign during the 2008 presidential race represents a significant development in the state-of-the-art in campaign communications. This development provided a means for the campaign to not only reach large numbers of potential voters but, more importantly, to keep them engaged in the process through an on-going electronic dialogue. Additionally, the low-cost and versatility of these emails, and the ability to localize their messages, seems to have offered a far more effective and efficient means of communication than television advertisements, the industry standard. However, perhaps the most significant development offered by the use of targeted emails is their viral nature, which effectively transformed supporters who forward these messages friends and colleagues into unofficial campaign volunteers.

Our preliminary analysis of these emails demonstrates the ways in which the Obama Campaign and his administration have utilized this communications strategy to speak directly to potential voters and constituents. In the modern era of the personalized presidency, the import of establishing and maintaining such a relationship with the American people cannot be understated. The frequency of these emails alone suggests the image of intimacy and personalization that accompanies this approach. Additionally, the ability of the campaign to raise money in small amounts, to an extent that is without comparison, was undoubtedly aided by the use of emails. Finally, not only were these emails utilized to engage potential voters and constituents with national issues, which were prominently discussed in these communications, but local issues were also targeted as well.

Continuing work on this project will undertake a more extensive micro analysis, which will look more closely at the messages themselves—in terms of form, content, and character—across a number of recipients. Email accounts for persons living in various parts of the country have been setup to receive email in order to deepen, as well as diversify, our data pool. Part of the analysis to follow will also include an examination of the consistency or variability of email messages in different areas of the country and efforts to determine if embedded URLs are encoded for tracking purposes. Additionally, the significance of the unique senders of these messages will be explored and correlated.

REFERENCES

[1] Baumgartner, Jody C. & Francia, Peter L. *Misconceptions about the E-Campaign: What the Internet Can and Cannot Do for Political Campaigns.* Conventional Wisdom and American Elections, 2nd ed. Lanham, MD: Rowman and Littlefield, 2010.

[2] Brownstein, Ronald. *The First 21st Century Campaign*. National Journal, April 19, 2008.

[3] Cornfield, Michael. *Game-Changers: New Technology and the 2008 Presidential Election*. The Year of Obama, edited by Larry J. Sabato. New York: Longman, 2010.

[4]   Doherty, Brendan. *Emailing for Dollars: Presidential Hopefuls, Online Strategy, and Fundraising Success in the 2003 Invisible Primary*. Midwest Political Science Association Annual Meeting, Chicago, 2004.

[5]   Germany, Julie Barko. *The Online Revolution*. Campaigning for President 2008: Strategies and Tactics, New Voices and Techniques, edited by Dennis W. Johnson. New York: Routledge, 2009.

[6]   Graf, Joseph. *New Media—The Cutting Edge of Campaign Communications*. Campaigns on the Cutting Edge, edited by Richard J. Semiatin. Washington, DC: CQ Press, 48-68, 2008.

[7]   Gulati, Girish, & Williams, Christine. *Closing the Gap, Raising the Bar: Candidate Website Communication in the 2006 Campaigns for Congress*. Social Science Computer Review, 24(4): 443-465, 2007.

[8]   Hirschorn, Michael. *Only Connect*. Atlantic Monthly, May, 2008.

[9]   Ives, Bill. *How Barack Obama is Using the Web to Further Engage Voters*. EContent 31(5), 2008.

[10] Madden, Mary. *America's Online Pursuits: The Changing Picture of Who's Online and What They Do*. Pew Internet and American Life Project, 2003. Retrieved January 12, 2011 from http://www.pewinternet.org/~/media//Files/Reports/2003/PIP_Online_Pursuits_Final.PDF.PDF

[11] Magleby, David B. *How Barack Obama Changed Presidential Campaigns*. Obama: Year One, New York: Longman, 2010.

[12] *Obama Presidency Will Be 'Wired'*. Information Management Journal, Jan/Feb 2009. Retrieved January 18, 2011 from http://findarticles.com/p/articles/mi_qa3937/is_200901/ai_n31426799/

[13] *Obama's Standing Army*. Editorial in The Economist. March 21, 2009. Retrieved January 18, 2001 from http://www.economist.com/node/13331151/comments

[14] Owen, John. *Obama's Election Success Represents a Landmark Victory for Digital Marketing*. Revolution, December, 2008.

[15] Rosenblatt, Alan. *Dimensions of Campaigns in the Age of Digital Networks*. Campaigns and Elections American Style, 3rd ed., edited by Thurber and Nelson. Boulder, CO: Westview Press; pp.207-226, 2010.

[16] Scherer, Michael. *The Permanent Campaign*. Time, January 26, 2009.

[17] Sifry, Micah L., & Rasiej, Andrew. *Welcome to the Age of the Sound Blast*. Politico.com, March 26, 2008. Retrieved December 1, 2010 from http://dyn.politico.com/members/forums/thread.cfm?catid=1&subcatid=4&threadid=545063

[18] Smith, Aaron. *The Internet's Role in Campaign 2008*. Pew Internet and American Life Project, 2009. Retrieved January 4, 2011 from http://www.pewInternet.org/Reports/2009/6--The-Internets-Role-in-Campaign-2008.aspx

[19] Williams, Andrew Paul, & Trammell, Kaye D. *Candidate Campaign E-Mail Messages in the Presidential Election 2004*. The American Behavioral Scientist, 49(4): 560-574, 2005.

[20] Wayne, Stephen J. *Campaign Communications: How Much Do They Matter?* Is This Any Way to Run a Democratic Election?, 4th ed., by Wayne, Washington, D.C.: CQ Press, 2010.

[21] Whitney, Carol A. *Campaign Ethics in a Changing World*. Campaigns and Elections American Style, 3rd ed., edited by Thurber and Nelson. Boulder, CO: Westview Press; pp.251-268, 2010.

Knowledge Process Outsourcing (KPO): Defining Characteristics, Ambiguous Flaws, and Future Trends

Christina Royer-Scheible, BNY – Mellon Bank, Philadelphia, PA
Snehamay Banerjee[*,] School of Business' Rutgers University, Camden, NJ 08102
Chintan Chhaya, School of Business, Rutgers University, Camden, NJ 08102

(ABSTRACT)

In the era of globalization, outsourcing and instant communication knowledge process outsourcing (KPO) is old news to the business world. KPO, a natural extension of business process outsourcing (BPO) has been studied by some researchers but not as extensively as BPO literature. While there are many similar definitions of KPO, these can be generally summarized as: *the transfer to a third party of complex and highly variable activities that require intellectual decision making.* (Michell 2005). While cost advantage is the obvious benefit of KPO, other benefits may include (depending on the type of KPO) reducing time-to-market for product development or services, managing risks of obsolescence and reducing the need to upgrade employee skills in areas such as technology and medicine. KPO has its potential risks too that include issues such as intellectual property, confidentiality, regulatory, geopolitical, reputation etc., (Currie et al., 2007).

Despite the potential risks and occasional political resistance KPO business is growing and likely to grow in future.   KPO activities under financial analysis and equity represents 60% of the KPO market, market research and business intelligence represents 20% of the market, Legal process outsourcing involving litigation review, IP protection and patent research represents 10% of the market, whereas pharmaceutical industry and others represent the balance 10% of the market (OutsourcePortfolio.com). However, the growth in KPO revenue is not meeting its projected expectation. It was projected that KPO revenue will exceed $17 billion in 2010 (Evalueserve Report 2004) and later NASSCOM predicted KPO business in India only will reach $15.5 billion in 2010 (Sathe, 2006). However, now KPO revenue for 2010-2011 is expected to reach $11.2 billion.

In this study we posit that KPO has moved from the phase of local exploitation, a stage where routine use of explicit knowledge by cheaper knowledge workers for cost advantage, to a more mature phase where the relationship between KPO providers and their partners, and the impact of this phenomenon on employees and corporate culture need to be evaluated for developing a realistic growth pattern for KPO.

Research Issues:

While most research conclude that cost savings is the driving force behind continuation and growth of KPO and that it has moved to a long term partnership and development of common goal between the parties involved. Some have presented KPO as a tool for creating national

economic competitiveness (Snieska and Draksaite, 2007). We analyze the relationship between and a firm and international partner based on the value of task involved (a higher value task requires a more skilled worker/professional with higher level of knowledge/expertise) and the level of control desired by the partner originating the relationship (i.e. for outsourcing, it will be the organization receiving the service) as presented below:

Control		
	Loosely Coupled	Tightly Coupled
Task Value — Low Value	BPO	Unlikely combination
Task Value — High Value	KPO	Subsidiary

Relationship Matrix with Offshore Partners

The extent and nature and extent of control between the partners will be moderated by factors such as maturity of the product or service, maturity of the relationship between trading partners and the nature of corporation receiving the service (such as domestic Vs multinational corporation). While the existing literature does not discuss such potential moderators and predictors for relationship between KPO providers and their clients, we show that such relationships do exist based on our field research.

Please contact the corresponding author for a full version of the paper and references.

ACCESSING THE IMPLICATION OF PROCESS COMPLEXITY ON IS AUDIT SERVQUAL: AN ENTERPRISE SYSTEM PERSPECTIVE

Joseph Nwankpa
Dept. of Management & Information System,
Kent State University,
Kent, Ohio
jnwankpa@kent.edu

Yaman Roumani
Dept. of Management & Information System,
Kent State University,
Kent, Ohio,
yroumani@kent.edu

ABSTRACT

The central theme of this paper is to examine the impact of Business process complexity on IS audit quality using the construct of SERVQUAL. From the IS auditors perception of an organizations ability to deploy appropriate compliance measures in an enterprise system, this paper contends that as business process complexity increases in an enterprise system environment the IS audit service quality will witness a corresponding decrease as IS auditors will grapple with understanding the business process hence lowering their expectations on audit reliability, assurance, empathy, and responsiveness.
Keywords: Business Process Complexity, Enterprise System, IS Audit, SERVQUAL.

INTRODUCTION

In the last century, the global business environment has witnessed the arrival of new sophisticated technological infrastructure aimed at alleviating risk, streamlining business processes and enhancing operational efficiency. Firms on their part have embraced these technologies in such a manner that they more than ever before rely on large-scale systems to coordinate complex business processes and sustain competitive advantages. Most renowned among them are enterprise systems, which are industry specific customizable software packages that integrate information and business processes in an organization [12, 29]. The use of technology has evolved from a data centric era to a process driven initiative where technologies are embedded with "best practice" process considerations. For example, enterprise systems have heralded an era where processes embedded in these technologies are increasingly becoming the panacea for efficiency as vendors surged to develop 'best practice' solutions for company's information processing needs. Nowadays, the search for optimal business processes ends up with enterprise systems in the form of packaged implementation, modular implementation or some degree of customization with existing in-house systems.

The consequences of these evolutions are seen in the role of information system auditors, which have subsequently evolved from complementary to mandatory service in an audit engagement. IS auditors are now critical members of the audit team and their expertise and knowledge are required more than ever before. Hence, as organizations continue to deploy robust and complex IT infrastructure, it becomes inevitable that 21st century auditors need to have the sufficient grasp of business processes to be able to make informed judgments and decisions critical to audit quality. Such thorough understanding of client's business processes help auditors assess potential business and financial risk and help evaluate control adequacy [5].

IS auditors' audit reports often serve as a proxy for an organizations effective use of enterprise systems and its alignment with business processes. Enterprise systems aim to mitigate risks by reducing complexity; complexity reduction is mainly at the system level rather than capturing evolving complexities at process and workflow levels. David compares enterprise systems to monolithic templates that are more self-serving to the system instead of being agile to changing process and governance complexities [10]. Simply put, due to time or cost considerations, companies often fail to incorporate and align actual business process complexities with their enterprise systems. Over time, these complexities devolve as a result of changes in processes and software configuration resulting to a declining state of complexity where the result of the system may no longer be reliable. Therein lies the problem as often, enterprise systems fail to capture the realities of process complexities, creating a disconnect and therefore, an issue of concern for IS auditors. With the overwhelming need for collaboration across distributed workflow and with tightly coupled complex linkages, complexity is increasingly becoming a natural part of business growth. By encapsulating myriad process complexities, enterprise systems offer IS auditors a constant challenge of assuring quality IS audit. Unless organizations can either control business process complexities or periodically customize their enterprise system software, IS audits will become onerous engagements for IS auditors who will find it more difficult in providing objective assessments and service quality. This issue begs the following research question: In auditing enterprise systems, how do IS auditors perceive the downstream effects of business process complexities on IS audit service quality?

This paper attempts to use SERVQUAL matrix to investigate IS audit quality. We seek to find out how business process complexity will influence IS audit service quality. Therefore, it must be noted that IS audit SERVQUAL is not the organization's (client) perception of the IS audit but the IS auditor's perception of the capability of the organization in deploying appropriate measures for maintaining compliance with its Enterprise Systems. Our paper extends the IS audit quality literature by examining a critical determinant of audit quality which has been largely ignored. Finally, we offer propositions regarding the consequence of business process complexity on each service quality construct and how it affects IS audit quality.

We begin this paper with a discussion of process complexity literature followed by a review of the enterprise system literature, including a brief review of IS audit service quality. We then introduce the IS audit service quality constructs and a theoretical framework, on which we propose propositions and a research agenda. The paper concludes with a discussion of its contributions to IS research.

THEORETICAL BACKGROUND AND CONCEPTUAL DEVELOPMENT

Process Complexity

Process complexity is defined as the degree to which a process is difficult to examine, understand or explain [4]. Although there remains an ongoing debate on definitional aspects of process complexity [2, 14], certain commonalities are traceable. A business process has been determined to be complex in two ways either 'by having a large number of steps, or by having a complex set of percentage routings and feedback loops' [13]. The nature of these complex processes tends to be hierarchical, cross-functional and usually associated with random changes thus creating very dynamic systems [13]. This was consistent with Simon found that complex systems are composed of interrelated sub-processes that are often associated with large corporations with complex and cross functional business processes[32]. Indeed these complexities emerge from the size of the system, the interrelationship of the system components and unpredictable behavior of individual system components [33]. Typically, the degree of process complexity is dynamic as it changes over time as organizations continue to align with changing business environment. It was determined that the greater the likelihood for vertical decomposition of a process the higher the level of process complexity [13]. Cardoso [4] went further to examine the complexity in business process from the control-flow perspective.

The paper identified four main complexity perspectives, which are activity complexity, control-flow complexity, data-flow complexity and resource complexity. According to Cardoso [4] activity complexity viewed complexity based on the number of activities within a process while control-flow complexity looks at the constructs such as loops, splits and joins. Data-flow complexity measures the number of formal parameters of activities and the mapping between activities of data and finally resource complexity, which measures the amount of resources used up in a process. The paper concluded that complexity increases over time as more processes are been added to accommodate changes in the business environment.

Building on definitions offered by existing literature, this paper defines business process complexity as a state of business comprising of multiple loosely and tightly-coupled functions and processes within and outside the business organization exhibiting dynamic and random interdependencies and cross-functional activities. The interdependencies and cross-functional activities are such that uncoupling them will lead to unforeseeable random outcomes.

While increasing business processes are growing realities in firms as they strive to adjust to unpredictable business challenges such increases in processes came with baggage of problems such as error detection. Hence, complexities not only create barriers to analyzing process effectiveness but could also obscure positive organizational reforms. [2]. Process complexity could propagate tremendous bottleneck resulting to limited understanding, leading to more errors and defects [4]. Moreover, Fathee et al [13] argued that process complexity tends to reduce the possibility of identifying risks associated with reengineering. As processes get more complex they tend to obscure the overall business process conversely making effective coordination difficult.

Enterprise Systems

Enterprise system refers to industry specific customizable software packages that integrate information and business processes in an organization [29, 12]. Firms implement enterprise systems to support business processes such as customer relationship management, supply chain

management and to integrate disparate business processes within the functional areas of an organization [24]. Organizations view enterprise systems as a solution for fragmented information, incompatible legacy systems and inefficient processes. However, these systems are fundamentally different from other information systems due to high complexities, risks, scope and the investment requirement [12]. Enterprise systems as associated with huge changes in the organization existing business processes as the "best practice" embedded in them are often inconsistent with an organization existing practices [1] hence creating the need for some customization.

Enterprise system supports the process-oriented view of an enterprise because of its ability to streamline information flow across traditional business functions using a common database. Hence, as business processes of hitherto disparate functional areas become more interrelated and integrated, the systems and processes that support the organization grow correspondingly more complex. Enveloping complex business processes, enterprise systems are examples of complex systems. The complexity of enterprise systems can be found on the lines of codes consisting of conditional branching and hierarchical interaction of objects used to manipulate information into logical steps [28]. Enterprise system vendors calm that complexity is removed with the integration capability of the enterprise system. However, Rettig [28] argues that, in reality, replacing legacy systems with sets of interconnected modules to run the organizational functions creates its own complexities.

As today's organizations continue to grapple with the task of adapting to the dynamic changes in the environment in order to retain the competitive advantage, it becomes imperative that firms need to constantly align and readjust business processes across multiple chains within the value chain. Processes are constantly redesigned and readjusted in response to customers changing demand for better products and services [16]. Many organizations that have implemented enterprise systems in a bid to achieve the so called 'best practice' business processes and services as in constant touch with enterprise vendors for continuous reengineering- engaging in an unending cycle of process reengineering in tandem with growing process complexities. Hence a post implementation evaluation of an enterprise system will show a system that has devolved into a 'black box' over time.

THEORETICAL FRAMEWORK

IS Audit Quality

Audit quality has always been a source of concern to various end-user of financial information as users seek assurance that audited information is free of material mismatch and error. Audit quality is defined as the probability that an audit will discover and report material mismatch or breach in a financial statement [11]. Several studies have tried to identify key determinants of audit quality such as auditor's independence, auditors reputation, client size, and audit fee. However, the results have been mixed and at best inconclusive. For example, while some researchers argue that large audit fees paid to auditors may increase the effort by auditors, hence improving audit quality [19], on the other hand the opposing school argues that large fees paid to auditors could make auditors more economically dependent on these clients leading to scenarios

where auditors are reluctant to make appropriate investigation for the fear of losing clients [3]. Moreover, such economic bonding could lead to "Same As Last Year" audit approach: a situation where auditors are merely repeating what they did last rather examining for potential new issues and concerns. In a bid to ensure audit quality, the Sarbanes-Oxley Act included the mandatory rotation of lead and reviewing audit partners after five consecutive years on an engagement while regulations are formulated to ensure independence. However, with the current trend toward process integration and technology adoption, the auditing landscape has inevitably shifted to IS auditors in determining the overall audit quality.

IS audit is a subset of the overall audit process which is aimed at determining the state of affairs of an organization and at facilitating good corporate governance [31]. According to Ron Weber, IS audit refers to the process of collection and evaluating evidence to make a determination as to whether the information system safeguards assets maintains data integrity and achieve organizations goals and objectives. Sanaya [31] identified the major elements of IS audit as physical review, system administrator review, application software review, network security review, business continuity review and data integrity review. However, in order for IS auditors to improve service to clients, they need to understand the business process and unveil any potential complexity. The question becomes how can we ensure that the quality of IS audits in not compromised or obscured by processes that have metamorphosed into black boxes in organizations.

Research Framework (Figure 1)

Servqual

SERVQUAL refers to a five dimensional construct used by customers to measure service quality [26]. The dimensions include reliability, assurance, tangibles, empathy and responsiveness. While SERVQUAL has generally been used to capture consumer perceptions of service, SERVQUAL is equally appropriate for gauging IS audits. Organizations implementing enterprise systems use enterprise systems to provide a portfolio of services including information control and compliance based on certain accepted standards (e.g. EU 8th Directive, COSO internal controls). The role of the IS auditor is to review, verify, and assure controls, compliance, and risk management. Because the IS audit is a third party engagement, the quality of the IS audit,

therefore, rests on how well the IS auditor feels that the organization has capably managed its enterprise systems. Therefore, it must be noted that IS audit SERVQUAL is not the organization's (client) perception of the IS audit but the IS auditor's perception of the capability of the organization in deploying appropriate measures for maintaining compliance with its Enterprise Systems. The assumption is that IS audit service quality is an implied function of the preparedness of the organization implementing enterprise systems to produce and provide information services.

Reliability

Reliability in the context of IS audit refers to the ability of IS audit report to be dependable and accurate. This means that IS audit report should conform to standards and regulations while reflecting the true state of the system. Conforming to standards with regards to enterprise systems means that IS auditors need to examine the business processes and procedures, system functionality, application security and data conversion and integrity [20, 30]. IS audit should indicate to what degree the system is able to safeguard organizations assets and prevent fraud. Hence audit reliability is closely associated with the evidence gathered by auditors. The quality of audit evidence is very important in audit as auditors are less likely to make error in judgment with high quality evidence. Typically auditors gather evidence through test of control, test of detail, analytical procedures etc. However, it is the task of IS auditors to understand of the system and business processes in order to indentify and extract high quality evidence hence improving the overall audit quality. For example an audit examines a business process from origin to its disposition and what to know how assess control and object allocation are treated. The challenge is to identify whether proper segregation of duties are contained in these processes. The problem is that as business processes become more integrated and complex, auditor's ability to isolate evidence becomes even more blurred. In fact, IS auditors would spend less time to complete reliability evaluation of system control with a well documented and designed system [8]. Yet many organization have documentation that do not reflect the current status of the system while many organizations have business process so tailored and customized to their needs that it becomes very difficult for an outside expert to comprehend. Moreover, as the degree of process complexity increases, the time and the number of expertise required will inevitably increase. With limited time on audit engagement and pressures from audit team and client, IS auditors are faced with situations where assessments are made with limited information and time. In addition, while IS auditors are expected understand the business process level controls designed to address business risks and risk of misstatement, however, these controls are not generic rather are based on management policies. As the complexity of a system increases due to interrelated sub-processes and third party integration, the effective allocation of internal control become very difficult to identify and assess by IS auditors. Consequently the result will affect the reliability of the IS audit report as the IS auditor will require more time, expertise and other resources if they are to understand the control mechanisms and business processes to make an inform judgment. With these increased process complexity the level of understanding of the potential business and control risks becomes more obscured and complicated. IS auditors are now faced with the dilemma of trying decomposing the system to an acceptable level of abstraction to gain adequate insight. Such process complexity would inhibit IS auditors information processing and system understanding hence limiting the level of reliance the IS auditor will place on the system. Hence we propose:

Proposition 1: An IS auditor will perceive a lower sense of reliability when auditing Enterprise Systems in firms with high business complexity than for firms with low business complexity.

Assurance

Assurance refers to the ability of the IS auditor to inspire confidence and trust among users of the audit information. This assurance will be accomplished if users of the audit information believe that the IS audit can potential detect errors, mismatch, recognize system flaws and unmask complexities that could potentially jeopardize audit results. The IS auditor's expertise and knowledge is typically assumed and expected to be at a very high level capable of seeing beyond the system and process complexities. Typically, when auditing an enterprise system, the IS auditor needs to provide assurance on process integrity, infrastructure integrity, application security and assurance that controls within the system are appropriate and effective. Yet the degree to which IS audit are able to perform these examination are increasingly becoming difficult as the level of complexity increases.

While section 404 of SOX requires management of SEC registered companies to annually make assertion of the effectiveness of internal control however, auditors still need to provide assurance on these assertions. Assurance can only come from an IS audit if IS auditors are able to decompose the system and understand the underlying business processes. As organizations implement enterprise systems with its inherent complex processes, auditors ability to inspire confidence and trust will diminishes as IS auditors cannot give complete assurance without first gaining comprehensive understanding of how the system processes information. Therefore as process complexity increases due to process integration and enterprise system implementation, IS auditors will need more verification and evaluation of information reliability to give assurance on the IS audit report. Assessing the strength of system controls is a necessary condition for assurance. Therefore as process complexity increases the level of assurance will reduce as IS auditors will be unable to obtain a sufficient understanding of how an entity use of systems may affect controls. Thus we propose:

Proposition 2: IS auditor will perceive a lower sense of assurance when auditing enterprise systems in firms with high business process complexity.

Tangibles

This refers to the underlying IT infrastructure, security facilities personnel and in-house IT department within the organization. In today's auditing landscape where the auditor rely heavily upon client documentation, control structure and business process these tangibles become so essential to IS audit service quality. For example a well documented modularity driven system with clearly identified business processes may require fewer assessment hour and IS audit effort while on the other hand a system with incomplete documentation and multiple layers of integrated IT artifact may require more time, effort and resource to understand. Moreover, the facilities and personnel will have significant impact on the IS audit. For instance IS auditors can work through and interview personnel in an in-house IT department compared to an organization whose IT service are out sourced to an outside party. Hence the role of tangibles becomes very important as business process get more complex as these tangibles act as necessary tools and

instruments that enables IS auditors obtain understanding of systems reliability and assurance. Hence we propose:

Proposition 3: An IS auditor will require the support of more IS tangibles when auditing Enterprise Systems in firms with high business complexity than for firms with low business complexity.

Empathy

This refers to the extent of improved value-added services, individualized services and care that accompany IS audit. This means that the service provider (IS auditor) understands the needs of the clients and is able to show care and personalized service in alleviating those needs [26]. There is need to ensure that IS audit is tailored to suit the specific business processes and Information system deployed by the organization. Individualized service is essential since business processes typically differ across organizations and even with organizations using similar enterprise systems, the level of integration and customization may differ among them. Moreover, empathy can be viewed from the lens of value added service derived from IS audit. Typically companies pay less attention to security implications of enterprise systems configuration during the deployment and implementation phase as such attention only increases implementation time and budget [17]. The challenge therefore for IS auditors is to recommending ways to improve weak controls and in setting up prevention mechanism to mitigate fraud rather than playing the role of detecting control flaws and vulnerabilities in the system. However as complexity increases IS auditors ability to provide these value added services will inevitably reduce. This is because as the IS auditor grapples with understanding and these process complexities, the audit focus will be directed toward assurance and information reliability rather than empathy. Empathy can be accomplished if IS auditors have sufficient understanding of the system as to be able to make recommendation for improvements. For example if IS auditors are struggling to understand the business processes within a client system then providing the client with recommendations for improvement become very unlikely as such service are conditional on having clear understanding of the system.

Proposition 4: An IS auditor is more to demonstrate a less degree of empathy when auditing Enterprise Systems in firms with high business complexity than for firms with low business complexity.

Responsiveness

Responsiveness refers to the auditor's perception of her ability to capture the changes in client environment. As business climate continue to evolve the role of IS audit will inevitably be increased. However, the question is to what degree are the processes and techniques evolving to capture these changing business conditions and these responsibilities. As IS audit consideration continue to increase as indicated in SAS94 what input are introduced to ensure the changing conditions are incorporated in the procedures, techniques, skills and conduct of IS auditors. For example there need to find out to what degree IS auditors are equipped to provide prompt services to clients with high business process complexity compared to clients with simple business process.

Proposition 5: An IS auditor is likely to feel less responsive when auditing Enterprise Systems in firms with high business complexity than for firms with low business complexity.

DISCUSSION AND IMPLICATION

Using SERVQUAL construct this paper offers a new perspective on how business process complexity created by companies could have significant effect on IS audit service quality. As business process complexity increases in an enterprise system environment the audit service quality will witness as corresponding decrease as IS auditors wrestle with understanding the business processes hence lowering their expectations on audit reliability, assurance, empathy and responsiveness. The implication of this is that companies need to consider ways to mitigate unnecessary process complexity as such levels of complexity can become draw backs to audit service quality. In addition, such complexity will inevitability increase audit fee as IS auditors will spend more time than necessary gathering evidence in a bid to comprehend the process flow. Furthermore, as complexity increases the need for more expertise within the audit team will arise leading to a corresponding implication with the audit fee.

One way to alleviate this concern could be through modular implementation of an enterprise system so that processes can be matched with a module hence making understanding of the process framework much easier. In addition, there is the need for organizations to have comprehensive documentation of existing business process. These documentations should be detailed and truly depict the functionality of the business processes. There should be a clear distinction between how business processes are expected to function and how such processes are actually functioning. The impact of integration of multiple business activities and effective allocation of internal control should be examined more carefully. Prior to re-engineering of business processes organization may need to consider how control mechanisms are allocated and whether such allocations are consistent with management policies. This new trend of business process re-engineering where every business activity that can be accommodated is added to the existing process without any consideration on the control mechanism, and process framework is a potential recipe for an incomprehensible process complexity and such additions end up creating business process monsters, which are neither helpful to companies nor to IS auditors. "A complex system that works is invariably found to have evolved from a simple system that works"- John Gaule

DIRECTION FOR FUTURE RESEARCH

This study examined the impact of business process complexity on IS audit process solely from the IS auditors perspective. Hence, the further studies examining more dimensions such as multiple stakeholders, using balance score card will be interesting. For example it would be interesting to examine whether process complexity has an impact in audit fee and audit completion time. In addition, there is also a concern of collinearlity among the variables used in the paper. We acknowledge that there may be correlations among these variables however we believe that each variable had its own merit. Therefore, future empirical research could be done by developing measurement scales to validate this proposed model of audit service quality.

REFERENCES

1. Al-Mudimigh, A., Zairi, M., Al-Mashari, M. *ERP software implementation: an integrative framework.* European Journal of Information Systems, 2001,10, 216.

2. Arteta, B.M and Giachetti R.E. *A Measure of Agility as the Complexity of the Enterprise System,* Robotics and Computer-Integrated Manufacturing, 2004,20, 495 – 503

3. Becker, C., M. DeFond, J. Jiambalvo, and K. R. Subramanyam. *The Effect of Audit Quality on Earnings Management.* Contemporary Accounting Research, 1998,15(1).

4. Cardoso Jorge. *Business Process Control-Flow Complexity: Metric, Evaluation, and Validation.* International Journal of Web Services Research,2008, 5, 2

5. Carnaghan Carla. *Business Process Modeling approaches in the context of process level audit risk assessment: An analysis and comparison.* International Journal of Accounting Information System 2006, 7, 170-204

6. COBIT. *Control Objectives for Information and Related Technology (COBIT),* Information Systems Audit and Control Foundation (ISACF), 2002, 3rd Edition

7. COSO. *Committee of Sponsoring Organizations of the Treadway Commission, Fraudulent Financial Reporting,* 1987–1997. An Analysis of U.S. Public Companies

8. Daigle, R.J, Kizirian, T and Sneathen D, L. *System Controls Reliability and Assessment Effort.* International Journal of Auditing 2005,9, 79-90

9. Davenport, T. Proceed innovation: Reengineering work through information Technology, Harvard Business School Press, 1993

10. David, J.S. *IT Evolution: Why ERP Systems Face Extinction*, Working Paper, Arizona State University, url: http://knowledge.wpcarey.asu.edu/article.cfm?articleid=1378 2007

11. DeAngelo, Linda E. *Auditor Size and Audit Quality*. Journal of Accounting and Economics Dec.1981, 183-199

12. Devadoss, P and Pan, S.L. *Enterprise Systems Use: Towards A Structurational Analysis of Enterprise Systems induced Organizational Transformation.* Communications of the Association for Information Systems, 2007, 19, 352-385

13. Fathee, M.M, Redd, R, Gorgas,D and Modarres, B. *The effects of complexity on business processes reengineering: values and limitations of modeling and simulation technologies.* Winter Simulation Conference, 1998, 2, 1339-1345

14. Gell-Mann, M and Lloyd, S. *Information Measures effective complexity and total complexity*. Complexity 1996,1(1): 44-52

15. Gibbs, J, and P. Keating. *Reengineering controls*. The Internal Auditor 1995,(October): 46-52.

16. Hammer, M and Champy, J. Reengineering the Corporation Harper Collins Publishers New York NY 1993

17. Hendrawirawan, D., H. Tanriverdi, C. Zetterlund, H. Hakam, H. H. Kim, H. Paik, and Y. Yoon. *ERP security and segregation of duties audit: A framework for building an automated solution.* Information Systems Control Journal, 2007, 2: 46–49.

18. Heuval, W.J. *Enterprise Application Integration and Complex Adaptive Systems*. Communication of ACM 2000, 45 (10), 59-64

19. Hoitash, R., U. Hoitash, and J. Bedard. *Internal Control Quality and Audit Pricing under the Sarbanes-Oxley Act.* Auditing: A Journal of Practice & Theory,2008, 27 (1): 105-126.

20. ISACA , IS Auditing Standards, Guidelines and Procedures, Information Systems Audit and Control Association, Rolling Meadows, IL, 2008

21. Information Systems Audit and Control Foundation (ISACF), 1998. Control Objectives for Information and Related Technology (CobiT), Rolling Meadows, IL: ISACF.

22. Jiang, J.J., Klein, G., and Crampton, S. *Measuring Information Systems Quality: SERVQUAL from the Other side.* MIS Quarterly, 2002, 26(2), 145-166

23. Kettinger, W.J. and Lee, C.C. *Perceived Service Quality and User Satisfaction with the Information Service Function.* Decision Science,1994, 25, 5, 737-766

24. McAfee, A.P. *Enterprise 2.0: The Dawn of Emergent Collaboration.* MIT Sloan Management Review, Spring, 2006, 47(3)

25. Nah, F. and Lau, J. *Critical Factors for Successful Implementation of Enterprise Systems.* Business Process Management Journal, 2001, 7(3), 285-296.

26. Parasuraman A, Zeithaml V.A and Berry L.L. *A Conceptual Model of Service Quality and its Implication for Future Research.* Journal of Marketing, 1985, 49, 41-50

27. Pitt, L.F., Watson, R.T., and Kavan, C.B. *Service Quality: A Measure of Information Systems Effectiveness.* MIS Quarterly,1995,19(2),173-187

28. Rettig C. *The Trouble with Enterprise Software.* MIT Sloan Management Review,2007, 49, 1

29. Rosemann, M & Watson, E.E. *Special Issue on the AMCIS 2001 Workshops: Integrating Enterprise Systems in the University Curriculum.* Communications of the Association for Information Systems, 2002, 8(15), 200 – 218

30. SAS No. 94, American Institute of Certified Public Accountants (AICPA), The Effect of Information Technology on the Auditor's Consideration of Internal Control in a Financial Statement Audit, Statement on Auditing Standards No. 94. New York, NY: 2001.

31. Sayana S.A. *IT Audit Basics: The IS Audit Process.* Information System Control Journal, 2002, 1, 20-21

32. Simon, H.A. The Science of the Artificial, Cambridge Massachusetts: The MIT Press Massachusetts, 1982

33. Sutherland, Jeff and Willen-Jan Van den Heuvel. *Enterprise Application Integration and Complex Adaptive Systems.* Communication of the ACM, 2002, 45(10)

ON GOAL-STRIVING AS A MEANS FOR INVESTIGATING THE TECHNOLOGY ADOPTION PROCESS

Alan A. Brandyberry
Department of Management & Information Systems
College of Business Administration
Kent State University
Kent, OH 44242
abrandyb@kent.edu
(330) 672-1146

ABSTRACT

Many prominent researchers have called for an evolution of how technology adoption research is conducted. Although the technology acceptance model (TAM) has provided much insight into the process, there is a developing consensus that the field needs to move on from research centered on this perspective. This research proposed here seeks to operationalize Bagozzi's Goal-Striving Model for a specific research problem. The research model and methodology will be discussed and some preliminary results may be presented.

BACKGROUND AND SIGNIFICANCE

For the past two decades technology adoption/acceptance research has been dominated by variations of the technology acceptance model (TAM) [9][10]. Although these studies have provided many important insights into the adoption decision [4][13][14], there has been increasing agreement among prominent researchers [4][7][13] that continuing to pursue technology adoption/acceptance research using variations of the TAM is unlikely to yield meaningful advances.
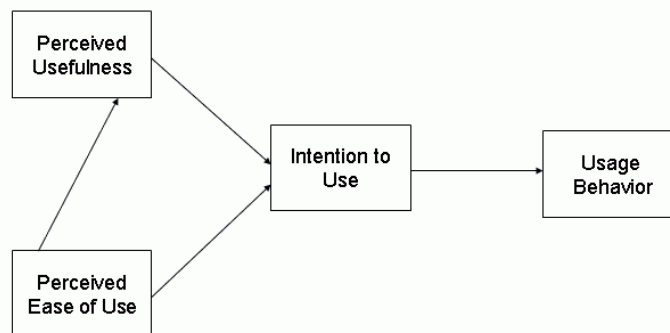
FIGURE 1: Technology Acceptance Model

However, the investigation of these behaviors is still of vast importance from both theory development and practitioner perspectives. The study of information systems is differentiated (from computer science and others) by the fact that the central theme is the study of socio-

technical systems; that is, the study of how human behaviors (both collectively and individually) interact with the technical environment. One of these behaviors that is core to the field and has dramatic practical implications is the study of why an individual chooses to use (or not use) a particular technology. The managerial implications derived from understanding this behavior have obvious importance to technology vendors who are interested in increasing the number of end-users who purchase or license their technological goods or services. A less obvious connection exists within organizations. Organizations often adopt technologies at the organizational level (Enterprise Resource Planning (ERP), Data Warehousing, etc.) and are interested in stimulating voluntary usage to obtain the most value possible from their investment. This is the difference between the binary adopt/reject decision of the organization and the degree of implementation within the organization (percentage of applicable processes that employ the technology).

It has been argued that the TAM's strengths and weaknesses both revolve around its simplicity and parsimony [4]. Its parsimony has allowed researchers to focus on two perceptions of potential adopters: perceived usefulness and perceived ease of use (see figure 1). The insights associated with these studies have dramatically increased our understanding of this complex association. However, many researchers have suggested that this line of inquiry has been 'played out'. Benbasat and Barki [7, p. 212] summarize their view concerning this as follows:

> "Unfortunately, we believe that, in spite of its significant contributions, the intense focus on TAM has led to several dysfunctional outcomes: 1) the diversion of researchers' attention away from important phenomena. First, TAM-based research has paid scant attention to the antecedents of its belief constructs: most importantly, IT artifact design and evaluation. Second, TAM-based research has provided a very limited investigation of the full range of the important consequences of IT adoption, 2) TAM-based research has led to the creation of an illusion of progress in knowledge accumulation, 3) The inability of TAM as a theory to provide a systematic means of expanding and adapting its core model has limited its usefulness in the constantly evolving IT adoption context, 4) The efforts to "patch-up" TAM in evolving IT contexts have not been based on solid and commonly accepted foundations, resulting in a state of theoretical confusion and chaos."

Bagozzi [4, p. 244] states that it "is unreasonable to expect that one model, and one so simple, would explain decisions and behavior fully across a wide range of technologies, adoption situations, and differences in decision making and decision makers." This state of affairs signals to many researchers in the field that we currently find ourselves at a crossroads. Research that introduces and/or validates new models, frameworks, or paradigms that seek to gain a deeper understanding of the adoption/acceptance process and extend the process to adoption outcomes (and thus the value to the individual/organization) is needed.

THE PROBLEM

As discussed above, researchers have pointed to several deficiencies of current technology adoption/acceptance research. Among these are a lack of study of the process of making adoption decisions and a lack of tying adoption decisions to organizational or individual goal

attainment (as motivation and outcome). This goal attainment represents a measure of the value of a technology. Although much of the research to date on the value of technologies has been focused on 'bottom line' indicators such as market cap, revenue enhancement, and/or cost containment, there are inherent difficulties in isolating the effect of a technology adoption from other organizational and market changes over time. It is often more meaningful to focus on intermediate performance enhancement such as attainment of specific goals (cf. [8]).

This decision process of an individual or organization to adopt a particular technology or set of technologies and the determination of the value of information technology to an organization represents two of the most important lines of inquiry in information systems research. Although related, these two lines have rarely been studied simultaneously in the extant literature. This is unfortunate as the value of a technology certainly affects adoption behavior and especially post-adoption behavior (perceived value of various types have been often used in research of adoption behaviors but this is often dramatically different than the realized value of an adopted technology). Conversely, the adoption process may significantly impact the value of a technology. For instance, if an adopter makes the decision to adopt a technology based on specific identified goals that would provide organizational or individual value if realized (rather than non-goal oriented decisions such as those that involve management or individual fashion), there is an increased probability of achieving realized value from that adoption decision.

Several suggestions have been made by researchers as to the direction adoption/acceptance research should take [4][7][13]. To date, none of these proposals have been well explored. This study seeks to adapt the model proposed by Bagozzi [4] which views the adoption/acceptance process as a subset of goal striving. This model specifically incorporates a temporal perspective of goal intention development and subsequent action intention development that lends itself well to studying the process of developing an intention to adopt/accept a technology. Additionally, action enactment (actually adopting and using) along with goal attainment and evaluation (D in Figure 2) is a further extension of the model. This also allows the study of post-adoption behaviors that can be enhanced by applying expectation-confirmation theory [15] and feedback mechanisms into the decision process model.
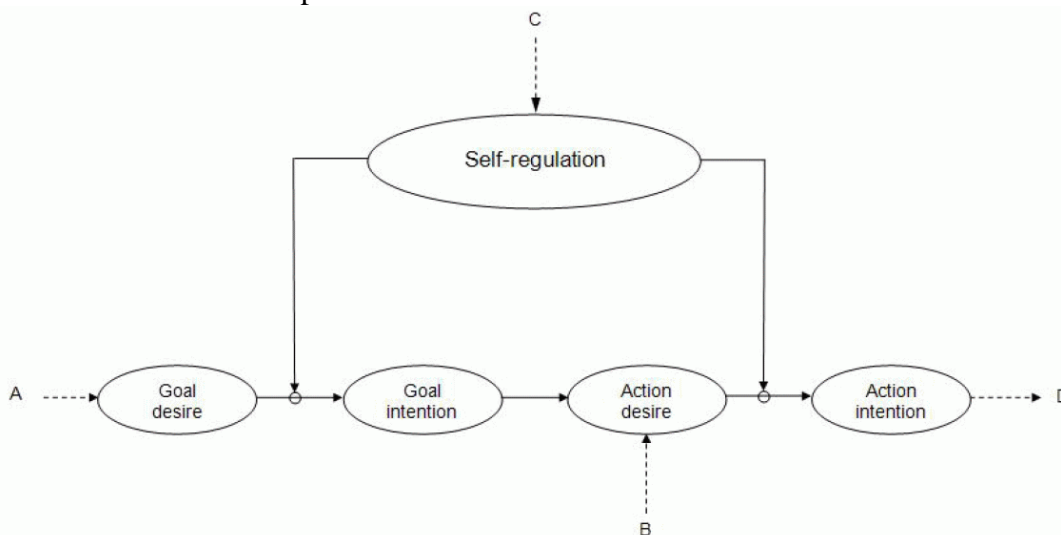
FIGURE 2: Bagozzi's [4] Goal-Action Decision Making Core

Following this, the primary research question proposed is: Can technology adoption/acceptance be appropriately modeled by a goal-striving model such as that proposed by Bagozzi? In addition: What insights into the process can be gleaned by the application of such a model?

APPROACH

Operationalizing the Model

Bagozzi contends that this model may be used to represent the decision process behind technology adoption and acceptance. While Bagozzi has a impressive research career that makes his name recognizable by many people in the fields of information systems, marketing, applied psychology, and sociology it is particularly notable, in this context, that he served on the dissertation committee that introduced TAM [9] and coauthored some of the seminal articles introducing TAM [10][11]. This contention is also strengthened by the relative success of similar models in modeling consumer behavior and other marketing-oriented decision processes [1][2][3][5][6][12].

The decision making core displayed in figure 2 represents the portions of the model that are universal (or nearly so) in scope. These goal-oriented items represent a basic decision making process that lends itself to many types of decisions.

The portions of the model designated by A, B, C, & D represent items that are contextual in nature. The causes represented by A are hierarchical, superordinate goals, values, & motives leading to a focal goal formation in the decision model (such as the goal of adopting a technology). Other items that may be considered in A are beliefs about the means of achieving the focal goal, perceived usefulness of obtaining the goal, perceived ease of obtaining the goal, relative advantage associated with the goal, ingrained attitudes of success, failure, and goal pursuit, amongst others. Causes designated by B influence the creation of a desire for an action-oriented behavior (purchase the technology, try the technology, etc.) and include items such as social norms, social identity, perceived behavioral control, & attitudes towards the action, amongst others. Outcomes designated by D represent goal striving, planning, monitoring, and evaluation. Finally, Bagozzi also contends that self-regulation plays a role in the process. Self-regulation is proposed to moderate any effects of desires on intentions with items represented by C being causes associated with levels of self-regulation [4].

Methodology

The proposed methodology to study adoption behaviors using this model entails a sequence of two separate studies. One study will investigate individual adoption processes among a sample of students. The other will investigate acceptance processes among potential users of a data warehousing system in a specific organization (this may be expanded to multiple organizations if others can be recruited).

The student-focused study will use multiple course sections of undergraduate students and MBA students to study their decision process in determining whether to adopt a specific technology. The chosen technology is a cloud computing application freely offered by Google™ to aid primarily in document accessibility and collaboration (known as GoogleDocs). It is anticipated

that students will have superordinate goals of career success, academic success, course success and other goals that will lead a portion of them to the focal goal of GoogleDocs adoption for a portion of their work. Superordinate goals and intrinsic factors and prior knowledge and use of GoogleDocs will be measured via questionnaires at the beginning of the study. After these baseline attributes are measured, the students will watch an informational video concerning the features of GoogleDocs. After viewing the video, the core aspects of the model will be measured concerning goal and action desires and intentions as they pertain to the possible adoption of the technology. Finally, the concluding portion of the study will measure actual adoption and usage, outcomes, evaluation, and belief modification associated with both adopters and non-adopters. All measurement will be done using questionnaires and adapted scales from previous studies. These will be adapted since no studies exist in the domain of technology adoption but similar studies have been conducted in marketing, health, and applied psychology. Methods to mitigate common method bias will be employed. Data analysis will be done using covariance-based structural equation modeling (EQS software).

The organization-focused study will employ a case-based approach to aid further exploration of associated factors. The participants are members of an organization with an existing data warehouse. Users are aware of the technology and most available features but some utilize the technology to a greater extent than others. A single questionnaire measuring the core model and the set of items represented by A, B, C, & D will be employed. After analysis (using the same methods as the student study) the respondents will be interviewed (structured and open-ended format). These responses will be analyzed using hermeneutical techniques to attempt to shed light on nuances that may have been missed or not fully implemented in the model.

The study of two technologies in two different environments is designed to enable a more comprehensive analysis and interpretations of the ability of the model to capture explain variance in these decision processes. External validity will be enhanced if there are commonalities across the two settings.

Current Status

The research model and data collection instrument are in the process of finalization. Data for the student study will be collected beginning in January 2011. The dates for collection of the corporate data are being negotiated. The full research model and possibly preliminary results of the student study will be presented.

<div align="center">REFERENCES</div>

[1] Bagozzi, R.P. *The Self-Regulation of Attitudes, Intentions, and Behavior*. Social Psychology Quarterly, 1992, 55,178-204.

[2] Bagozzi, R.P. *Consumer Action: Automaticity, Purposiveness, and Self-Regulation.* in N.K. Malhotra (ed.), Review of Marketing Research vol. 2, Armonk, NY: Sharpe, 2006, 3-42.

[3] Bagozzi, R.P. *Explaining Consumer Behavior and Consumer Action: From Fragmentation to Unity*. Seoul Journal of Business, 2007, 12(2), 111-143.

[4] Bagozzi, R.P. *The Legacy of the Technology Acceptance Model and a Proposal for a Paradigm Shift.* Journal of the Association for Information Systems, 2007, 8(4), 244-255.

[5] Bagozzi, R.P. & Dholakia, U. *Goal-setting and Goal-striving in Consumer Behavior.* Journal of Marketing, 1999, 63, 19-32.

[6] Bagozzi, R.P., Dholakia, U.M., & Basuroy, S. *How Effortful Decisions Get Enacted: The Motivating Role of Decision Processes, Desires, and Anticipated Emotions.* Journal of Behavioral Decision Making, 2003, 16, 273-295.

[7] Benbarat, I. & Barki, H. *Quo vadis, TAM?* Journal of the Association for Information Systems, 2007, 8(4), 212-219.

[8] Brandyberry, A. A., Rai, A., & White, G. *Intermediate Performance Impacts of Advanced Manufacturing Technology Systems: An Empirical Investigation.* Decision Sciences, 1999, 30(4), 993-1020.

[9] Davis, F.D. *A Technology Acceptance Model for Empirically Testing New End-User Information Systems: Theory and Results.* Doctoral dissertation, MIT Sloan School of Management, Cambridge, MA, 1986.

[10] Davis, F.D., Bagozzi , R.P., & Warshaw, P.R. *User Acceptance of Computer Technology: A Comparison of Two Theoretical Models.* Management Science, 1989, 35, 982-1003.

[11] Davis, F.D., Bagozzi , R.P., & Warshaw, P.R. *Extrinsic and Intrinsic Motivation to use Computers in the Workplace.* Journal of Applied Social Psychology, 1992, 22, 1111-1132.

[12] Dholakia, U.M., Bagozzi, R.P., & Gopinath, M. *How Formulating Implementation Plans and Remembering Past Actions Facilitate the Enactment of Effortful Decisions.* Journal of Behavioral Decision Making, 2007, 20(4), 343-364.

[13] Fichman, R.G. *Going Beyond the Dominant Paradigm for Information Technology Innovation Research: Emerging Concepts and Methods.* Journal of the Association for Information Systems, 2004, 5(8), 314-355.

[14] Goodhue, D.L. *Comment on Benbasat and Barki's "Quo Vadis TAM" article.* Journal of the Association for Information Systems, 2007, 8(4), 220-223.

[15] Oliver R. L. *Effect of Expectation and Disconfirmation on Postexposure Product Evaluations - an Alternative Interpretation.* Journal of Applied Psychology, 1977, 62(4), 480.

A Design Theory for Strategic Decision Systems

ABSTRACT

In this paper, we examine the development of information systems to support the strategic decisions for a retail chain expanding nationally. The company's strategic decisions were clearly an emerging knowledge process (EKP), and insights into how the decisions emerged, and supporting systems developed, are offered. Additionally, we find that existing EKP design theory does not apply well to the systems that were developed, and that an alternate design theory has greater perceptive capability. We conclude that design theory for an EKP will differ depending upon the type of emergent process.

INTRODUCTION

Strategic decisions are described by Mintzberg and his associates as decisions of importance to an organization, and strategic decision making as an unstructured process characterized by "novelty, complexity, and open-endedness" (Mintzberg 1978, Mintzberg, Raisinghani and Theoret 1976). Mintzberg describes organizations as often being only vaguely aware of a decision's starting and ending points, and the decision process proceeding iteratively with discontinuities. Mintzberg also developed classifications for strategic decisions, on a continuum ranging from voluntary decisions stimulated by new opportunities to mostly involuntary decisions stimulated by crises. Here we are focused on the former, those strategic decisions stimulated by new opportunities such as decisions to enter new markets or expand share in an existing market.

1

Strategic decisions have been characterized as having no precedents, being ill structured, and not easily modeled or analyzed. They are complex, nonlinear, and fragmented; thus not well suited for decision support models (Bennett 1998). Strategic decisions are a form of an emergent knowledge process (EKP). EKPs have no known best sequences of steps, evolve dynamically, and entail unpredictability in the job roles and required knowledge (Markus, Majchrzak and Gasser 2002). Other processes termed EKPs include organizational design and new product development. Strategic decisions, and especially those which have little precedent and are motivated by new opportunities, are prime examples of EKPs (Markus, Majchrzak and Gasser 2002).

The highly unstructured and unpredictable nature of EKPs are likely reasons why development of information systems to support EKPs has proven more difficult than developing more traditional information systems such as transaction processing systems. Correspondingly, design theory has been well known and established for these more traditional types of information systems, but has yet to be fully developed for all EKPs, and especially for strategic decision applications. Design theories are prescriptive (Walls, Widmeyer and El Sawy 1992), and in the case of information systems, describe what the systems should do and how they should be developed. Information system design theories act as guides for developers involved in creating systems, and include goals, artifacts to be produced in the process of reaching those goals, and hypotheses about the effectiveness of these artifacts (Hevner, March and Park 2004).

Design theory has existed in other disciplines, such as architecture, for some time. Design theory has also been developed for several classes of information systems, including

2

executive information systems (Walls, Widmeyer and El Sawy 1992) and systems to manage organizational competence (Lindgren, Henfridsson and Schultze 2004). Although there has been some design theory for some specific EKPs (Markus, Majchrzak and Gasser 2002), there appears to have been little proposed for strategic decisions, which are also EKPs.

Many software vendors have created products intended to support some aspect of strategic decisions, but the results to date have led to a large but fragmented market for such decision-aiding software. While not representing the entire range of strategic decision-making software products, one particularly active area with a wide strategic scope is that of Business Intelligence (BI) software. The market for BI software – often defined as including OLAP (on-line analytical processing), data mining, and modeling tools – was estimated by International Data Corporation (IDC) at $15 billion in 2005 with a forecasted annual growth rate of 10% (Computerworld 2006). Of course, not all applications of BI software are for strategic decision making; and these estimates are for commercial software only and do not include the considerable number of 'home-grown' and independently developed strategy-aiding applications.

Most importantly, BI software does not often address the EKP aspects of strategic decision making. Data mining tools, for example, produce only highly structured results by analyzing large quantities of data. Other software such as Enterprise Resource Planning (ERP) software usually is focused on enterprise and inter-enterprise operations (such as supply-chain management), and less on strategic decisions. While these types of information systems and software tools have made great strides since the time of Mintzberg's research, they have not yet been capable of truly supporting the EKP aspects

of strategic decisions. In addition to its research contribution, a design theory for this class of EKP would be of benefit to practitioners, and possibly spur the application of information systems to new and higher levels of business planning and decision making.

This research presents a case study of a company endeavoring to create information systems to support their strategic decisions and, from the case-based evidence, extracts design theories for strategy-aiding information systems. This company is a large European manufacturer of confectionary products that was seeking to expand its presence in the United States. As the company prefers not to be named, we will refer to them as Venus Chocolates. The strategic decisions were focused on company could transform a very small but successful group of six retail stores in the U.S. into a nationwide chain of perhaps 300 or more stores. The decision process to be followed was – as strategic EKPs are – highly unstructured, complex, with no best set of steps known *a priori*. This paper describes how existing design theory that is applicable to other types of EKPs was not entirely applicable in developing the information systems required to support these strategic decisions. From the case evidence, the paper extracts and proposes new elements of design theory for this class of decisions.

The remainder of this paper is organized as follows. First, the theoretical background of design theory, and design theory for EKPs, is described. The case study methodology is detailed, followed by a description of the case itself. Design theory derived from this case is next, followed by discussion and conclusions.

THEORETICAL BACKGROUND

IS Design Theory

4

Information system (IS) design theories are not new and have been well established for many classes of information systems. IS design theories are distinct from other IS theories in that they are prescriptive rather than descriptive or predictive (Walls, Widmeyer and El Sawy 1992), focusing on how problems can be solved or goals achieved. IS design theories are based on stated requirements, for which a process is prescribed. That process produces one or more artifacts intended to best meet the stated requirements.

When Walls, Widmeyer and El Sawy (1992) developed an IS design theory for executive information systems, they noted that commercial software vendors were responding to market needs, rather than proceeding from specific design theory. They argued that design methods based on well founded IS design theory, consisting of kernel theory, meta-requirements, and system artifacts, would lead to better software. We believe the situation today is similar for business intelligence software, particularly when applied to strategic decision support.

Just as design is both a verb and a noun, both processes and artifacts are encompassed by IS design theory. The artifacts can include constructs, models, methods, and instantiations (March and Smith 1995). As defined by March and Smith, constructs are the terms and language by which concepts are communicated. Models provide representations of reality and define solution spaces. Methods are documented processes, and instantiations are the implementations of information systems.

Relational database theory is considered to be an IS design theory (Walls, Widmeyer and El Sawy 1992), as is the System Development Life Cycle (SDLC). Each of these prescribes processes by which to achieve a desired result (for example,

5

normalization to produce a consistent database or the waterfall method to develop a system), and the artifacts to be produced (for example, a database or a completed system, both instantiations).

IS design theory provides guidelines for system developers, and as a result has sometimes been incorporated as part of best practices. A best practice used by developers can include all the elements of design theory. For example, a best practice for creating object-oriented software might include a process such as the Rational Unified Process (IBM 2006), and produce UML artifacts such as class, sequence, and collaboration diagrams. As with design theory, these best practices can prescribe both process and artifact.

Best practices by themselves, however, do not constitute IS design theories. Best practices are established and routine (Hevner, March and Park 2004). These writers further argue that to be relevant, IS design theory research needs to either address new, unsolved problems, or to solve problems in innovative, more efficient ways. For example, the research by Markus, Majchrzak and Gasser (2002) was sparked when existing design theory was found lacking for the application they were developing.

Developing IS design theories is an iterative process (March and Smith, 1995). The two processes identified by March and Smith are "design" and "build". Theories and their measurable hypotheses must be developed, systems built using those theories as guidelines, leading to IS design theories being empirically validated and modified. The criteria for evaluating design theories can be complex and partly pragmatic – a matter of what 'works' (Hevner, March and Park 2004). The intent of our research is to evolve design principles, which in turn may be evaluated by future research.

To be relevant, IS design theory research needs to address *wicked* problems (Rittel and Webber 1984). Hevner, March and Park (2004) typified these problems as having:

- Unstable requirements,

- Ill-defined environmental contexts,

- Complex interactions among subcomponents of the problem and solution,

- Inherent need for flexibility in design process and artifacts,

- Dependence on human cognitive abilities (creativity) for success, and

- Dependence on social abilities (teamwork) for success.

IS Design Theory for EKPs

An arena for IS design theory research that meets these criteria is emerging knowledge processes (EKPs). In the seminal work by Markus, Majchrzak and Gasser (2002), EKPs are characterized by:

1) "an emergent process of deliberations with no best structure or sequence",

2) "an actor set that is unpredictable in terms of job roles or prior knowledge," and

3) "knowledge requirements for general and specific distributed expertise."

Strategic decisions have been described as emerging (Dutton and Jackson, 2002). Markus, Majchrzak and Gasser (2002) specifically included strategic business planning as a type of EKP. The strategic decisions faced by Venus Chocolates meets these three EKP criteria.

1) Venus Chocolates was pursing a new opportunity in a previously unknown channel of distribution. When the vision of an extensive chain of retail outlets was first proposed, the organization was at a loss as to how to proceed. For Venus Chocolates, the retail venture was truly uncharted terrain with no known path or sequence of steps to be taken.

2) The roles of the individuals involved unfolded over time. The set of decisions and systems to support those decisions evolved to include several functional areas, and entailed many different roles, all largely unpredictable at the outset. The entire Venus Chocolates retail organization was being started from scratch. Not only were the job roles involved with strategic decisions unpredictable, novel job roles themselves kept emerging over time.

3) The strategic decisions faced by Venus Chocolates required intensive expert knowledge. Some of this knowledge was tacit, some poorly organized, some difficult to access and much was distributed across functional areas and job roles. The systems supporting these decisions had to harness such tacit and dispersed knowledge for effective strategic decision-making.

The retail venture of Venus Chocolates represented an EKP and the design theory developed presented is for systems to support that EKP. Design theories and their artifacts may help further define the core of the IS discipline (Weber 2003). The contribution of this research is to develop a first iteration of an IS design theory for EKP systems that support strategic decision making.

RESEARCH METHODOLOGY

Case Study Research

Case studies have been an accepted methodology for IS research for some time (Orlikowski and Baroudi, 1991), and case studies in particular are an appropriate methodology for studying IS design theories (Hevner, March and Park 2004). The methodology used here is that of a historical case study, based on analyzing IS design for a specific company and situation. Although analyzing data from the past eliminates the possibility of co-creating a design theory while designing and deploying a system – like Markus, Majchrzak and Gasser (2002) did – historical case studies provide great scope for reflection and are useful as a method for extracting knowledge from events that have already occurred (O'Brien, Remenyi and Keaney 2004). Case study research is most appropriate when a phenomenon cannot be studied outside its natural setting, when no control or manipulation is necessary (or feasible), and there is not an established theoretical base (Benbasat Goldstein, and Mead 1987).

This research presents a case study of Venus Chocolates, a European confectioner seeking to expand its presence in the United States through the introduction of a chain of retail stores. One of the authors was involved as a contract employee for Venus Chocolates during the years in which they were embarking upon this effort. This contract was for a sequence of projects, undefined at the outset, each of which was to develop systems and solutions for strategic decisions. This author had no research agenda during the time of this contract; however, virtually all memos, personal note journals, documents, and software were retained – thus paving the way for this historical analysis.

Data Support

There are five sources of evidence mentioned by Yin (1984) that can work well for case studies: documentation, archival records, interviews, direct observation, and physical artifacts. This research included data from each of these sources.

Documentation included project reports, memos, and presentations. Many of these presentations were to senior management about proposed strategic directions and plans. Also among the documentation were specific outputs and reports from systems designed to support those strategic decisions, as well as internal reports such as financial statements. The archival records included formal meeting minutes, informal notes taken during meetings and discussions,journal notes, and e-mails. Aside from interviews during the projects, for which informal notes were the primary source of data, interviews with other participants were conducted afterwards for fact verification. Evidence from direct participant observation in meetings, presentations, and observations while working within the organization was available for this research. Finally, physical artifacts retained included the actual software and data. The data sources available for this analysis were wide and deep.

Validity

Validity is an issue to be addressed by any research, and for qualitative research there are four criteria by which validity may be ascertained (Guba and Lincoln 1994). Those criteria are credibility, transferability, dependability, and confirmability.

Credibility: In much of qualitative research, assessing the degree of credibility, which is analogous to internal validity, depends to a great degree on the participant (Truchon 2005). For this case study, credibility is enhanced by the fact that these strategic support

systems were actually used, and used successfully. Credibility of this research is related to the fact that the design theory proposed here was developed from experience with systems that were in frequent use and underwent many cycles of revision.

Transferability: Similar to external validity, transferability signifies the extent to which research can be generalized (Guba and Lincoln 1994). This research consists of analyzing only one case, but the case was prototypical. We believe that at a minimum, this research is applicable to large international companies attempting to penetrate the U. S. market through rolling out specialty retail stores, and is likely applicable to many other strategic decision systems as well.

Dependability: The design theory developed in this case study is based on experience with an EKP over a period of approximately two years. By its very nature, an EKP develops and changes, and the systems described in this case study were each developed over several iterations characterized by false starts and trial and error; and not as a single fully-formed system or artifact. The dependability of this research is grounded in the length of time, number of system elements, and system development iterations on which this design theory was developed. The iterative learning reinforces the dependability of the theory developed.

Confirmability: Confirmability in qualitative research is analogous to objectivity and refers to the degree to which the results could be confirmed or corroborated by others (Meyers 1997). An author who was not involved with the Venus Chocolates project analyzed the data sources from a detached standpoint. This author played the role of 'devil's advocate', especially in offering alternative explanations for phenomena found in

the data. Eventual convergence of the detached and the immersed perspectives gives us confidence in the confirmability of the findings.

CASE STUDY

Until the point at which this case study commences, Venus Chocolates had been distributing its line of premium confectionary products in the U.S. through a wholesale channel. They had developed U.S. manufacturing capacity in New England for one of their major product lines, but relied on imports from Europe for other lines.

Venus Chocolates was aware that per capita consumption of their product category was far lower in the U.S. than in Europe, and was convinced an opportunity was present. The success of the small chain of six U.S. retail stores that had been opened as an augmentation to the wholesale business also influenced them. Even though only a small fraction of the overall business, these six stores were highly profitable and the expansion potential of this success was hard to ignore. Guided solely by this vision, and confident that great potential for a retail chain must exist, an individual with extensive retail experience was hired in the newly created position of Vice President of Retail.

Bounding the Strategic Decisions

The set of strategic decisions necessary to fulfill this vision were only vaguely defined, as noted above, and were complex and intertwined. Management was apprehensive that a significant rollout of retail stores would greatly impact, and might dwarf, their existing operations and organization. They were concerned Venus might be unable to meet the manufacturing, distribution, and infrastructure requirements of a rapid, large-scale retail network rollout. But they also felt a keen sense of urgency about entering this new channel, knowing the window of opportunity might close quickly, and that competitors

12

were considering similar types of expansion. The set of strategic decisions was clearly on the end of the strategic decision spectrum Mintzberg described as "voluntary and stimulated by new opportunities."

Fulfilling this retail potential meant not only the creation of an entirely new retail organization and a new channel of distribution, but the transformation of an existing organization and operations. A sentiment voiced by a member of management was "We don't even know what we don't know," evidencing the emergent nature of the decisions to be made. It was at this point that one of the authors of this paper began his relationship with the company.

Venus Chocolates embarked upon a cycle of meetings held approximately every quarter, and many informal meetings were held as well. The purpose of these meetings was to review progress and results, and to formulate new strategic decisions. One consequence of these meetings was that the strategic direction of the retail stores rollout was prone to dramatic changes. Thus, any strategic decision system under development could be scrapped or significantly revised.

One result from an early meeting was the determination that estimating the size of the potential retail market was an appropriate starting point, and main driver, for strategic decisions. Figure 1 shows an ad-hoc diagram informally created during a meeting. This figure shows how market potential estimates were considered to have direct impacts on existing infrastructures and the retail operation necessary to reach that potential.

Some strategic decisions that hinged on the size of the potential retail market included product mix and marketing strategy as well as the numbers, profiles, locations, and rollout timing for the new stores. Some of the interrelated strategic decisions for the

13

existing operations included capacities and schedules for manufacturing, storage, and distribution. For example, if domestically made products (as opposed to European made and imported) were to be a major element in the retail product mix, how much more manufacturing capacity would be needed? How much additional storage space for raw materials and finished goods would be required? If some capacities were to be exceeded prior to the introduction of additional capacity, could outside facilities be leased, or would European capacity be available? If so, what would be the additional timing, transportation, and storage requirements for these imports?

The concepts illustrated in Figure 1 pointed to sets of interrelated and circular decisions – a product mix planned for the retail stores needed to be based on manufacturing and distribution capacities, but planning those capacities required assumptions about specific product mix. Both depended on the size of the potential retail market, how fast the chain could be expanded, and the degree of interaction – positive and negative – between the retail and wholesale channels.

--- Insert Figure 1 about here ---

Even though new, the decisions could not be based on a vacuum: emergent knowledge would have to be created to make these decisions, Figure 1 helped provide context and more importantly, a set of boundaries for the upcoming strategic decisions. Defining and drawing boundaries is one indicator of success for companies in dynamic, high-velocity markets (Eisenhardt and Martin, 2000).

Strategic Decision Systems

A software tool to estimate market potential was the first strategic decision system created. This software, which consisted of a series of models, estimated total market size

based on a data from the existing small chain of retail stores, demographics, trends in product consumption, and specifics about store profiles (examples of store profiles include mall stores, freestanding stores, and stores in strip centers).

This system, termed the Market Potential Estimator (MPE), and each of the systems to follow, had high visibility within the organization. Not only were various members of management the primary users, but also the outputs of these systems were used extensively in the quarterly meetings. The formal quarterly meetings, and the more frequent informal meetings, created needs for new information and strategic decisions. As a result, the development process for these systems was conducted with a significant sense of urgency.

This environment and set of circumstances had several direct effects on system development for the MPE and each of the systems that would follow:

1) Specifications for system functionality were highly informal. Other than drawings made during meetings, there were few, if any, written system requirements. During meetings, management could articulate desired functionality only in general terms, mostly verbally, and had no time to either elaborate this functionality in detail, or to review and approve system functionality and design. Gathering complete, or even mostly complete, requirements was infeasible, as was any development methodology that relied on having them.

2) Designing the system for code reuse was impractical. With such informal specifications, designing for modifications and reuse could be based only on estimates of the future requirements for the software. Building any additional functionality, which may or may not be needed, meant delaying the completion

15

and use of the system, which would have been unacceptable. The systems were

also developed within the context of the quarterly meetings, which could change

the direction and requirements for these systems very quickly. An example of

how this was implemented can be found in the creation of a very rigid data

structure for demographic data. A more complete design would have included a

general purpose data structure flexible enough to handle general demographic

files containing new data items. This is a typical tradeoff between specific

functions developed quickly, and generalized but flexible functions developed

less quickly. The emergent nature of these strategic decisions meant that any time

or effort to build flexibility into the software had to be considered in light of the

possibility that the entire software system would be scrapped or completely

reconfigured.

3) Traditional design artifacts, such as class or entity-relationship diagrams, were of

minimal value. Such artifacts had some value as documentation, but with such

informal specifications, they could not be linked to any system requirements

artifacts such as use-cases. These artifacts were also subject to large (and likely)

changes in response to changes in system functionality, calling into question the

value of efforts to create a set of traditional artifacts.

4) Existing design theory for other types of systems was of little help, as

development methodologies and artifacts were of only marginal use. Markus,

Majchrzak and Gasser (2002) reported similar findings when applying existing

design theory to their EKP. The development methodology, which needed to

build loosely defined systems for moving targets, was itself emergent, and

characteristic of what Truex, Baskerville and Klein (1999) advocated for systems development in emergent organizations. In these organizations, Truex, Baskerville and Klein viewed traditional IS development goals – lengthy analysis and design, user satisfaction, abstract requirements, and unambiguous specifications – as poor investments of time/effort, improbable in terms of achievement, imaginary, and ineffectual.

With the MPE indicating significant retail potential, the build-up of a retail organization began and the need for more knowledge emerged. To address those needs, one system was developed to determine the best locations for stores (BLS), another system to analyze the economics of the retail expansion (termed the Business Proposition Analysis, or BPA), and yet another to determine logistical requirements (DLR).

Emergent Nature of Strategic Decisions and Systems

The MPE indicated significant clusters of stores were viable in the northeast and other locations nationwide. Based on this, the need to develop a highly specific store rollout plan emerged, including the opening schedule, location, and store profile for each new retail store. The BPA was designed and developed to answer this need. Once BPA was in place, more needs emerged for decisions to site individual stores, and to start transforming the wholesale infrastructure to incorporate retail operations. This required two additional systems to be developed: the Store Location Model (SLM) and the Network Model (NM).

The retail opportunity was pursued aggressively. Within six months, 13 new retail stores were up and running, tripling the size of the chain from the beginning of the year.

17

These systems were revised and used repeatedly to develop plans that included opening

more than 40 new stores during the next year.

Intertwined Strategic Decisions and Systems

The EKP of strategic decision-making drove the development of systems to support those

decisions (see Figure 2). Just as the next set of strategic decisions could not be stated

until they emerged, the functionality and scope of new systems could not be

predetermined. To varying extents, each new system incorporated outputs from previous

systems. The systems were intertwined just as the strategic decisions were. For example,

the BPA system evaluated plans for store openings based on store profiles and cluster

locations. Those profiles and cluster locations, however, were determined by the Store

Location Model (SLM), which in turn used results from the MPE. When the first system,

the MPE, was developed, concepts such as using multiple store profiles, or opening stores

in specific cluster sizes, were not yet seen as key considerations. The MPE was used to

help define these, and each of the subsequent systems required output that the MPE had

to offer.

--- Insert Figure 2 about here ---

Need for New Systems

New strategic decisions emerged as management evaluated results from the retail chain

and compared them to what had been expected. Quarterly review meetings often resulted

in the investigation of new or modified directions based on field results. In turn, a stream

of new systems was needed to support those decisions. As the organization grew and

changed, the job roles proved unpredictable. Existing design theory – usually requiring

specifications to be gathered up front, formal development methodologies, design

18

artifacts, or planned stages – would have been largely inapplicable for these systems. A brief description of some of these systems follows:

o **Model Stock (MSS)** – this system was used to determine desired inventory positions for each store during each week, particularly during seasonal peak weeks. It was also used to determine corresponding storage and transportation requirements needed to reach and maintain those target inventory levels. Model Stock used results produced by the BPA and the Network Model as part of its inputs, along with sales transaction data from each store. This is one example of many linkages between the strategic decision systems.

o **Ramp-up System (RUS)** – through the use of the BPA, one factor that emerged as a highly significant determination of the success of each store was how fast revenue ramped up. This factor was confirmed in the field. The ramp-up system used sales transaction data from new stores in order to develop aggregate revenue forecasts for each existing and planned store. The BPA was modified to use the output from the Ramp-up system (RUS). Therefore feedback loops were created between these systems.

o **Forecasting System (FS)** – using the aggregate revenue estimates from the Ramp-up system (RUS), the Forecasting system (FS) produced estimates of required shipments of each individual product. This is an example of a strategic decision system with linkages to operational systems.

o **Break-even System (BES)** –This system was used in the field to analyze potential new store locations, based on specific proposed lease terms and revenue estimates.

19

- Storage Capacity System (SCS) – the U.S. facility in which Venus Chocolates manufactured products also served as a storage facility for raw materials and all finished goods, both imported and domestically produced. Highly seasonal peaks, coupled with a quickly expanding retail chain, meant that the capacity of the existing facility would be exceeded within two years, a great concern. The Storage Capacity system used outputs from the BPA and Network Model (NM) to help determine when the existing storage facility would reach capacity, and how large a new facility would be needed.

Need for Revised Systems

Initial results from the Storage Capacity system indicated the need for an enormous storage facility. A new group of users, the retail managers, questioned the size and financial capital to build this facility. An engineering consulting firm was engaged, and this firm found that the strategic decision system was functioning properly but the underlying assumptions made by management and used by the BPA were incorrect. This necessitated a revision of Venus Chocolate's strategy. A new strategy was developed, supporting by the re-use of many of the strategic decision systems already in place. Ultimately, a much smaller facility was constructed, illustrated the intertwined nature of strategic decisions and these systems.

This was traced to specific assumptions used in the BPA. The logic used by the systems was reasonable; the assumptions in question were critical. An engineering consulting company was called in to review the assumptions and design the new facility, providing a reality check and demonstrating another use of these strategic decision systems. Ultimately, ground was broken on a new, smaller storage facility.

As the retail chain grew, new strategic decisions emerged that required new knowledge and new supporting information systems. Some emergent strategic decisions also prompted the need to revise, potentially extensively, one or more existing systems. For example, after using these systems for almost a year, results suggested that the mix of retail outlets should include kiosks. The strategic decision then entailed deciding whether or not to open kiosks for the next peak season, and if so, how many, what product mix, and so on. As with many of the strategic decisions Venus Chocolates made, the job roles involved were unpredictable – even nonexistent –at the outset. As with many strategic decisions that emerged, this led to substantive modifications to the suite of systems that were in place.

Kiosks were completely outside the scope of management's thinking during the first year of retail expansion. It was only after experience with new stores was gained that the idea gained consciousness, and it is very unlikely that kiosks would have been mentioned in any use-cases or requirements-gathering effort. Even had they been considered, however, little about operating kiosks was known. For example, it turned out that each kiosk in operation would require a nearby store in order to provide additional storage. A kiosk could only exist as a satellite to another retail store, which was a completely unknown requirement when stores were first being rolled out. Even if kiosks had somehow been included as a store profile at the outset, they still would have needed major modification to incorporate them into the systems later on. This contingency is only one of possibly hundreds that would have been needed in order to determine a complete set of requirements.

Some revisions also required data from new sources to be created and utilized. For example, after observing the performance of stores in new markets, one hypothesis that management wanted to evaluate was that awareness levels of Venus Chocolates' products were a key factor in first year store sales. In order to test this hypothesis, surveys were conducted in several markets. The data from these surveys needed to be incorporated in the BPA. At the time the BPA was developed, neither this hypothesis nor these surveys were envisioned. As with the kiosks, a substantial upfront design effort in anticipation of potential surveys would have been counterproductive, indeed infeasible.

Need for System Linkages

The iterative nature of the use of these systems made clear that each of these systems needed to be linked. This was not anticipated at the outset, but ultimately a web of systems to support strategic decision-making was created. As each system was revised, or new systems were created, they were interfaced to one another. In the same way as the need for each system emerged, the needs for each interface between these systems emerged. The configuration of these systems and their linkages is shown in Figure 3.

--- Insert Figure 3 about here ---

Venus Chocolates continued using this set of strategic decision systems to further expansion until the chain reached almost 80 stores, all in clusters on the East Coast and stretching from Maine to Florida. As new retail stores continued to be opened during the subsequent two years, new data was inputted into these strategic decision-making systems, and these systems were used, revised, and reformulated repeatedly. The initial set of the systems had become a stable set of tools, and some emergent aspects of the

22

strategic decisions eventually became routine. The number of new revisions and new systems that were needed dwindled. At this point, the author who had been a contract employee started to pursue other projects and lost daily contact. Ultimately, as more retail chain results and experience were gained, management concluded that the fastest way to expand, especially on the West Coast, was through acquisition. A significant competitor was acquired by Venus Chocolates, and shortly afterwards, the initial set of systems was abandoned.

DESIGN THEORY

The case study in the previous section outlines Venus Chocolates' strategic decisions and systems. Data accumulated during the time those systems were built was carefully analyzed, much as secondary data might be, for observations related to design theory. These observations were rigorously examined for validity and alternative explanations, and some were eliminated or modified. Some of these observations were related directly to design principles, while some were more related to system requirements.

We developed design theory consistent with those observations, and were able to frame these observations as either a design principle or requirement. We were also able to identify elements of kernel theory as components of this design theory. While this kernel theory may not be complete, the elements are related to the requirements, which in turn drive the design principles. This structure of kernel theory, system requirements, and design principles was derived from the seminal work by Markus, Majchrzak and Gasser (2002), particularly Figure 7 of this work, and also from work by Hanseth and

Lyytinen (2004). These elements of kernel theory, the system requirements, and design principles are summarized in Table 1.

--- Insert Table 1 about here ---

Kernel Theory

Much of the kernel theory is derived from the management literature. A model of strategic decision-making effectiveness by Dean and Sharfman (1996) found procedural rationality to be a significant determinant of such effectiveness, and concluded that managers who gathered more information and used more analytical techniques made more effective decisions. Venus Chocolates' management did follow a reasonably rational process, and was continuously hungry for new information.

With little knowledge or experience in retail operations and marketing, Venus Chocolates' management was presented with a steep learning curve. In such dynamic situations successful firms often seek to develop new, situation specific knowledge quickly through experimentation and trial and error. These firms tend to bound potential solutions, operate with deadlines, and develop simple rules, but do not create complex routines (Eisenhardt and Martin 2000).

The speed with which strategic decisions are made has been theorized to impact the effectiveness of those decisions, as has the extent to which a multitude of decision alternatives considered, and the degree to which the decisions are integrated both across functional areas, and with operational systems (Eisenhardt 1989). Support for these theories could be found in Venus Chocolates' systems.

Strategic decisions are often approached iteratively (Mintzberg, 1975), and can be sequentially linked by nesting, snowballing, or recurrence (Langley et. al. 1995). If nested, a series of sub-decisions lead to a larger decision, in a fashion similar to evaluating linked nodes of a decision tree. Snowballed decisions are a series of smaller decisions leading to a larger decision, one that potentially spans many functional areas. Recurrence simply refers to repetitive decisions. Venus Chocolates' strategic decisions were from each of these categories.

Each of the theories described bear on the system requirements for systems to support strategic decisions, and therefore considered to be kernel theory components for the design theory proposed here.

There also exists theory for strategic decisions by international firms seeking to penetrate markets in new countries by opening retail stores. There is evidence that firms in Venus Chocolates' situation performed better in the long run when strategic decisions were made to open a substantial chain of stores as early as possible, in a store format new to the host country but familiar to the corporation, and without partners or acquisitions (Gielens and Dekimp 2001). Venus Chocolates' actual results would seem to support these findings.

There also exists a body of theories specific to retail chains. One example is for retail store locations in a dynamic environment (Ghosh and Craig, 1983), which includes a competitive equilibrium model to determine store locations. Many such theories could be applicable to Venus Chocolates' situation. However, the value of a design theory is derived from being applicable to as wide a class of problems as possible (Walls, Widmeyer and El Sawy 1992). We believe the proposed design theory to be applicable

25

to a wider range of problems than those of international companies seeking to penetrate new markets by introducing retail chains, and so believe it is unnecessary to include theories specific to retail chains as kernel theories.

System Requirements

The system requirements were related to kernel theory elements – management did choose a rational process for strategic decision-making, despite the pressure, and they were the primary system users. They were anxious users, often eager to seize on results produced by systems for actual implementation, seemingly almost as soon as they emerged from the printer.

At first glance, the lack of a requirement for system performance might seem at odds with the kernel theory indicating that speed of decision is performance. However, in this system requirement, performance relates to response time. Unlike transaction processing systems, responsiveness is not a factor in strategic decision systems – a system that takes an hour to deliver an accurate and reliable response is far preferable to one that returns a poor answer in a second.

While Venus Chocolates' systems did not initially involve many functional areas, they grew incrementally until they involved every operational function and managers from each of those functions. As has been noted, many of these systems were interfaced with each other, and to a lesser extent, to operational systems.

Design Principles

The analysis of the data from this case produced a series of observations. Each observation was examined for alternate explanations and validity, and many were modified, combined, or discarded. These design principles were derived from the

26

resulting set of observations, and are summarized in Table 1. This table also shows the relation of each design principle to a specific system requirement, and the link of each system requirement to a specific kernel theory.

User satisfaction is based on content and not form

Focused on results only, the users were utterly unconcerned with underlying technology, including that of user interfaces. The degree to which they were satisfied with the systems was related only to how accurately the systems could guide them in their decisions, and to some extent, the ultimate success of those decisions.

Encapsulate defaults in a simple user interface

Users were unconcerned with the user interfaces other than that they should be immediately useful. Simple user interfaces, each which encapsulated defaults that could be manipulated through sub-menus and screens, were more effective than heavily instrumented, but possibly more efficient, user interfaces. Two examples of actual user interfaces are shown in Figure 4.

--- Insert Figure 4 about here ---

Use prototypes sparingly and only for logic validation

The level of pressure and anxiety to make decisions meant that users had little patience for system demonstrations or prototypes. This meant there were few opportunities to 'get it right', and each prototype, in a sense, counted as an opportunity. The best use of prototypes was for validating logic and accuracy rather than inspecting user interfaces or system operations.

System development standards should be a secondary consideration

27

In addition to possibly being discarded, just as the requirements for these systems are emergent, the need to maintain, reuse, or extend these systems was difficult to predict. One of the purposes of system development standards is to enable maintainability. Given the emergent and potentially short life cycle of these systems, any extra effort to enforce development standards – which should be differentiated from other good software development practices – would not have paid off.

Develop each system in a functional area and plan for linking these systems together

EKPs, and the strategic decisions presented here, are often approached in a cyclic fashion (Mintzberg 1976). This cyclic nature was reflected in the development of systems built to support those EKPs, in incremental development for each system, and also for feedback loops between systems (this is partially captured in Figure 2). Developing links between systems, as shown in Figure 3, provided a form of integration for the decision-making processes, and to an extent enabled systems to be more loosely coupled and more easily modified.

Plan for system outputs that can be printed or otherwise incorporated into documents and presentations

The outputs of these systems needed to be communicated and presented, especially at quarterly meetings. Systems needed to produce output that could be readily formatted for printing, or transfer to other documents, for this purpose. Systems were more effective producing reports, graphs, or charts then intermediate data files that could be subsequently manipulated.

Expect to proceed with incomplete requirements

Supporting emerging processes meant, by definition, that a complete set of requirements could not exist prior to the start of system development. Attempting to gather even a reasonably complete set of requirements prior to actual system development would have meant that no systems would have been built.

Develop each system incrementally with a small number of functional increments

It was better to build a small functional increment, even based on incomplete requirements, than to build prototypes. This was true because of the unpredictable nature of the emerging system requirements, and because of the limited number of opportunities for presenting users with new or revised systems. The set of functional increments formed a base for subsequent system development, and also for linkages between these systems.

Componentization is not important

Componentization is a standard practice in many system development efforts. While componentization would have enhanced potential reuse for Venus Chocolates, it was far less viable because the frequency of component reuse was unpredictable. As EKPs, these systems may have been significantly revised or scrapped altogether on short notice. An up-front design effort to componentize for reuse would not have been well justified.

<div align="center">

Comparison to Existing EKP Design Theory

</div>

Before proposing any new design theory from this case study, related design theory that already exists needs to be examined. The closest existing design theory is that for EKPs, which was developed by Markus, Majchrzak and Gasser. They also examined existing design theory, specifically for semi-structured decision-making processes, and found a

high enough degree of inapplicability to warrant proposing a new design theory for EKPs.

Markus, Majchrzak, and Gasser formulated six design theory principles for systems to support EKPs. The applicability of each of these design theory principles to the case study presented here varied considerably - some could not have been applied to Venus Chocolates, while some were consistent. This prior design theory differed significantly from that proposed here in several ways:

- One prior design principle called for the conceptualization of "each user-system interaction as a customer engagement process", and that "naïve users should be repeatedly sought out through a process of 'onion-layering' the design team." In the case of Venus Chocolates, there was only a single core of users, which was the evolving management team, and there were no users that could be considered as naïve, either about systems or their functional area. This design principle would not have applied at Venus Chocolates.

- Another design principle stated that "developers should expect to need many functional prototypes," and use "radical iterations." This is at odds with the requirement for strategic decision systems to produce rapid results. For strategic decision systems, user interfaces are of relatively little significance; but the logic, models, data, and system interfaces are critical. A limited number of prototypes or functional increments was acceptable to Venus Chocolates' management, but they were extremely anxious for finished software. In addition to being time consuming, a stream of prototypes would have possibly eroded Venus Chocolates' confidence in the systems.

30

- "Radical componentization" and componentizing "everything, including the knowledge base" were part of a design principle, intended to create systems that were extremely flexible. Of course developing with components leads to flexibility and facilitates software reuse. But if components are to be effective, they need to be carefully designed upfront. Not only does this take more time, but the emergent nature of an EKP such as strategic decisions means that the design may not be known at the point the components are needed. While it is certainly true that componentization should be used for EKP systems, and good software engineering practices will aid reuse, it is infeasible to expect to componentize much, let alone everything.

But as might be expected with a prior design theory for EKPs, there was considerable consistency on several design principles. Some of those called for systems requiring little or no training, designed for offline action, and using a dialectical rather than consensus approach to design. However, the degree to which several of the prior design principles had significant incompatibilities with those developed from the Venus Chocolates case led us to conclude that a new and different design theory was justified for this class of problems.

Evaluation Metrics for the Proposed Design Theory

For any design theory, metrics for evaluation need to be developed that are not tied to the design theory itself (Hevner, March and Park 2004). One possible metric is the success of the systems developed using that design theory. In the case study presented here, a design theory was not explicitly tested, but was developed from observing practices and artifacts. One basis for assessing the efficacy of this proposed design theory is the success

31

of the systems that were produced with that design theory. Venus Chocolates utilized these systems extensively for strategic decisions, starting from an evaluation of potential market size, through store rollouts, and through infrastructure transformations. They used these systems cyclically, reformulating decisions and creating new or modified systems on an ongoing basis. The systems developed were successful, providing some validation for the design theory proposed by this research.

DISCUSSION AND CONCLUSIONS

This research presents the results of analyzing a case study in order to develop IS design theory for systems which an EKP, specifically that of strategic decision-making. Design theory that is accepted in practice, and design theory that has been proposed, are examined. Past EKP-oriented design theory is found not to be completely applicable to this case.

A weakness of a single case study is that the generalization of findings can be difficult. These results, however, are applicable to international companies seeking to open retail stores in U.S. – and beyond this also. Many of the strategic decisions faced by the company in this case study are typical of all cross-channel efforts, which are increasingly common in today's Internet-enabled economy. The strategic decisions from this case are representative of strategic decisions stimulated by new opportunities, and we believe the design theory proposed here may be applicable to many strategic decision support systems.

However, as a first iteration of an IS design theory, based on a single case, we are not able to determine the limitations on the range of systems to which this design theory could be relevant. Future research could be directed at testing the design theory

32

developed here, and determining how well it could be applied to other systems supporting strategic decisions. Such research could help find the range of strategic decisions to which this design theory could be applied.

Finally, we were struck by the degree to which the observations from this case did not match existing design theory, even for design theory for systems supporting EKPs. However, that design theory arose from experience with a system supporting organizational design, not strategic decision-making. We were intrigued that while both strategic decisions and organization design are EKPs, applicable design theories had considerable variation.

Based on this, we speculate that perhaps there is simply not one overarching design theory for all systems supporting EKPs. There may be a need for multiple design theories depending on the characteristics of a particular EKP. We believe a productive area for future research would be to determine which characteristics of EKPs would drive the need for divergent design theories, as well as what classes of EKPs would be formed based on those characteristics. We also believe that further development of the design theory presented here, and the development of potentially new design theory for any related classes of systems to support EKPs, would be an interesting and productive area for future research.

TABLES AND FIGURES

Table 1: Design Theory Components for Strategic Decision Systems		
Kernel Theory	**System Requirements**	**Design Principles**
Successful strategic decision makers use more, not less, information	Users will be managers and functional area experts	User satisfaction is based on content and not form
Rational procedures lead to more effective strategic decisions		Encapsulate defaults in a simple user interface
Faster decision-making which evaluates many possible alternatives leads to better performance	Accuracy and reliability are paramount; system performance is not	Use prototypes sparingly and only for logic validation
		System development standards should be a secondary consideration
Integrated strategic decisions lead to faster decision-making	Systems will need to involve several functional areas	Develop each system in a functional area and plan for linking these systems together
		Plan for system outputs that can be printed or otherwise incorporated into documents and presentations
Strategic decisions are iterative and can be linked through nesting, snowballing, or recurrence	Systems will need to accommodate the emergent nature of strategic decisions	Expect to proceed with incomplete requirements
		Develop each system incrementally with a small number of functional increments
		Componentization is not important

34

Table 2: Summary of Case Evidence Supporting Design Principles

System Requirement	Discovered Design Principle	Case Evidence Supporting the Discovery of Principle
Users: managers, functional area experts	Content, not form as basis for user satisfaction	• short snippet • short snippet
	Simple user interface encapsulating defaults	• short snippet • short snippet
Accuracy, reliability trump system performance	Use prototypes sparingly and only for logic validation	• short snippet • short snippet
	System development standards should be a secondary consideration	• short snippet • short snippet
Involve multiple functional areas	Development and later linking of individual functional systems	• short snippet • short snippet
	Incorporation of printed system outputs into documents and presentations	• short snippet • short snippet
Accommodate the emergent nature of strategic decisions	Proceeding with incomplete requirements	• short snippet • short snippet
	Incremental system development, small number of functional increments	• short snippet • short snippet
	Componentization not important	• short snippet • short snippet
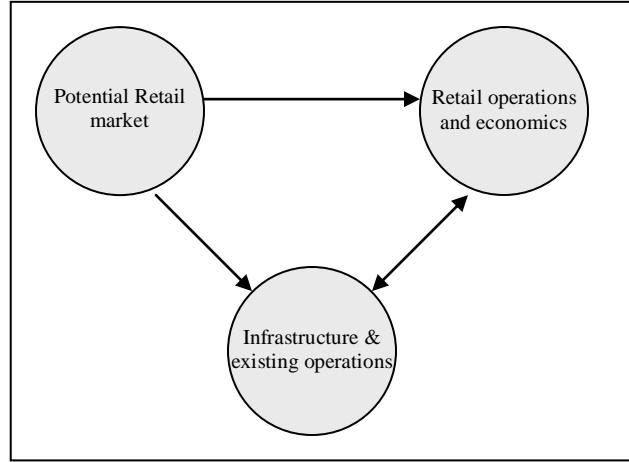
Figure 1: Conceptual Interrelation of Strategic Decisions
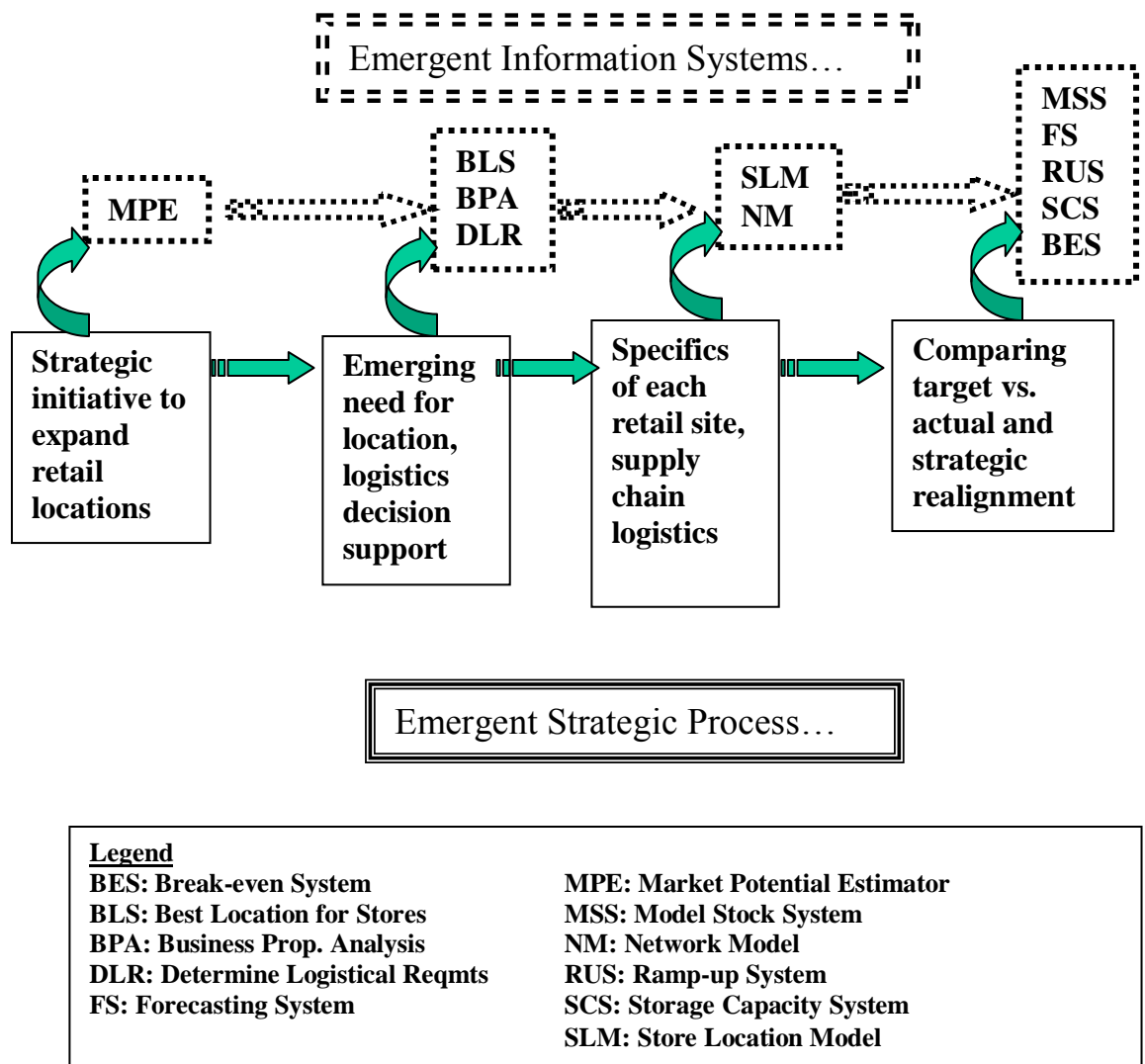
Figure 2: The Emergent Character of Strategic and IS Process
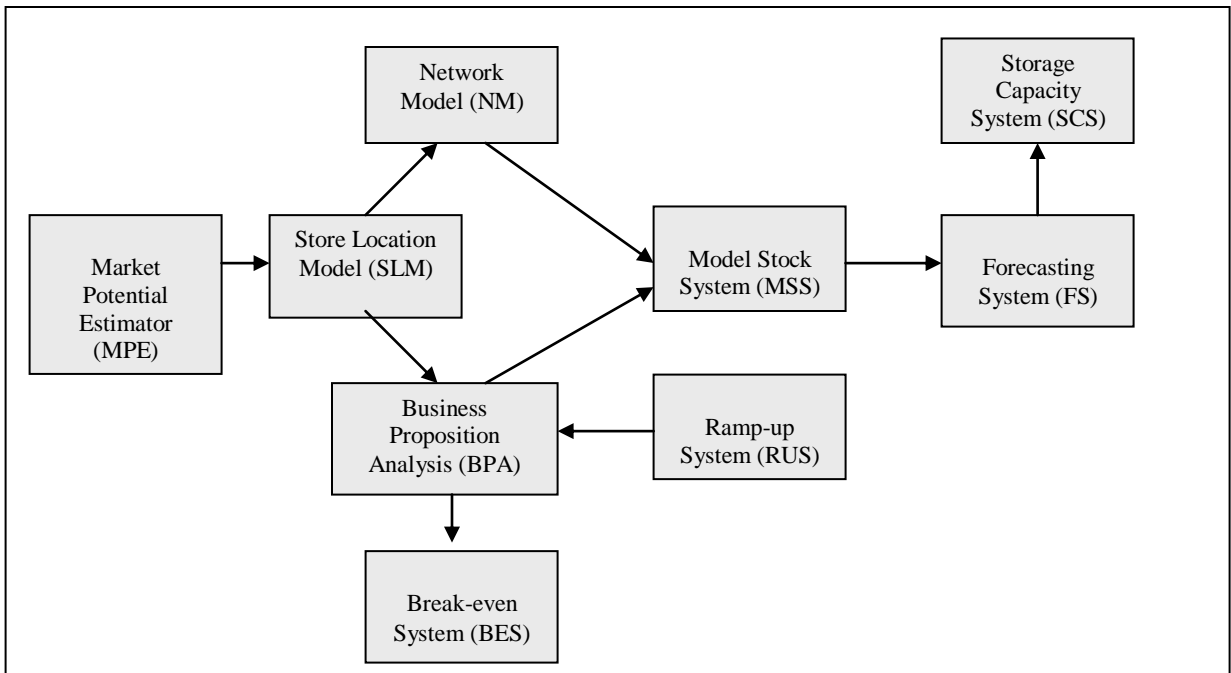
Figure 3: System Linkages
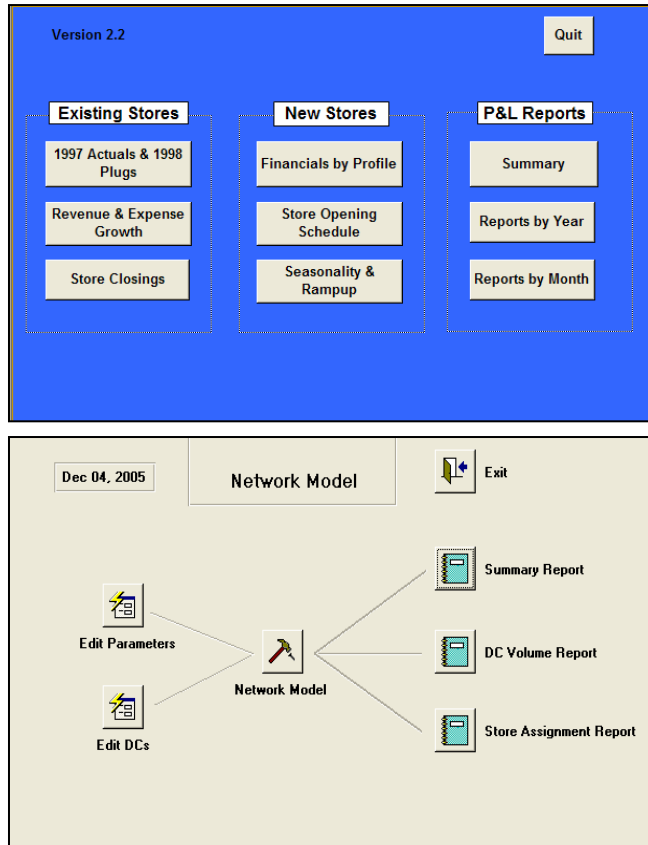
Figure 4: Sample User Interfaces from Venus Chocolates systems

REFERENCES

Benbasat, I, Goldstein, D., and Mead, M. "The case research strategy in studies of information systems", *MIS Quarterly* (11:3), September 1987, pp. 368-386.

Bennett, R. "The importance of tacit knowledge in strategic deliberations and decisions", *Management Decision* (36:9), 1998, pp. 589-597.

Bourgeois, L. and Eisenhardt, K. "Strategic decision process in high velocity environments: Four cases in the microcomputer industry", *Management Science*, (34),

Computerworld "Get the Picture", news story by Connie Winkler, January 9, 2006, accessed February 13, 2006, <http://www.computerworld.comprintthis/2006/0,4814,107502,00.html >

Dean, J. and Sharfman, M. "Does decision process matter? A study of strategic decision making effectiveness", *Academy of Management Journal*, (39:2), 1996, pp 368-396.

Dutton, J. and Jackson, S. "Categorizing strategic issues: Links to organizational action", *Academy of Management Review* (12:1), 1987, pp. 77-90.

Eisenhardt, K. "Making fast strategic decisions in high-velocity environments", *Academy of Management Journal*, (32:3), 1989, pp. 543-576.

Eisenhardt, K. and Martin, J. "Dynamic capabilities: What are they?", *Strategic Management Journal*, (12), 2000, pp 1105-1121.

Gielens, K. and Dekimpe, M. "Do international entry decisions of retail chains matter in the long run?", *International Journal of Research in Marketing*, (18), 2001, pp. 235-259.

Guba, E.G. and Lincoln, Y.S. "Competing paradigms in qualitative research" in *Handbook of Qualitative Research*, N.K. Denzin and Y.S. Lincoln (eds.), Sage, Thousand Oaks, CA, 1994, pp. 105-117.

Hanseth, O. and Lyytinen, K., "Theorizing about the Design of Information Infrastructures: Design Kernel Theories and Principles" *Sprouts: Working Papers on Information Environments, Systems and Organizations*, (4:4), (2004), pp 207-241. http://sprouts.case.edu/2004/040412.pdf

Hevner, A., March, S., Park, J., and Ram, S. "Design science in information systems research", *MIS Quarterly* (28:1), March 2004, pp. 75-105.

Kvassov, V. "Strategic decisions and intelligent tools", *Proceedings of the 33rd Hawaii International Conference on System Sciences*, Hawaii, January 2000, IEEE, p. 2022.

IBM "IBM Software – Rational Method Composer – Rational Unified Process", product information from IBM software, accessed March 15, 2006, <http://www-306.ibm.com/software/awdtools/rup/>

Langley, A., Mintzberg, H., Pitcher, P., Posada, E., Saint-Macary, J. "Opening up decision making: The view from the black stool", Organization Science, (6:3), 1995, pp. 260-279.

Markus, M. L., Majchrzak, A., and Gasser, L. "A design theory for systems that support emergent knowledge processes", *MIS Quarterly* (26:3), September 2002, pp. 179-212.

Meyers, M. "Investigating information systems with ethnographic research", *Communications of the Association for Information Systems* (2:23), December 1999, pp. 1-20.

Mintzberg, H. "Patterns in strategy formulation", *Management Science* (24:9), May 1978, pp. 934-948.

Mintzberg, H., Raisinghani, D., and Theoret, A. "The structure of 'unstructured' decision processes", *Administrative Science Quarterly* (21:2), June 1976, pp. 246-275.

O'Brien, J., Remenyi, D., and Keaney, A. "Histiography – A neglected method in business and management studies", *Electronic Journal of Business Research Methods*, (2:2), 2004, pp 135-144.

Orlikowski, W. and Baroudi, J. "Studying information technology in organizations: Research approaches and assumptions", *Information Systems Research* (2:1), March 1991, pp. 1-28.

Rittel, H. and Webber, M. "Planning problems are wicked problems", <u>Developments in Design Methodology</u>, John Wiley & Sons, New York, 1984.

Tillquist, J., King, J., and Woo, C. "A representation scheme for analyzing information technology and organization dependency", *MIS Quarterly* (26:2), June 2002, pp. 91-118.

Trochim, W. "The research methods knowledge base", 2nd Edition, Internet WWW page, at URL <http://trochim.human.cornell/edu/kb.index.htm>, revised Jan 16, 2005, accessed March 23, 2006.

Truex, D., Baskerville, R., and Klein, H. "Growing systems in emergent organizations", *Communications of the ACM* (42:8), August 1999, pp. 117-123.

Walls, J., Widmeyer, G., and El Sawy, O. "Building an information system design theory for vigilant EIS", *Information Systems Research* (3:1), March, 1992, pp. 36-59.

Weber, R. "Still desperately seeking the IT artifact", Editor's Comments, *MIS Quarterly* (27:2), June 2003, pp. iii-xi.

Yin, R. *Case Study Research, Design and Methods*, Sage Publications, Thousand Oaks, CA, 1984.

Globalization 5.0: Brining Computer Literacy Curriculum into the 21st Century

Donna Schaeffer

Cynthia L. Knott

Michelle Liu

Northeast Decision Science Institute (NEDSI)

Montreal, Canada

April 14th-16th, 2011

Abstract

We are entering a time that can be called Globalization 5.0, where size doesn't matter.

In *The World is Flat,* Tom Friedman describes Globalization 1.0, which began around 1492, when the world went from size large to size medium. The growth of multinational companies, Globalization 2.0, the world went from size medium to size small. And then around 2000 came Globalization 3.0, in which the world went from being small to tiny. Friedman took his ideas to Globalization 4.0, where the balance of power is changing

The year 2010 and beyond has taken us to a place where technology is so ubiquitous that time and space almost don't exist, and the size of the world is irrelevant. In this context, computer literacy curriculum needs to be experiential, interactive, and relevant.

In this paper, we describe Marymount University's recent efforts to globalize its curriculum. The University requires exposure to global perspectives through classes or study abroad. There is a movement towards "Inquiry-guided learning" where the students learn *how to* learn. These elements were designed into a Global Perspectives in Computing class for freshmen in the School of Business.

Introduction

The emerging Globalization era poses unique challenges for business school faculty, especially those teaching technology and operations management. The dynamic and evolving nature of the disciplines along with economic, cultural, and technological changes fueled by Globalization obliges us obliged to keep abreast with those rapid changes and teach the subject with the tools that enable our students to learn best. In developing a computing course that addresses market needs and increases students' competencies, we needed to better understand the traditional freshmen. The current generation of digital natives will become the workforce and compete with their peers in a global village. We were interested to find their levels of technical skills and their views of globalization. We examined their knowledge of technology and the ways they learn and think using visual and audio technologies. We were surprised by the lack of any depth to their understanding of how technology works and how it may be used innovatively in a business context.

What is Globalization 5.0

In *The World is Flat,* Tom Friedman describes Globalization 1.0, which began around 1492, when the world went from size large to size medium. Civilizations have traded with one another since ancient times, and for millennia, international trade was accomplished via physical products being transported across the oceans.

Friedman describes the growth of multinational companies during the middle part of the 1900s as Globalization 2.0. Then, the world went from size medium to size small.

Technologies helped organizations in various parts of the world become business partners.

Globalization 3.0 came around 2000 and the world went from being small to tiny. Friedman recognizes that trade became harder post-9/11, but the economic growth of India, China, Brazil, and Russia created a robust international economy. Globalization 3.0 is characterized by the diversity of the participants.

Economists around the world have extended Friedman's ideas to Globalization 4.0, where the balance of power is changing. In Globalization 4.0, the "state" is an active player in the economy in a range of different ways (DeVos, 2010). In a speech given to the Foreign Press Association in 2010, Daniel J. Brutto, President of UPS International, described how technologies have changed the ways shipments are handled and provided a transparency. Organizations of all sizes and in any locale can now offer their products competitively.

The year 2010 and beyond has taken us to a place where technology is so ubiquitous that time and space almost don't exist, and the size of the world is irrelevant. We have entered, and more importantly, our students will be working and living in an era of Globalization 5.0.

How Globalization is Realized in Marymount's General Education Curriculum

The following is a list of the University requirements. This is part of the core curriculum and is mandatory for all students at Marymount University.

University Requirements

Global Perspective

GP — One designated course, study abroad, or foreign language course is required.

Writing Across the Curriculum

WI — Three designated writing-intensive courses are required, in addition to the Liberal Arts Core requirement of the written communication sequence.

Ethics Across the Curriculum

ETH — One designated course is required, in addition to the Liberal Arts Core requirement of a moral principles course.

Experiential Learning

EXP — An internship or research experience is required. Most students meet this requirement through the internship in their majors.

Inquiry Learning

INQ — Three courses in the major are designated as inquiry-guided learning courses. In addition, DSC 101 DISCOVER First-Year Seminar is an inquiry course. (http://www.marymount.edu/academics/core/requirements.aspx)

From the list, it is clear that the changes made to the undergraduate IT course, Information Technology in a Global Environment, are critical. Inquiry learning is also a key element in the curriculum and has also been incorporated in the syllabus as well.

All programs are looking to add a course with this "global" focus. Once a course shifts to meet these requirements, it is intended that any student enrolled at the University can take a course outside of their majors to meet this standard. Given we have only implemented this course in the last year, we have only seen a few students from other

majors, however, we expect this to increase in the future. Although traditionally only

Bachelor of Business Administration (BBA) students took the introductory IT course,

with this new requirement, we hope to get a variety of students that can add a different

perspective to the discussion. This will be beneficial to both the BBA students as well as

the others from the School of Arts & Sciences, the School of Education and Human

Services and the School of Health Professionals.

Why/How We Designed IT 110 to be a Global Course

The traditional model for an introductory IT course was to focus on definitions

about systems analysis and design. The new course incorporates current trends and

information that span across societies and environments. The shift was critical to prepare

our students for the fast pace and changing world that has less and less defined borders.

Technology used to be isolated to each business entity. Now however, technology

focuses on how these entities are integrated using the new methods of networking and

computing.

As previously reported in the paper by Knott & Schaeffer, *Globalization--How Has it*

Changed Teaching Strategies?, to be published in the Spring of 2011, the following is a

description of the course, the objectives and broad purpose of the IT 110 course:

COURSE DESCRIPTION

Introduces students to the role of information technology in today's global business,
political, and government environments, and in society in general. It examines the
role of technology globally, particularly as it is used for cultural awareness, business
development, political change, and social improvement. The negative aspects of
technology (e.g., dumping of end-of-life hardware in developing countries) are also
discussed. The course studies infrastructure (hardware and software, networks, the

Internet), communications, software and website development, databases, and information security and privacy.

COURSE OBJECTIVES

Upon successful completion of this course, students will be expected to:

a. Analyze information systems based on the idea that information is a critical resource to the organization that in turn influences the management of other resources;
b. Document the concepts of information within the organization i.e. what questions should be asked to collect information, why and with what priority.
c. Establish how information systems fit within an organization to support its processes and division of work;
d. Explain the essentials of information systems and subsystems in terms of purposes, processes, boundaries, and value;
e. Identify facts or ideas on the importance of advances in technology in the design and development of new information systems;
f. Summarize the role of people in the operation of information systems.
g. Plan how information systems can be developed to accomplish organizational objectives; and
h. Discuss the ethical concerns associated with information systems including security and privacy.

BROAD PURPOSE OF COURSE

The broad purpose of the course is to explore the role of computers in this global age and understand how computers are used in business and in the global society. The basic tenet is that information is a critical resource and that using technology effectively to communicate and to disseminate information can influence many aspects of the life of citizens throughout the world. For example, students should become more aware of the importance of social media and the ethical implications of the use of technology.

During the transition from the old model to the new model, the course was entitled, ISY 110/301 Foundations of Information Systems. The following are the previous broad purpose of the course and the course objectives:

BROAD PURPOSE OF COURSE

The purpose of this course is to provide the foundation for information systems, look to future implementation and applications of information systems, and develop and present projects.

Information systems are to achieve the goals of the organization is not, of course, something brand new; it has been required since time immemorial. But within the past two decades we have seen an extreme change in the importance of information and the quantity of information that is needed by organizations. And all of this has been facilitated by the amazing advances in information technology. This course will prepare you for terminology, understanding of concepts, and provide creative ideas for how to harness information systems in the future.

COURSE OBJECTIVES

Upon successful completion of this course, students will be expected to understand:

- That information systems are based on the idea that information is a critical resource to the organization that in turn influences the management of other resources;
- The concepts of information within the organization i.e. to understand what questions are being or might be asked, why and with what priority.
- How information systems fit within the organization and support its processes and division of work.
- The essentials of information systems and subsystems in terms of purposes, processes, boundaries, and value.
- The importance of advances in technology in the design and development of new information systems.
- The role of people in the operation of information systems.

- How information systems are planned and developed to accomplish organizational objectives.
- The ethical concerns with information systems including security and privacy.

When you compare the current and the past purpose and objectives, the shift from merely covering information technology in general, has been expanded to include a global component that is incorporated into each aspect of the foundational skills.

Not only has the introductory course shifted, other core BBA courses have incorporated technology as an integral part of the objectives. Technology is now a reality and has changed each core area in business. For example, Accountancy still teaches students how to use t-charts for accounting entries, but now, due to technology, the field has moved to using spreadsheets, data analysis and statistical procedures that are now available using technology. Another example is Operations Management, which incorporates technology by analyzing data using simulations and modeling. These techniques are crucial for any business to gain and maintain a global competitive advantage in today's economy.

Specific IT 110 Assignments

The key to making the course successful is to not only teach the fundamentals of information technology but to have the students do assignments that incorporate these with respect to a global focus. The main deliverable for the course is the final group project. Smaller assignments are given throughout the semester that build up to developing a country paper and presentation that incorporates all of these things into one project. These smaller assignments include things like creating a blog, using social

networking to meet other people around the globe and creating a multimedia presentation

that incorporates audio, visual pictures and video to share with the entire class.

The group project assignment is the following:

Group Project

Working groups of 3 to 5 team members, complete the following project.

You are an intern in the International Studies Office at Marymount University. You need to present your proposal for a study-trip to a country of your choice. One person in the group must have been communicating with one or more persons in the county during the semester. Your proposal must include facts, cultural information, photos, and video from the country you select. You can present your proposal as a multi-media PowerPoint presentation or make a video which you upload to the Marymount Channel on YouTube. Guest speaker will be invited to provide detailed instructions on how to do multi-media PowerPoint presentations or YouTube videos. The deliverables are a report and a presentation at the end of this semester.

The previous ISY110/301 courses final assignment was the following:

Research/Final Paper: The purpose of the paper is multi-fold: 1) to allow students to explore the several information tools used in business; 2) to enhance student's writing ability; 3) to develop student's research ability; 4) to support student's oral presentation skill.

The student teams will propose a new technology solution for the company. In order to make a proposal for technology solution, student must understand the company, the company business, and the company customer. This is the intent of the research.

The researched company must have an international presence. That is, the company must be in the US and in at least one country other than the US.

The research paper will use all tools explored in the laboratory portion of this course as well as material learned from the lecture part of the course.

MS Word: Student teams will produce the research paper using MS Word. A table of contents will be generated using MS Word as well as bibliography and appendices if appropriate.

MS Excel: Student(s) will capture twelve months stock market activity for the researched company. These stock market numbers are entered into an MS Excel spreadsheet. Student teams will create columns to calculate gain/loss demonstrating change in color. Students will graph market changes. Both of these Excel outputs are inserted into the paper where appropriate.

In addition, student teams will create a spreadsheet identifying and itemizing the cost of the technological change proposed. This spreadsheet is inserted in to the paper where appropriate.

MS Access: Student(s) will create an Access database with twenty names of vendors for products proposed for the technological change. . The names can be fictitious. The database created by students will have at a minimum the following: first name of contact person, last name of contact person, company name, company address, business unit of contact person, telephone contact numbers (this may include business, cell, and fax), products offered, and product pricing. Student is welcome to create additional fields as deemed appropriate. This table is inserted into the paper at an appropriate point. The final database will demonstrate ability to search for contact person by first name, last name, company, and one other field.

MS Powerpoint: Students will create a final presentation using MS Powerpoint. The presentation is presented the last day of class to all classmates. (Schedule will be determined at the later date.) All team members are expected to have an active part in the oral presentation. Any one of the team members is expected to have sufficient knowledge to answer questions during the question and answer portion of the final presentation. Powerpoint presentations are posted to Blackboard Discussions area for downloading on day of presentation. It is assumed presentations will span two days. If this occurs, all presentations are due at the same time.

MS Project (time allowing): Student(s) will develop a project timeline for completing implementation of technological change. The technologic change will have a timeline indicating resources needed and scheduling of tasks and milestone.

From the two course project descriptions, it is clear that the previous course

focused on using specific software applications. Students were required to implement the

other areas of study into these applications to analyze a specific company. As mentioned

before, these skills are now incorporated into the classes that discuss each topic, such as Project Management, Accounting, Marketing and Finance. Since these skills are now covered in other core courses in the BBA, it has allowed for the IT course to expand the focus to not only specific software applications, but on a broader use of technology with respect to a given country and how that country is progressing with respect to the rest of the world.

Conclusions--Best Practices and Lessons Learned

After teaching this global perspective course for the first time to several sections of freshmen, we are able to reflect on our experiences, analyze students' feedback, and synthesize best practices and lessons we have learned.

In summary, we had three findings:

First, in order to accomplish a large group project assignment, each team was required to create, update, and maintain its own blog during the whole semester. Most of the students had difficulties understanding how blogs work and how to create a blog. As one of the most powerful Web 2.0 technologies, blogs have been used as a significant way of commentary for both organizations and individuals. However, our students did not blog or did not even know what blog is. An interesting observation is that although the majority of the business freshmen use Facebook, they are new to blogs.

Second, we need to emphasize critical thinking and real-world problem solving during the early stage of their four-year college education. Most of students lacked critical thinking skills and the ability to solve problems independently. Ironically, the first

source they turn to, no matter whether it is discussion question or case study question, is Google. If they cannot find the links directly from Google search results, the students would immediately gave up instead of thinking through the problems by themselves or trying different search techniques.

One sub-dimension question derived from this problem could be described as how to teach our student to find the relevant information efficiently as well as effectively by using a search engine. We need to impart the idea to the students that doing online research is not the same as, and is much more demanding and challenging, than entering a keyword in Google search box and accepting all the information returning by the search engine. Instead, online research needs systematic techniques and involves decision making processes. How to use focus words to narrow an online search? How to weigh primary source and secondary sources? How to create queries using the logical operators? How the vertical searching is different from the horizontal searching?

Once you locate the information, can you trust all the sources? In summary, we need to bring the above questions to the students' attentions and teach them those skills. The core idea is that students need not only search engines but also their own intelligence and judgment to locate the relevant information effectively.

Third, the students lack soft skills such as collaborating with team members, communicating with people from other cultures, and encompassing different opinions and views. We believe we need to impart the importance of team spirit to our students.

Finally but not least important, faculty members need to keep learning and updating their own knowledge base. This poses challenges to us as professors teaching information technology. We have obligations to know the dynamic and evolving subject and to teach our disciplines through the tools by which our students are best able to learn.

In the future, after we have taught the course for several semesters, we will analyze the evaluations from students to determine what their experiences were in a globally focused course. We also plan on implementing a separate survey to specifically address the changes in the course and the feedback about the content. As we move further into the 21st centuries, we would expect to see more and more courses incorporating this new "global" environment. We plan on analyzing the requirements of other core courses to determine if they follow suit with this focus.

References

Brutto, Daniel J. "Globalization 4.0 and the New Logistics." Speech given to Foreign Press Association, New York, September 22, 2010. Available via http://www.pressroom.ups.com/About+UPS/UPS+Leadership/Speeches/Daniel+J.+Brutto/Globalization+4.0+and+the+New+Logistics

De Vos, Marc. *After the Meltdown – The Future of Capitalism and Globalization in the Age of the Twin Crises.* Shoehorn Books, 2010.

Friedman, Thomas. *The World Is Flat.* Farrar, Straus and Giroux, 2005.

http://www.marymount.edu/academics/core/requirements.aspx

Syllabus for IT 110, *Information Technology in the Global Age*, Knott, Fall 2010.

Syllabus for ISY 110/301, *Foundations of Information Systems*, Knott, Spring 2007.

EXPLORING THE OPEN SOURCE SOFTWARE DEVELOPMENT METHOLOGY AS BARRIER TO ADOPTION

Joseph Nwankpa
Dept. of Management & Information System,
Kent State University,
Kent, Ohio
jnwankpa@kent.edu

Yaman Roumani
Dept. of Management & Information System,
Kent State University,
Kent, Ohio,
yroumani@kent.edu

ABSTRACT

The open source development methodology was seen by many as the development paradigm that would reshape the software industry as it boasted unique and unconventional characteristics that traditional development methodology could not match. However, despite all its potentials, the adoption of open source software at the end-user level remains moderate or at least failed to keep pace with the initial expectations. The aim of the paper is to identify and discuss major challenges for open source software, so as to facilitate end-users adoption. The paper argues that nature of OSS development methodology encourages the elimination of key activities during the design and development phase thus creating implicit adoption barriers.

Keywords: Open Source Software, Proprietary, Adoption

INTRODUCTION

A few years ago open source software (OSS) was seen as the panacea to proprietary dominance in the software industry. In fact, at a time the OSS model was thought by many as the model that would reshape the software industry as it boasted unique and unconventional characteristics that traditional software development model could not match. It was anticipated that open source software would genuinely challenge proprietary software market share. This new software development paradigm consisted of community of volunteers working together and collaborating to develop an end-product at no cost [5, 9].The attention and interest in OSS lead to bold predictions at a time that OSS development paradigm would transform the software industry. OSS boasted unique characteristics such as reduced cost, rapid development cycles, high reusability and freedom to choose [7] compared to those of proprietary software. OSS seemingly

became the solution for companies seeking to optimize their IT budget while re-examining the return on investment. This new philosophy was heralded at a time as the gateway to adoption for companies looking to save huge licensing cost. In fact, earlier studies on OSS attempted to understand the motive and incentives behind community of developers who worked selflessly across geographical boundaries to develop free end-products [4] while some other studies examined the applicability of open source phenomenon outside the software domain [13].

However, despite all these potentials, the acceptance of open source software at the end-user level has remained moderate or at least failed to keep pace with initial expectations. At the user end, open source software continues to struggle to attract user-base. Indeed, some studies have examined this worrisome trend. On the one hand, proponents of open source model are quick to blame the big corporations and their proprietary walls as barriers to greater utilization of OSS. On the other hand, some argue that poor user interface, improper documentation, feature-centric development and programming for self are potential factors that hinder open source acceptance to the general public [12]. For instance, it has been noted that lack of apparent ongoing support for OSS project constitute significant barrier to adoption as users are increasingly apprehensive of the level of support for these products [9]. Nonetheless, with many companies such as Red hat providing support services for key OSS application, the level of adoption still remains slow. The severity of this is argument is brought to bear when we factor in the fact that OSS are suppose to be high quality, cheaper and more flexible when compared to proprietary applications.

This paper seeks to provide additional insight by examining the developmental methodology and processes of OSS. We argue that the developmental structure of OSS negates one of the fundamental processes in system development thus, leading to final products that may not represent the needs of mainstream end-users or greater user community. The central premise is that the duality role in OSS development where developers are also users contradicts the requirement analysis phase in system development. The resultant effect is developers who are technology savvy will rely on their software needs as a representation of mainstream user needs. Consequently, software developed from this process may struggle to meet the demands of users thus creating a huge barrier to adoption.

OPEN SOURCE PROJECT

Open source projects are typically initiated by individuals or small groups with an idea for something that is interesting to them. The common goal of an open source project is to create software that is useful or interesting to those who are working on it rather than to fill a commercial void [2]. Conversely, the open source development project may indeed be based on what developers find interesting rather than what is essential to wide-ranging users [14]. The project developers and designers are also the potential users of the software hence these projects are regularly driven by a need to use basis. Usually, the project initiators generally become the owners of the project. Open source projects typically engage in no active recruiting beyond simply posting their intended goals and access address on a general public website such as Sourceforge and Freshmeat. As a result, many open source projects have been created because some developers felt the existing project could not satisfy their software needs.

Due to the community approach to open source projects, it is inevitably that a gap will exist between community of developers who are typically technology savvy and non-developers who may not contribute codes, but are in need of software that provides a specific service [15]. If this is the case the question becomes to what degree can the software developed within the open source community accommodate the general users need? Many projects end up with software projects that satisfy the developers need, yet are so far from the need of programmers outside the project let alone the general users. This has created a scenario where some open source projects lack the depth to attract any meaningful user outside the domain of the project developers. Hence, creating a situation where open source projects continue to witness huge growth yet limited used and acceptance.

The quality assurance component of open source is predicated on developers working together as a community and making positive contribution. In his paper "The Cathedral and the Bazaar" Raymond argues that high level of quality demonstrated by open source software is partly due to the high degree of peer review and user involvement. Open source software claims methodological superiority over proprietary model because of its ability to attract programmer from across the globe [8]. Yet arguably quality can also be view from product usability and how the product fits the task that it is built to perform. The user centered design of OSS makes it difficult to articulate the quality dimension or at least to compare OSS to proprietary software that is design with main stream users in mind.

Technology Adoption

Technology adoption has been widely investigated within the Information System literature. For instance, the technology acceptance model (TAM) introduced by Davis has been widely applied to understand the attitude one has about the use of technology and subsequently, used to predict the adoption and use of information technology [1]. TAM suggests that two particular beliefs, perceived usefulness and perceived ease of use influences user's computer acceptance behavior. Perceived usefulness is defined as the prospective user's probability that using a particular system would enhance his or her job performance. Perceived ease of use refers to the degree to a prospective user believes that using a particular system would be free from effort [1].

Perceived usefulness of a project within the open source community can be measured by the size and duration of the project. This is usually the case as more interesting projects with higher potential will attract more developers and are usually more sustainable. Moreover, a study by Krishnamurthy found that the number of developers associated with a project was positively correlated to the age of the project [6]. The study also found that projects with more developers were viewed and downloaded more often. However, perceived ease of use is more complicated and difficult to attain using the open source model. One major shortcomings of open source software is its inability to create user friendly platform. Arguably, this is because open source model was designed to serve users who are computer savvy hence did not required an elaborate user friendly design. However, as open source software to moves toward mainstream users; these concerns become evident limiting open source ability to compete favorability with proprietary software. To mitigate these concerns, some companies such as Red hat, provide support services that makes it easier to install, use and maintain these software licensed under the GPL. However, this business model is only possible if critical mass and some level of user threshold can be achieved.

Open Source Development Approach

Open source development processes are fundamentally different from traditional software development in a variety of ways. However, one that clearly stands out is the requirement analysis aspect of the system needed to be developed. Conventional system development methodology outlines and encourages active interaction and feedback between users and system analysts in a bid to ensure that system meets end-users needs. In fact, new system development methodologies such as agile software development, rapid application development are aimed at generating feedbacks and integrating end-users with the designers and developers in an attempt to ensure that final product is consistent with users needs. It goes without saying that software companies go through series of prototype testing to elicit feedback from users in order to develop software that are consistent with end-users.

Nevertheless, for OSS software, the requirement analysis begins and ends with the developers who initiated the project and subsequent developers that will join the project mid way into development stage [15]. Hence, by so doing OSS development process omits requirement analysis a key activity in any software development. Most projects make limited use of software engineering approaches and practices such formal design procedures, specification and system prototyping [9]. The resultant effect of this process structure can be seen in two folds. On the one hand, developers are encapsulated by their pre-defined views of the software requirements which may not be able to capture the overarching requirements of member of the open source community who may not be developers but are in need of the software. On the other hand, the process ensures that developers' ability to get positive feedback from non-developer members of the community is indeed eliminated. Software application by their nature can incorporate multiple functionalities, hence developing software capable of meeting multiple users requirement is not an unattainable goal. In fact, this is why proprietary software continues to have significant advantage over OSS. While proprietary software are able to serve the need of wide variety of users thus gaining the necessary critical mass and network externality needed to gain dominance, OSS which by its nature is tailored to meet the needs of few developers working on the project find it very difficult to gain wider use and acceptance. One the most critical factors cited as a barrier to OSS adoption is lack of support. Users are quick to argue that OSS lack reliable support services that accompanies commercial software. Such concern begs the issue of who was the intended user of the software and what nature of support is need. The duality role in OSS where developers and project initiators are indeed users ensures that support is as good as the system developed. These developers understand the system configuration and may not see the need for trivial support services either embedded in the system or outside the system. In fact, the project intended users based on the OSS perspective are technology savvy thus view support services as waste of manpower time and effort.

Furthermore, the dual role of developers in the OSS paradigm does create multiple project spread with the open source community. The number of open source projects keeps multiplying as developers continue to create project that will meet their specific needs and not necessarily the needs of wider open source community. An examination of projects within software categories in open source hosting sites reveals as many as a dozen projects with little variations. For example, Sourceforge.net which is the dominant open source project hosting website lists more than150, 000 projects and more than one million registered users. These projects span the open source

community with little variable creating a confusing situation where similar projects with limited capabilities and functionalities are developed and maintained within the open source community. Hence not only does duality role affect comprehensive understanding of software requirement and feedback, but the process indirectly encourages a constellation of user requirements leading to a plethora of projects. Giving these conditions, it is difficult to see how OSS can compete with commercial software or for that matter gain wider acceptance and adoption.

Duality role and technology adoption

Technology adoption has always been based on the ability of the technology to meet the needs of potential users [3, 11]. Software providers try to understand existing void in technology usage and then design software capable of matching or to some degree filling the gaps. Hence, requirement analysis is a critical process in software development methodology as organizations and software providers attempt to ensure that product is consistent with the need. Requirement analysis in system engineering involves those process that go into determining the needs of users. It represents a description of how a system should function and a description of the system properties and attributes. The software requirement analysis process thus covers the task of modeling, analyzing these requirements as well documenting them as a basis for system design. Studies have shown that improper attention to software requirement analysis during a software development is a leading cause of software that fails to deliver the needs for which they were designed. However, the OSS development methodology creates a unique structure where the developers and end-users are indeed one and the same. This unusual situation means that developers may not require requirement analysis as a way of understanding the end-users needs since they are also the intending users. Thus the critical aspect of requirement analysis in a software development is accelerated almost to the point of total elimination in the OSS development methodology. Hence we propose:

Proposition 1: The dual role in OSS development methodology where developers are also end-users will likely reduce the amount of requirement analysis in a project.

The perceived usefulness of a technology has been identified as one of the primary antecedents to intention to adopt a technology [1]. Hence, systems and software applications are typically tailored to meet the requirement of potential users. In the case of OSS, developers' needs are represent the needs of the overall potential users since the developers are indeed motivated to design and develop a software application that will meet their current software need. However, once the software is designed and developed and licensed under open source licensing scheme, it is made available to the public who are free to download, use at limited or no cost. But the extent to which such software application will meet the demands of the wider open source community as well as users outside the open source community depends largely on the degree to which the initial developers requirement conform to those of potential users outside the project. More often than not, the needs and requirement these sophisticated technology savvy programmers are usually inconsistent with those of end-users who typically lack such level of expertise in software design and development. Therein lies the problem with OSS as potential users outside the boundaries of the project struggle to grasp how such projects can indeed be useful to them. In fact, these potential users struggle to understand these projects goals, usefulness and fit with their current system needs. This argument becomes more evident when one examines many projects spread across open source hosting websites, which were designed to meet the need and requirement of a couple of developers or project owners. Such projects are indeed dead today

because first, they do not reflect or represent the needs and requirement of users outside the boundaries of the project and secondly, they scope of the project was too narrow to support any potential user base. Hence we propose:

Proposition 2: The dual role in OSS development methodology where developers are also end-users will likely increase the number open source projects inconsistent with wider user requirement.

The very nature of software products enables multiple functions to be embedded in a single software application. In fact, commercial software providers are able to leverage this multi facet ability of software to develop applications that can accommodate the different user requirements. Thus proprietary software providers are able to widen the software scope and requirements while capturing the needs of multiple users.OSS are also rich in functionality a study by Nichols and Twidale found that usability of OSS was a key barrier to adoption and not functionality of the software [10]. The study noted that OSS development methods followed a process of continuous improvement with emphasis on functionality and not usability. However, proper usability design can only be achieved if potential end-users requirements are taking into account when developing and designing the software. Hence we propose:

Proposition 3: Having procedures in OSS development methodology that encourage the understanding of wider user requirements will increase usability and enhance adoption.

DISCUSSION AND IMPLICATION

This paper offers a new perspective on how OSS development methodology creates adoption barriers for users outside the community. We argue that the reason OSS continues to struggle with end users adoption has more to do with the development process than the usability of the software. Hence if OSS are to exploit their full potentials in term of wide spread acceptance, they need to address development procedures and processes. By ignoring requirement analysis which is a critical aspect in software engineering, open source developers and designers alienate users outside the project boundaries and community. Thus, while many OSS projects may meet the needs of the few developers and contributors of the project, the same cannot be said of potential external end-users who will grapple with the software relevance and fit with their current software needs. In addition, inadequate requirement analysis does create issues of project proliferation within the open source community. The limited requirement analysis during the design phase of the software development means that many OSS projects are just tailored to meet few developers need creating scenarios were projects are started in a bid to meet those needs not included in the existing OSS projects. In fact there are lots of projects that are not fundamentally dissimilar yet continue to exist alongside each other. The chances of these projects attracting enough developers and contributors remain questionable at best. Indeed with multiple projects, ability of the open source community to pull talents and contributors across geographical boundaries may not be sustained thus leading to abandoned projects, low quality software, reliability issues and end-users rejection of OSS.

In order to gain the full potentials of OSS in terms of widespread adoption, the open source community need encourage processes that are supportive of requirement analysis. The initial design should be more involving than it is currently as this will ensure that existing projects meet the demands of wider end-users and hence encourage adoption.

References:

1. Davis, F. Perceived *Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology.* MIS Quarterly, 1989,13(3), 319-339

2. Godfrey M.W and Tu Q. *Evolution in Open Source Software: A Case Study.* Proceedings of the International Conference on Software Maintenance (ICSM 2000) San Jose California 2000,131-142

3. Goodhue, D.L and Thompson, R.L. *Task-Technology Fit and Individual Performance*. MIS Quarterly, 1995,19(2), 213-236.

4. Hertel, G, Niedner, S and Herrmann, S. *Motivation of Software developers in Open Source Projects: an Internet-based survey contributors to the Linux Kernel* Research Policy, 32, 1159-117

5. Hippel E.V and Krogh G.V. *Open Source Software and the "Private –Collective" Innovation Model": Issue for Organization Science* Organization Science,2003,14(2) 1154-1165

6. Krishnamurthy, S. Cave *or Community? An Empirical Examination of 100 Mature Open Source Projects*. First Monday, 2002, 7(6)

7. Lerner, J and Tirole, J. *Some Simple Economics of Open Source.* Journal of Industrial Economics 50(2) 197-234

8. Michlmayr M, Hunt F and Probert D. *Quality Practices and Problems in Free Software Projects.* Proceedings of the First International Conference on Open Source Systems Genova, 11th -15th , July 2005, pp 24-28

9. Mockus A, Fielding R.T and Herbsleb J.D. *Two Case Studies of Open Source Software Development: Apache and Mozilla* ACM Transactions on Software Engineering and Methodology 2002, 11(3), 309-346

10. Nichols, D.M and Twidale, M.B. *The Usability of Open Source Software.* First Monday, 2005, 8(1)

11. Rogers, E (2003) "Diffusion of Innovations" New York Free Press

12. Vixie, P. Software Engineering, Open Sources: Voices from the Open Source Revolution", first ed. O_Reilly, 91–100.1999

13. Weber, S. The Success of Open Source Harvard University Press, Cambridge, Massachusetts 2004.

14. Wu, M.W., Lin, Y. D. *Open Source Development: An Overview*. IEEE Computer, 2001, pp. 33-38.

15. Zhao, L and Elbaum, S (2003) *Quality Assurance Under The Open Source Development Model.* The Journal of Systems and software 2003,66, 65-75

Importance of the Modularity of Web-Based Contents: Lessons from a Causal Model

Silvester IVANAJ

ICN Business School – CEREFIGE

13, rue Michel Ney, 54000 Nancy - FRANCE

E-mail: silvester.ivanaj@icn-groupe.fr

Wilfrid AZAN

Ecole de Management de Strasbourg - HuManiS

Rue de la Fonderie, 68100 – Mulhouse - France

azan@unistra.fr

The use of web-based training and education is increasingly impacting the way organizations acquire knowledge (Alavi & Leidner, 2001). This new way of learning has hidden barriers of time and space (anytime anywhere learning). Another potential advantage of e-learning is that it offers personalized and contextualized learning. This is mainly thanks to the use of the modularity of learning content which consists on dividing it into a set of small functional units (also called learning objects). The modular cutting of learning content permits better learning which relies on the capacity to decompose the complexity of the knowledge to acquire [Langlois (2002) ; Cohendet al. (2005) ; von Hippel (1990)].

In front of this new type of learning, the E-learners' behavior, is quite different compared to those of traditional learning.

In this paper we study the behavior of students who took three information technology courses at a large business school. Each course is divided into units and for each unit a pre-evaluation test is offered. In function of the results obtained from this test, a personalized route is automatically proposed to the students.

Studying the way that students behave taking these web-based courses, we propose a hypothesized causal model which includes four variables: respect of the learning architecture, individual differences in cognitive abilities, learning output and satisfaction. We tested our model using a sample of 205 students. The sample represents students with varying scholastic levels. Their ages vary from 20 to 23 years old. Their level of experience with E-learning is practically identical considering the date at which E-learning was implemented at this particular school. To validate these variables we used both exploratory and confirmatory factor analysis techniques. The goal of exploratory factor analysis is to eliminate items that present unsatisfactory psychometric qualities. Cronbach's alpha coefficient was retained as a measurement of internal consistency for this stage. Confirmatory factor analysis was used in order to confirm the factorial structure.

We found that the respect of the learning architecture proposed, leads to a quick progression of in particular e-learners who have a low level of proficiency before the learning process is started (Narayanan, Balasubramanian & Swaminathan (2009)).

Our findings should help institutions to better understand E-learners' behavior, and to improve the quality of their online courses. They also suggest a need for further research on other important constructs of E-learners behaviors.

References

Alavi, M. (1994). "Computer-mediated collaborative leaming: An empirical evaluation", *MIS Quarterly,* 18(3), p. 159-174.

Alavi, M.; Wheeler, B. ; Valacich, J. (1995). "Using IT to reengineer business education: An exploratory investigation oi collaborative telelearning". *MIS Quarterly.* 19(2), p. 293-311.

Alavi, M.. & Gallupe, B. (2000). "*The transformation of business education",* New York: Prentice Hall.

Alavi, M. ; Leidner, D. (2001). "Research commentary: Computermediated learning—A call for greater depth and breadth*", Information Systems Research.* 12 (1) March, p. 1-10.

Au, N., Ngai, E.W.T. ; Cheng, T.C.E. (2008). "Extending the Understanding of End User Information Systems Satisfaction Formation: An Equitable Needs Fulfillment Model Approach", MIS Quarterly, march, 32 (1), p. 43-66.

Baldwin, C.Y.; Clark, K.B. (2000). *Design Rules: The Power of Modularity*, Cambridge, MA: MIT Press.

Cohendet P. ; M. Diani ; C. Lerch (2005). « Stratégie modulaire dans la conception. Une interprétation en termes de communautés", *Revue Française de Gestion*, N° 158, septembre-octobre, p. 121-143.

Daft, R.L.; Lengel, R.H. (1986). "Organisational Information Requirements, Media Richness and Structural design", *Management Science*, 32 (5)5, p. 56-572.

Narayanan S., S. Balasubramanian, J. M. Swaminathan (2009). "A matter of balance : Specialization, Task Variety Individual Learning in a Software Maintenance Environment", *Management Science*, 55 (11), p. 1861-1876.

Von Hippel, E. (1990). "Task partitioning : an innovation Process Variable", *Research Policy*, (19), p. 407-418.

Experiences Engaging Undergraduate Business Students in Research Activities at a Small Liberal Arts College

James A. Griesemer, D.P.S.
Division of Business
Mount Saint Mary College
330 Powell Avenue
Newburgh, New York 12550
(845-569-3120)
grieseme@msmc.edu

ABSTRACT

Engaging undergraduate students in research activities has been identified as an innovative strategy for improving education in America's colleges and universities (Boyer). The value of engaging students in research is reflected in the *U.S. News & World Report* ranking of colleges and universities where undergraduate research and opportunities for creative expression are now distinct categories. While in science and engineering programs research participation is an important step in undergraduate student development and is highly promoted by the National Science Foundation (Stoblein and Kanet) this is not the case for business students whose involvement in research activities have been traditionally limited to graduate studies (Chang).

The author's College has as part of its mission the development of students into life-long learners. It also has long standing tradition of engaging its undergraduate natural and social science students in research activities. The College knows research activities sharpens students objectivity, enhances their ability to find and interpret relevant information, and improves their written and oral presentation skills.

This article focuses on the author's first experiences engaging undergraduate business students in research activities. The results exceeded his expectations and he is now planning to introduce research activities in his other courses.

KEYWORDS: Undergraduate business research

PRIMARY TRACK: Pedagogy

INTRODUCTION

Introducing undergraduate research activities has been identified as an innovative strategy for improving undergraduate education in the U.S. (Boyer). Though large universities have been engaging undergraduate business students in research activities for many years, many small liberal arts colleges have just begun offering such opportunities. This is because the relationship between research and the goals of a liberal arts education first need to be addressed by the institution. Unlike larger colleges and universities, small liberal arts colleges are focused mainly on teaching with individual faculty valued as generalists. With their higher teaching loads and committee assignments, faculty at small liberal arts colleges often find it difficult to do their own research and find time to mentor undergraduate students in research. Also, research facilities at many small liberal arts colleges may be also modest at best (Kierniesky). While balancing a scholarly agenda with heavy teaching commitments easily consumes faculty's available time, Elgren and Hensel state that utilizing the curriculum to better prepare undergraduates for independent research serves them well and can also prepare them to contribute to faculty's own scholarly work. The synergy between these two activities has been referred to as act of "enlightened self-interest" (Mills). Additionally, involvement in research projects can make a student's resume gain more attention from potential employers and graduate schools (Stobleing and Kanet).

APPROACH

Involving students in a class in research activities requires the instructor to modify their teaching strategy in regards to context which is dependent on the student's interests,

the student's role as data gatherer, analyzer, or presenter, and finally the instructor's role as educator, mentor and facilitator (Stoblein and Kanet). The ASHE Higher Education Report states the total experience of undergraduate research activity can be divided into three phases: beginning, middle and end. The beginning phases involves setting the objectives and goals for participation in undergraduate research, the middle phase is the actual engagement of students in research activities, and the end phase is the presentation of findings (ASHE Higher Education Report). The pedagogical sequence for involving undergraduates in research used at the University of Dayton includes five key steps: (1) planning - including designing course topic research projects, (2) coaching - including supplying students with a compact manual on what constitutes good research, (3) evaluating – including the grading rubric used for the course, (4) disseminating students' findings, and (5) assessing – collecting feedback on the research experience (Stoblein and Kanet).

STUDENT EXPERIENCE

Like most small, liberal arts colleges the author's college is working to improve the academic quality of its student body. Engaging undergraduates in research activities is seen as one way to accomplish this important goal, and doing so has been part of the curriculum for natural and social science majors for quite some time.

It was decided to introduce research activities in the Introduction to Management Science course and these experiences are from the Fall 2009 traditional semester. The course is a business elective usually taken by juniors and seniors, and sometimes has a lower enrollment. Eight students enrolled in the course and one dropped due to concerns over the needed mathematics and Excel programming skills required.

The first challenge was to obtain the students buy-in to doing research activities. This required explaining in detail what makes good research in a business context. The students had concerns about working alone so it was decided to let them work as a team. The second challenge was teaching them enough about the tools of management science such as linear programming, integer programming, mixed integer programming so they could participate in the research project selection.

At the same time potential venues for presenting the students' findings were being investigated. The Chairs of Divisions of Natural Sciences and Social Sciences at the College offered to let the students to join their Fall semester poster sessions. Another venue found was participating in a student poster session at the annual meeting of the Environmental Consortium of Mid-Hudson Colleges and Universities which was being held late in the semester at a nearby college.

The students decided to focus on the student poster session at the annual meeting. This set a must have project completion date and also helped them to focus on a research project having an environmental focus. The students next used a basic brainstorm technique to identify possible projects. Since time was short they decided to build on a challenging case study in their textbook based on the well known coal allocation model developed by Duke Energy (Anderson). Their idea was to add an environmental constraint to the model that addressed meeting gaseous sulfur emissions.

The students were helped to organize the needed research activities. Two students were assigned to solve completely the case study. The other students researched the EPA's air quality standards, how sulfur content in coal is related to gaseous emission content, and coal usage by electric generating plants in the US. At every class meeting, students reported on their progress including any difficulties they were experiencing. Rapidly the students identified where more resources were needed.

Once their research was almost done the students focused on their poster and what information it should contain. Quickly they found they had more background information, data, and findings than space on their poster. They were advised to have the poster serve as an overview and to write a detailed summary of their work that could handed-out to interested visitors. Their focus was then on how to produce the poster and they located the needed special paper and roll-fed plotter.

The poster and hand-out were completed about one and a half weeks before the deadline. But upon review it was discovered the word environmental was misspelled in the title. To remedy the problem one of the students had to learn how to run the plotter since the operator had gone on vacation. The student then decided who would make the various presentations. This was followed by several practice presentations and question and answer sessions. The result was the students had a very professional looking poster and well developed complementary hand-out.

FACULTY EXPERIENCE

This first attempt engaging business undergraduates in research activities changed the classroom behavior of both the instructor and students. The instructor saw his role evolve from being primarily a presenter of knowledge to being a more of a facilitator and mentor. This role change was also accompanied by changes in the pedagogy followed in the course, and most importantly, in the growth of the students from passive to active learners.

Lessons learned from this first attempt to engage undergraduate business students in research activities include the following:

- Instructors must design into their course syllabi time and opportunities for research activities. Though this may initially be viewed as taking time away from important topics, these opportunities can be viewed as alternative ways to cover these topics.
- Often ideas for student research can be obtained by the instructor from being active in local sections of professional societies. This helps keep the research relevant to current issues.
- Students have different academic strengths and skill sets, and working together allows students the opportunity to use them to their project's advantage.
- Smaller in size and/or scope research projects are better than one long project that stretches over several semesters. Students need to experience the beginning, middle, and end of the project.
- Students should work in small groups especially if the deadline to present their findings is only two or three months.
- Class size needs to be limited to a reasonable number. If no graduate assistants are available this experience suggests no more than twenty students in the course.
- Both the instructor and the students need to realize their roles in the course, and behavior in the classroom will change significantly. The complexity and number of student questions rose significantly resulting in a much more dynamic learning environment.
- Instructors must be prepared to allocate more time supporting courses with research activities.

These findings support the argument that engaging undergraduate business students in research activities should only be done after individual institutions address the relationship between research and a liberal arts education.

CONCLUSION

The experience of adding research activities to an undergraduate business course completely changed how a course, Introduction to Management Science, was designed and conducted. The student learning process also changed as students learned the meaning of good business research, how it is conducted, and how to present research findings at several different types of venues.   The instructor learned business students appreciate an opportunity to investigate a problem for a extended period of time under the guidance of a faculty member.   This course concluded with the students identifying projects for next year's class.

ACKNOWLEDGEMENTS

REFERENCES

Anderson, D. (2008). An Introduction to Management Science – Quantitative Approaches to Decision Making. 12th Edition, Mason, Ohio.

ASHE Higher Education Report (2007). *Engaging Students in Research and Creative Activities.* Vol. 33 issue 4, p. 57-74.

Elgren, T. and N. Hensel (2006). *Undergraduate Research Experiences: Synergies between Scholarship and Teaching.* peerReview, AAC&U.

Boyer Commission on Educating Undergraduates in the Research University (1998). *Reinventing Undergraduate Education: A Blueprint for America's Research Universities.* Stony Brook, NY: State University of New York.

Chang, W. (2005). *Impact of constructivist teaching on students' beliefs about teaching and learning in introductory physics.* Canadian Journal of Science, Mathematics and Technology Education, 5(1), 95-109.

Kierniesky, N. (1984). *Undergraduate Research in Small Psychology Departments.* Teaching of Psychology. Vol. 11 No. 1.

Mills, N, (2005). *Undergraduate research in chemistry at Trinity College: The value of enlightened self-interest.* Presented at the American Chemical Society National Meeting.

Stoblein, M. and J. Kanet (2008). *Developing Undergraduate Student Experiences in Operations Management.* Decision Sciences Journal of Innovative Education. Volume 6 number 2.

Experiential Learning and Organizational Culture

Jack Rappaport
Management Department
La Salle University
Philadelphia, PA 19141
E-mail: rappapor@lasalle.edu

Stephen B. Richter
Department of Accountancy and MIS
Villanova University
Villanova, PA 19085
E-mail: stephen.richter@villanova.edu

Experiential Learning and Organizational Culture

Abstract

Rappaport and Richter [36] [37] described the results of an experiential learning experiment that uses the racetrack as an application to teach statistics and decision making. A full blown field trip to the racetrack was implemented as part of this process. In this paper we study this experiential learning case within the context of organizational culture. We develop the case using stakeholder analysis and influence diagrams. A model is developed which uses a dynamic process whereby organizational change can be understood from a cultural standpoint.

Experiential Learning

The concept of experiential learning has been widely accepted as a useful approach to learning in many academic fields. In our example the objective is to enhance the teaching of either a statistics or a problem solving class by using the application of horse racing. The racetrack can be seen as a natural laboratory within which to study the various concepts of risk and decision making. Experiential learning is learning through reflection on doing, which is often contrasted with rote or didactic learning. An example of experiential learning is going to the zoo and learning through observation and interaction with the zoo environment as opposed to reading about animals from a book. Thus one makes the discoveries and experiments with first hand knowledge, instead of reading from a book. Experiential learning theory draws on the works of prominent 20th century scholars who gave experience a central role in their theories of human learning and development; notably John Dewey, Kurt Lewin, Jean Piaget, William James, Carl Jung, and others. For example, John Dewey [17] conceived of education as a continuing reconstruction of experience, whereby the process and goal of education are one and the same. The field trip extends the experiential approach to both statistics and problem solving classes to engage the students in a real world experience by taking them to the racetrack. The concept of the racetrack field trip can be seen within this philosophy of education, since it allows the students to experience the concepts of risk and return within an extended social setting; the concepts of risk and speculation are made more apparent as the students are immersed in a real gambling environment.

The Academic versus the Practitioner Approach

A natural source of conflict in any experiential learning experiment can result from the dichotomy that often exists between the academic approach and the practical approach in any given discipline. Research oriented faculty are generally more concerned with the academic component of any teaching experiment whereas teaching oriented faculty and the students in general may be more willing to experiment with a more hands on approach; administrators may go either way depending upon their background and the current climate within the school. Academic research in the field of horse racing is quite sophisticated and can be separated into two major groups; the first group deals primarily with the development of decision strategies for making horses race selections. Many of the models are very sophisticated from an academic standpoint since there has been a significant contribution to this field from many high quality academic researchers [6] [22]

[2] [11] [25] [34] [38] [41] [43] [47] [50]. Another group of articles are concerned with the study of risk; virtually all of the empirical work analyzing risk preferences in the horse race betting markets shows the representative bettor to be risk loving. This has resulted in what is called the "favorite-longshot" bias; an established feature of the betting markets whereby longshots win less often than their subjective probabilities imply and favorites win more often. There has been a considerable amount of research studying this phenomenon [1] [12] [35] [39] [44] [46] [49] [55] [56].

On the popular side, there are a large number of publications and information sources available to the average horseplayer to help them better understand the sport and make better decisions. These kinds of publications would be considered trade publication by most academics. The most important source of information is the Daily Racing Form (DRF), a newspaper-style publication which can be an important tool for the handicapper or horseplayer. There are also numerous books on the subject that provide a variety of theories on how to best interpret the information provided by the DRF. For example a number of well known handicappers such as Andrew Beyer [3] [4] [5] have written books on this topic. The DRF also publishes books that study special aspects of horse race handicapping, such as how to interpret such factors as speed and pedigree [7] [8] [23] [24] [26] [27] [30] [48].

Experiential Learning and Organizational Culture
An underlying assumption of our approach is that in many cases the successful implementation of an experiential learning experiment cannot be fully realized without an understanding of the organizational culture. Culture is widely thought to be the "natural outgrowth of the social interactions that we call organizations" [51]. Organizational culture is generally regarded as consisting of both beliefs and social practices, including communication, socialization, and shared experiences. Culture can also be conceptualized in terms of homogeneity-heterogeneity across a social system, a formulation that provides an easy connection to demographic homogeneity-heterogeneity [28] [32]. Martin [29] describes several different views that can be used to operationalize organizational culture: namely, integration, which posits culture as shared meanings and values; differentiation, which assumes that there exists subgroups within any organization that differ in their shared meanings; and finally fragmentation which suggests that culture is a differential network of meanings that are interrelated but are ill-defined and inconsistent.

An anecdotal story relating to our racetrack field trip illustrates how organizational culture can be a relevant issue for faculty who engage in experiential learning. After implementing the field trip, which consisted of 130 students attending the racetrack for the afternoon, the chair of the English department wrote a fairly caustic (and slightly sarcastic) letter to both the Dean of arts and science as well as the Dean of business, suggesting that the field trip was inappropriate. He had been concerned about a field trip of a geology class to a mine earlier in the semester and now wanted the administration to review the university policy with regard to all field trips. The tone of his letter clearly showed a cultural bias against the racetrack and this seemed to be the main force motivating him to write the letter, even though he already had a concern about the appropriateness of field trips in general. Apparently he thought that a general bias against the racetrack by the faculty and the administration would be the critical factor in getting his attention.

This incident shows that organizational culture can be a key determinant of how educational processes evolve and can be the basis that we can use to understand how the various constituencies or stakeholders interact and have influence on each other. These constituencies could include not only students, faculty and academic administrators, but also external stakeholders such organizations that are potential employers for the students, as well as alumni and any potential benefactor for the school. Cultural issues can then play a role as they focus on understanding how the important participants or potential participants can be either positively or negatively impacted by the project or the experiment.

Influence and Social Networks
Because of the informal nature of an educational institution, it is likely that innovative projects are implemented more as a result of organizational culture and the process of socialization than any formal power and authority structure. For this reason we decided that the application of influence or social networks would be the best approach to take. We also borrow some of concepts developed by Harrison and Carroll who studied how the structure of influence patterns within the organization affects the intensity of socialization pressures experienced by the individuals (in contrast to the more typical assumption that all the individuals have equal influence on each other [19] [42]). An interesting aspect of their work is a model that represents influence with continuous variables and assumes an endogenous process of influence change over time. Both of these features are attractive in the sense that they reflect realistic assumptions about the processes involved in organizational culture [9] [20] [21].
Influence networks have been used extensively by industry practitioners and consultants. For example, Ross Dawson, recognized globally as a leading futurist. Entrepreneur, and strategy advisor uses organizational network analysis to better understand how large technology purchasing decisions are made [14] [15] [16]. In one example he uses a network approach to understand how the various organizational roles are combined, what inputs they receive and what are the roles in the process of vendors and other external parties.

Stakeholder Analysis
Stakeholder analysis can be useful in understanding the major constituencies involved in any case on organizational change or innovations. Much of the literature on stakeholder analysis is concerned with developing various mapping processes which can help to visualize and understand the extent to which the different stakeholders have on the ultimate success of the project. The potential list of stakeholders will almost always exceed what is feasible and sensible to consider in the mapping process, so the challenge is often to focus on the "right stakeholders" and then use a mapping tool to help visualize this critical subset of the whole community. The most common presentation styles use a matrix to represent two dimensions of interest; some of the commonly used dimensions include, "power" (high, medium or low), "support"(positive, neutral or negative or "influence" (high or low).
Mitchell, Agle et al [31] proposed a classification of stakeholders based on power to influence, the legitimacy of each stakeholder's relationship with the organization, and the urgency of the stakeholder's claim on the organization. The results of this classification

can help address the concept of "salience"- i.e. the degree to which managers give priorty to competing stakeholder claims [31]. Fletcher Guthrie et al [18] defined a process for mapping stakeholder expectations based upon value hierarchies and Key Performance Areas (KPA). Savage, Nix et al [40] offer a way to classify stakeholders according to their potential for threat and their potential for cooperation. Turner, Kristoffer and Thurloway [52] developed a process of identification, assessment of awareness, support and influence leading to strategies for communication and assessing stakeholder satisfaction.

A useful approach to stakeholder analysis is based is upon the extensive body of work focusing on influence networks. This research focuses on the importance of relationships through the study of "influence networks", "social networks", and "social capital", viewing projects as temporary knowledge organizations (TKOs) and more recently the idea of CRPR (Complex Responsive Processes of Relating) [54]. All of these theories emphasize the critical importance of the relationships between different stakeholders both within and around the project team.

The Stakeholders

In this section we list the important stakeholders involved in the process and how each of these could be either positively or negatively impacted by the project. We categorize them as either "primary stakeholders", "secondary stakeholders", or "key stakeholders" (who can also belong to the first two groups). The primary stakeholders are those that are ultimately affected, either positively or negatively by the outcome of the project. The secondary stakeholders are the "intermediaries", i.e. the persons or organizations who are indirectly affected by the outcome of the project. We also categorize the secondary stakeholders as either internal or external with respect to the organization (i.e. the university). The key stakeholders were chosen to be the faculty developers and the Dean since they are the ones that have the most prominent role in the decision making process and the most at stake in terms of the outcome of the experiment.

Table 1: Primary Stakeholders

Primary Stakeholders	Potential Benefits	Potential Costs
Students	Enhances learning	Distract students
	Increases the motivation to learn	Potential to miss other classes
	A free trip to the racetrack/casino	
Faculty Developer	Enhances the teaching environment	Conflicts with university administrators
	Increases the motivation to teach	Distrust from other faculty
	Write papers on the application of	
	experiential learning	
Dean	Foster faculty development	Risk of student and faculty complaints
	Foster innovative teaching	Conflicts with other administrators
	Potential for industry contacts	

Table 2: Secondary Stakeholders (Internal)

Secondary Stakeholders (Internal)	Potential Benefits	Potential Costs
Business School Faculty	Help foster collegiality	Potential for students to miss classes
Arts and Science Faculty	Increase support for field trips	Different educational philosophy
		Potential for students to miss classes
Dean of Student Affairs	Improve the social atmosphere	Excessive drinking and/or gambling
		Parental concerns
NCAA Compliance Officer	Little or no potential benefit	Perceived violation of NCAA rules
Greek Organizations	Potential to fund raise by participating	Little or no potential cost
	in a handicapping contest	
Member on the Board	Enhance horse ownership	Opinions of other board members

Table 3: Secondary Stakeholders (External)

Secondary Stakeholders (External)	Potential Benefits	Potential Costs
Director of Racing	Course on racetrack management	Conflict with boss
The Racetrack	Increase awareness of horse racing	Potential for funding costs
	Increase attendance on a slow day	
Horse Owners/Breeders	Enhance participation in the sport	Time and costs
	University lectures	
Horse Trainers	Enhance participation in the sport	Time and costs

Influence Maps

We use a series of influence diagrams or maps to represent the nature of the influences between the various stakeholders. In order to reduce the complexity of the problem, we simplify the overall problem by developing influence diagrams for three separate sub-problems corresponding to the following environments: the classroom, the university at large, and the external environment. Associated with each diagram is a qualitative description of the process of cultural transmission and socialization that occurred as a result of the influences represented in the diagrams. The process is to be viewed as dynamic and evolving over a period of time.

The following diagrams represent the influence maps for each of the three scenarios or contexts mentioned above.

Figure 1
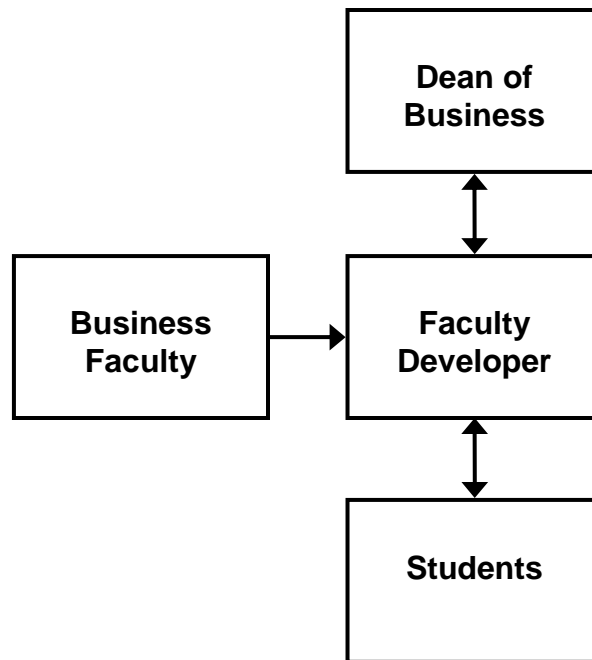
This diagram maps the basic influences among the stakeholders with respect to the learning environment in the classroom and at the field trip. The key component of this influence map is the interaction between the faculty developer and the students in the classroom environment. The integration was carried out in a measured way and occurred over a period of many semesters prior to the field trip (i.e. it was assumed that students from prior semesters had some influence on students in the current semester). In the beginning it was used as a method to teach basic topics such as expected value, standard deviation, risk and decision making and later the focus was expanded by simulcasting live races into the classroom and even making a few wagers during the class. Needless to say, the reception of the students was generally very favorable, since students are often bored with the traditional approaches to teaching.

This process can be seen as a form of cultural transmission from the faculty member to the students. When the idea of the field trip first surfaced there was a critical mass of positive reactions as a result of this socialization process. A critical amount of peer pressure was also a positive force.

The age of the students was generally a positive factor for the overall success of the experiment, since young people generally are more willing to accept new ideas and practices particularly since many college students have an interest in gambling (it should be noted that the legal gambling age for horse racing is 18 not 21 and that horse racing is theoretically the only legal form of internet gambling). Ironically, the racetracks have not been able to attract younger people and it is well known that the racetrack population is aging (on a typical weekday you might see almost no participants under the age of 30 a given track or OTB center). The fact that the age was a net positive factor illustrates the importance of using demographic variables such as age in the study of organizational culture. In his essay, Pfeffer [33] made a very persuasive case for the study of organizational demography.

The interactions between the other business faculty and the faculty developer also played a role in the process of stakeholder influences. The prior and current research and publications of the faculty developer in the horse racing field created a positive synergy with many of his peers and also created a positive climate within the teaching environment. This process could be viewed in the larger contest of academic institutions whereby many research oriented faculty create peer socialization pressure on their colleagues to continue to publish.

The Dean was more of a passive stakeholder in the process willing to accept any innovations as long as there were no significant negatives. The positive feedback on the course evaluations and the lack of any significant negative feedback from either the students or the other faculty kept him in a steady state of neutrality with respect to the teaching experiment.

Figure 2
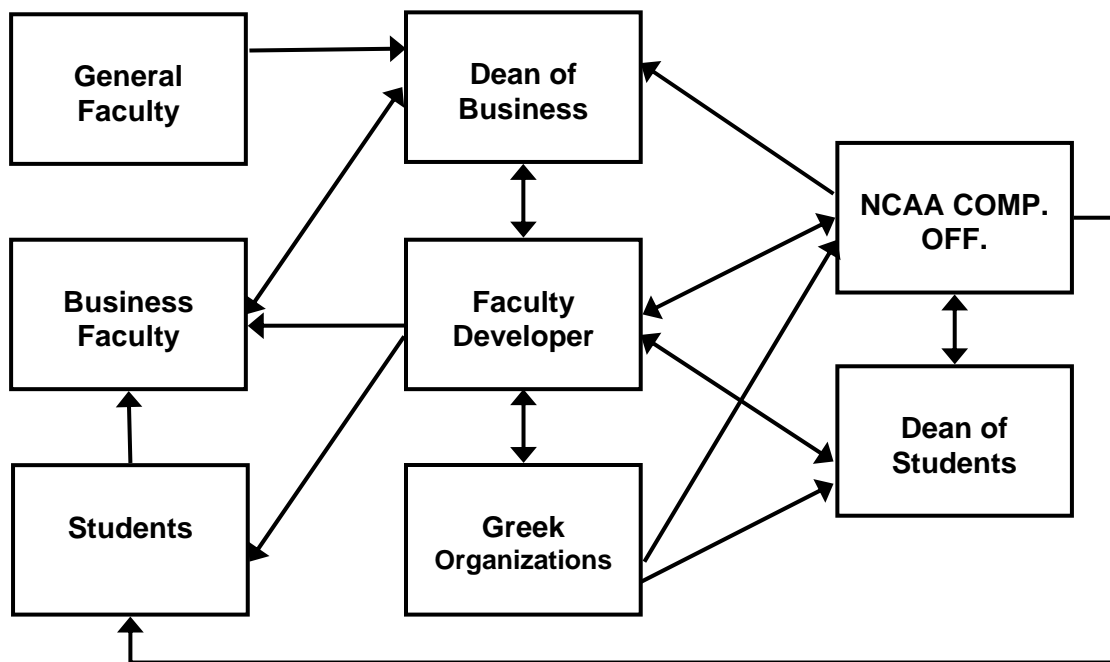
This diagram represents the influences among the stakeholders in terms of the acceptance of the field trip within the larger campus community. The influence process within the larger campus showed some sources of conflict between some of the stakeholders. Perhaps the strongest source of conflict resulted from the concern of the NCAA compliance officer that the field trip could be perceived as a violation of the spirit of the NCAA rules concerning the involvement of student athletes in any form of gambling. Although student athletes were officially permitted to gamble on the races, the NCAA clearly frowns upon any form of involvement in gambling.

Beliefs and assumptions are often difficult to measure but they provide the foundation from which behavior and action spring [13]. The attitude of the compliance officer was influenced by the statements and policies of the NCAA but possibly also by her own belief and value system. The Dean of students seemed to have a similar view on the matter, and indeed stated directly that her own personal decision would be to opt against such a trip.

One of the most important cultural trait is a sense of mission. Successful organizations have a clear sense of purpose and direction that defines organizational goals and strategic objectives [13]. The attitude of the compliance officer may have been affected not only by her personal belief system but also her interpretation of the mission of the university, which as a Catholic university is based upon the notion of community spirit and service. This source of conflict was mitigated by the development of a handicapping contest where the prize money was to be donated to charity (and that no money was to be directly distributed to any student). Several of the Greek organizations agreed to participate in the contest which also created some positive synergies with the student affairs area. This approach seemed to alleviate much of the tension between the compliance officer and the faculty developer (in fact it was the compliance officer who suggested the charity concept). The faculty developer also cooperated with her by supplying her with a list of all the student athletes that would potentially go on the trip.

Another potential source of conflict was from some of the arts and science faculty since many of the students would have to be excused from their classes in order to go on the trip (note the letter written by the chair of the English department that was mentioned above). Their attitude may also be the result of a cultural bias against more practical learning approaches. This source of conflict was moderated by building the appropriate network with the business school faculty who experienced the largest amount of class excuses. As the interest in the trip increased, the business school faculty became more and more receptive to the idea, until the point where some of them actually informally cancelled classes.

Figure 3
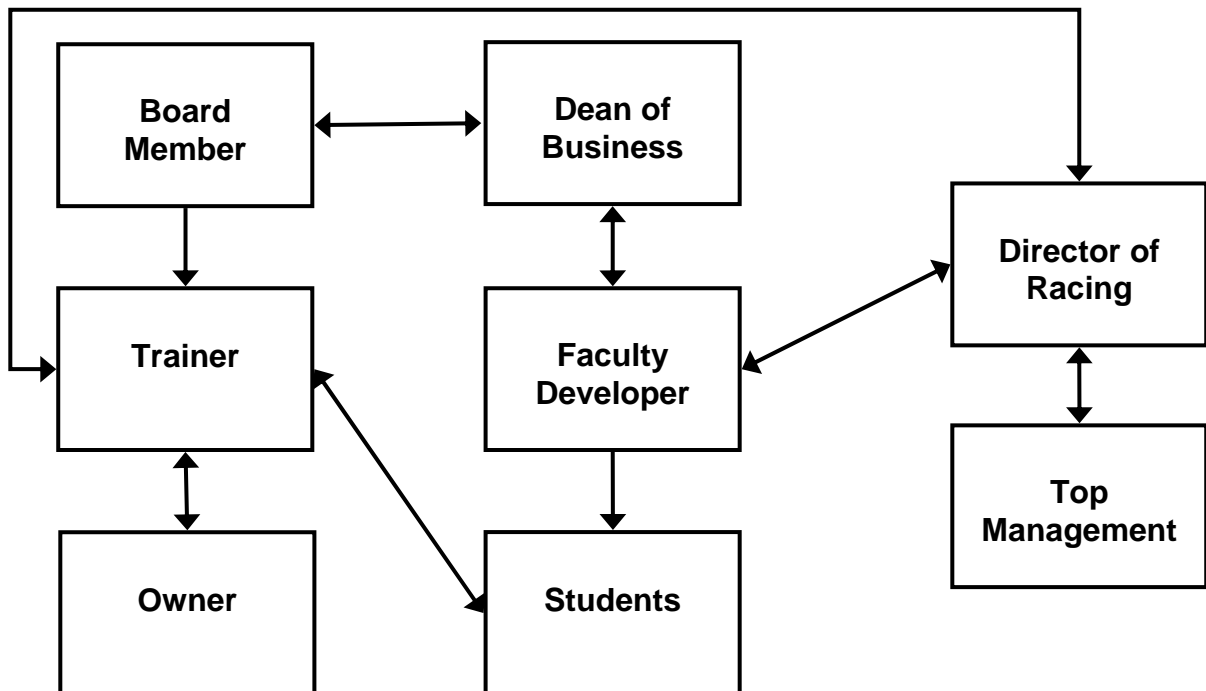
The third and final influence diagram models the relationships between the various stakeholders in terms of the development of the project with the external community and potential fund raising sources. The faculty developer was fortunate to obtain a $2000

contribution from a weathly horse owner and breeder to help fund the costs of the field trip (i.e. money for buses and prize money for the handicapping contest). This clearly created many positive synergies because it provided an atmosphere of legitimacy to the entire experiment. Another positive development occurred when the faculty developer connected with one of the board of trustees who owned racehorses with her husband. He contacted one of the leading trainers at the track to implement some visits to the barn area for selected students which created more networking possibilities at the track and the potential for a stronger relationship with the racetrack management. No doubt the Dean was impressed when at an informal dinner with the board, this particular board member expressed a great deal of enthusiasm about the experiment.

These influences show the relevance of diversity and general demographic factors in the transmission of organizational culture [10] [53]. The positive effect of the involvement of horse owners resulted in part from their socio-economic status. The influence of the university board member was particularly important since the effect of her socio-economic status was strengthened by her status as a board member. The trip to the barn area also shows the relevance of diversity and demographics since the culture of the backstretch of the racetrack is quite distinct and is generally restricted to only the owners, trainers and the general barn workforce. The focus on this aspect of horse racing definitely created more positive synergies between the various stakeholders.

Dynamics over time
The stakeholder influences can be viewed as a recursive process whereby the direct influence that one stakeholder has on another can cause a chain reaction of influence on other stakeholders which can eventually cycle back to the original stakeholder. For example, some of the more aggressive students convinced many of the business faculty of the positive nature of trip in order to be excused from class; this positive influence was transmitted between the faculty and eventually to the Dean; this then created a cycle whereby more faculty and students were positively influenced. Of course this kind of cycle could also occur in the negative direction which was beginning to occur because of the initial negative reaction of the NCAA compliance officer.

The process can seen as evolving over the three stages discussed above (the classroom, the general university and the external environment). A so-called "snapshot" of each stakeholder can be taken in each environment, and the sequence of these snapshots can represent a dynamic process that evolves over time. The net result was positive and after the field trip the Dean suggested that the faculty developer make a presentation to the business school faculty concerning the results and future plans of the experiment. Apparently he wanted the faculty to know that the experiment evolved much beyond the classroom and had implications for the general university and the enhancement of particular relationships with the external community.

Needless to say, the faculty developer gained a tremendous amount of insight by his involvement in this dynamic influence cycle. The influence relationships gave him the necessary insight to decide how to better allocate his resources as well as his political capital. He concluded that the best overall strategy was to gain the acceptance of the students, create a charity fund raising component to the handicapping contest and emphasize the barn visitations to complement the gambling aspects.

Conclusions

The paper studies an interesting case that is useful from many different perspectives. Overall, it illustrates the importance of studying the dynamics of an organization when embarking on an innovative teaching experiment. Many academics often ignore the political ramifications of their teaching approaches and can be unnecessarily frustrated at the implementation stage; this phenomenon can also apply to research, seeking grants, or any academic project that can potentially interact with variety of stakeholders. Of course the most important goal for academics is to focus on the purely intellectual aspects of their work; but ignoring the impact that their work may have on their students, colleagues, the university, and the external community may limit its effectiveness and not enable the work reach its greatest potential.

The case also illustrates how influence diagrams can be a useful in studying the interrelationships amongst the various stakeholders with respect to a particular project. A unique approach is developed which analyses the overall influence process by dividing it into smaller components. The analysis of the case is primarily based on cultural issues which are often neglected in organizational studies. Finally, the case shows that experiential learning can be enhanced by using a diverse set of applications. There is a great deal of potential for faculty to engage in novel experiments as a way to increase the interest of the students and also to interact with the external environment.

References

[1] Asch, P., Malkiel, B.J., and Quandt, R.E., (1984), "Market Efficiency in Racetrack Betting", The Journal of Business, Vol. 57, No.2, April

[2] Bain, R.S., Hausch, D.B. and Ziemba (2006) "An application of expert information to win betting on the Kentucky Derby, 1981-2005", The European Journal of Finance, Jun, Iss.4

[3] Beyer, A. (2007) The Winning Horseplayer: An Advanced Approach to Thoroughbred Handicapping, Houghton Mifflin

[4] Beyer, A. (1994) Picking Winners: A Horseplayer's Guide, Houghton Mifflin

[5] Beyer, A. (1995) Beyer on Speed: New Strategies for Racetrack Betting, Houghton Mifflin

[6] Bolton, R.N. and Chapman, R.G. (1968) "Searching for Positive Returns at the Track: A Multinomial Logit Model for Handicapping Horse Races", Management Science, Aug, Vol 32 Iss 8, Linthicum

[7] Borg, N., Off The Charts (2003) Daily Racing Form Press

[8] Cardello, J., Speed to Spare (2003) Daily Racing Form Press

[9] Carroll, G.R. and Harrison, J. R. (1998) "Organizational Demography and Culture: Insights from a Formal Model", Administrative Science Quarterly 43

[10] Chatman, J.A., Polzer J.T. et al (1998) "Being Different Yet Feeling similar: The Influence of Demographic and Organizational Culture on Work Processes and Outcomes", Administrative Science Quarterly, Vol 43, No 4

[11] Chezum, B. and Wimmer, B.S. (2000) "Evidence of adverse selection from thoroughbred wagering", Southern Economic Journal, Stillwater, Jan, Vol66 Iss 3

[12] Coleman, C. (2004) "New light on the longshot bias", Applied Economics" London, Mar Vol 36 Iss 4

[13] Denison, D. (2001) "Organizational Culture: can it be a Key Lever for Driving Organizational Change?" in Cooper, C.L., Cartwright, S., Early, P.C. (eds.), The International Handbook of Organizational Culture and Climate, New York: Wiley

[14] Dawson (2002) Living Networks, Financial Times/Prentice Hall

[15] Dawson (2005) Developing Knowledge-Based Client Relationships, Elsevier

[16] Dawson (2009) Implementing Enterprise 2.0, Advanced Human Technologies

[17] Dewey, J. (1938) Education and Experience, New York: Simon and Schuster

[18] Fletcher, A., et al. (2003) "Mapping stakeholder perceptions for a third sector organization", Journal of Intellectual Capital 4(4)

[19] Friedkin, N.E. (1998) A Structural Theory of Social Influence, Cambridge: Cambridge University Press

[20] Harrison, J.D. and Carroll, G.R. (1991) "Keeping the Faith" A Model of Cultural Transmission on Formal Organizations" Administrative Science Quarterly 36

[21] Harrison, J.D. and Carroll, G.D. (2001) "Modeling Organizational Culture: Demography and Influence Networks" in Cooper, C.L., Cartwright, S. Earley, P.C. (eds.), The International Handbook of Organizational Culture and Climate, New York: Wiley

[22] Hausch, D.B. and Ziemba, W.T. (1990) "Arbitrage Strategies for Cross-Track Betting on Major Horse Races", The Journal of Business, Chicago Jan, Vol 63 Iss 1

[23] Heller, B. (2004) Go For The Green, Daily Racing Form Press

[24] Illman, D. Betting Maidens & 2-Year-Olds (2005) Daily Racing Form Press

[25] Iyengar, G.N. and Cover, T.M. (2000) "Growth optimal investment in horse race markets with costs", IEEE Transactions on Information Theory Nov, New York

[26] Keppler, D. Trainer Angles (2006) Daily Racing Form Press

[27] Klein, S., The Power of Early Speed (2005) Daily Racing Form Press

[28] Koene, B.A., Boone, C., Soeters, J.L. (1997) "Organizational factors influencing homogeneity and heterogeneity in organizational cultures" in S.A. Sackmann (ed.), Cultural Complexity in Organizations, Newbury Park, CA: Sage

[29] Martin, J. (1992) Cultures in Organizations: Three Perspectives, New York: Oxford University Press

[30] Michaels, N. (2005) Handicapping Contest Handbook Daily Racing Form Press

[31] Michtell, R.K., B.R. Agle, and D.J. Wood (1997) "Toward a Theory of Stakeholder Identification and Salience: Defining the Principle of Who and What really Counts", Academy of Management Review 22(4)

[32] O'Reilly, C.A. and Chatman, J. (1996) "Culture as Social control: Corporations, cults and commitment" Research in Organizational Behavior, 18: Greenwich, CT: JAI Press

[33] Pfeffer, J. (1983) "Organizational demography", Research in Organizational Behavior, 5: 299-357, Greeenwich, CT: JAI Press

[34] Phatarfod, R. (1999) "Betting strategies in horse races", Asia-Pacific Journal of Operational Research, Singapore May, Vol 16 Iss 1

[35] Quandt, R.E. (1986) "Betting and Equilibrium", The Quarterly Journal of Economics, Vol. 101, No. 1, Feb

[36] Rappaport, J. and Richter, S.B. (2009) "Teaching Note: Decision making using the racetrack betting markets", International Journal of Applied Decision Sciences

[37] Rappaport, J. and Richter, S.B. (2010) "An Experiential Learning Case: A University-Based Field Trip to the Racetrack", Proceedings, Northeast Decision Sciences Institute, Alexandria, Va.

[38] Rosenbloom, E.S. (2003) "A better probability model for racetrack using Beyer speed numbers", Omega Oct, Vol 31 Iss. 5

[39] Rhoda, K.L., Olsen, G.L., Rappaport, J.M. (1999) "Risk Preferences and Information Flows in Racetrack Betting Markets", The Journal of Financial Research, Vol XXII, No 3, Fall

[40] Savage, G.T., T.W. Nix, Whitehead and Blair (1991) "Strategies for assessing and managing organizational stakeholders", Academy of Management Executive 5(2)

[41] Schnytzer, A. and Shilony, Y. (1995) "Inside information in a betting market", The Economic Journal, London, Jul, Vol 105 Iss 431

[42] Scott, J. (1991) Social Network Analysis, London: Sage

[43] Shin, H.S. (1991) "Optimal Betting Odds Against Inside Traders", The Economic Journal, London, Vol 101 Iss 408

[44] Shin, H.S. (1992) "Prices of State Contingent Claims with Insider Traders, and the Favorite-Longshot Bias", The Economic Journal, London, Mar, Vol 102, Iss 411

[45] Snyder, W. (1978) "Horse Racing: Testing the Efficient Market Hypothesis", Journal of Finance, 33

[46] Sobel, R.S. and Raines, S.T. (2003) "An examination of the empirical derivatives of the favorite-longshot bias in racetrack betting", Applied Economics, London, Mar 4 Vol 35 Iss 4

[47] Stein, W.E. and Mizzi, P.J. (2003) "An analysis of exotic wagers in a parimutuel setting", Applied Economics, London, Mar 4 Vol 35 Iss 4

[48] Stitch, L. (2004) Pedigree Handicapping Daily Racing Press

[49] Swindler, S. and Shaw, R. (1995) "Racetrack wagering and the "uninformed" bettor: A study of market efficiency", Quarterly Review of Economics and Finance, Greenwich Fall 1995 Vol 35 Iss 3

[50] Thalheimer, R. and Ali, M.M. (1995) "Exotic betting opportunites, pricing policies and the demand for parimutuel horse race wagering", Applied Economics, London Aug, Vol 27 Iss 8

[51] Trice, H. and Beyer, J. (1993) The Cultures of Work Organizations, Englewood Cliffs, N.J., Prentice Hall

[52] Turner, J.R., V. Kristoffer, et al., Eds. (2002) The Project Manager as Change Agent, London, McGraw-Hill

[53] Watson, W.E., Kumar, K., Michaelsen, L.K. (1993) "Cultural Diversity's Impact on Interaction Process and Performance: Comparing Homogenous and Diverse Task Groups", The Academy of Management Journal, Vol 36, No 3

[54] Weaver, P. (2007) A Simple View of Complexity in Project Management, Proceedings of the 4th world Project Management Week, Singapore

[55] Weitzman, M. (1965) "Utility Analysis and Group Behavior: An Empirical Study", Journal of Political Economy, 73

[56] Williams, L.V. and Paton, D. (1997) "Why is there a favorite-longshot bias in British racetrack betting markets?", The Economic Journal, London, Jan Vol 107 Iss 440

CAN LEAN SERVICE APPLICATIONS BE LEARNED THROUGH THE USE OF A MANUFACTURING SIMULATION?

Susan E. Pariseau, Merrimack College, 315 Turnpike Street, North Andover, MA 01845, (978) 837-5417, susan.pariseau@merrimack.edu

ABSTRACT

This preliminary study seeks to determine whether students exposed to lean manufacturing can apply these concepts in a service environment. Undergraduate students are provided an opportunity to learn lean principles using the Time Wise physical simulation in which clocks are assembled using a multi-stage process. To determine if they can apply the principles to services, students are asked to write an essay demonstrating the application of lean principles in a service setting.

INTRODUCTION

This paper describes the use of a hands-on manufacturing simulation, Lean 101, developed by Time Wise Solutions® (see http://www.timewisems.com/time_wise_solutions.html), to teach lean principles while providing a backbone for the OM course and connecting the various topics. The methodology used with Lean 101 in the introductory operations management course is one of guided exploration where the amount of direction is dependent upon the particular group of students. Before and after each round of the simulation, students complete exercises, related to the simulation, but with the purpose of teaching a variety of OM topics. At the beginning of the class following each round of the simulation, student teams are also required to make brief presentations focused on describing waste found during their analysis and discussion. They must also provide actionable suggestions for change that will improve performance by eliminating the waste identified.

The purpose of this paper is to determine whether students can apply lean principles in settings outside of the context of the manufacturing simulation used in class. Lean is presented and discussed primarily through the manufacturing simulation. The final examination essay requires students to provide and categorize an example of waste in a service environment. A modification to eliminate the waste and make the service leaner must be provided.

OM COURSE

The simulation is used in a two-credit introductory operations management course generally taught to second semester sophomores or first semester juniors in a business school. Prerequisites for the course include four-credit courses in introductory business, business statistics and accounting. The sophomore core at the institution is comprised of two-credit courses in marketing, finance, management information systems, organizational behavior and operations management and a four-credit course in Law and Ethics. When the curriculum was designed, many of the three semester hour courses were converted to two semester hours to allow for three new courses that serve to integrate the functional areas (freshmen, junior and senior-level courses). The upper-level discipline-specific courses were also converted from three

to four semester credit hours in order to provide greater depth for those students concentrating in a particular discipline. Thus, the faculty teaching the sophomore core was required to determine what "every business student needed to know" and to design the new core courses around this smaller number of topics. In operations management, the selected topics included the introduction, OM strategy, forecasting (with a link from statistics), process, quality, supply chain management, just-in-time and inventory management. The simulation is used to tie the course together so that students see a cohesive whole rather than a series of disjointed and seemingly unrelated topics.

TIME WISE SIMULATION

The Time Wise Lean 101 simulation is used by Manufacturing Extension Partnerships (MEPs) to provide small and medium manufacturing enterprises (SMEs) with training and assistance in the application of lean concepts in a make-to stock environment. This hands-on simulation involves all participants in a manufacturing environment where they are required to make and ship a variety of clocks. All the materials necessary to make the clocks are included in the simulation kit which may be purchased from the developer, Time Wise Solutions®.

In the corporate environment, the simulation is normally run during a single eight-hour day where lectures introducing lean concepts are interspersed with four 15-minute rounds of the simulation. In the academic environment, changes may be made to incorporate more of an exploratory approach where students are given freedom to make changes they believe will improve the process to provide on-time delivery of quality clocks.

In my operations management course, each unique class section (I teach four sections each semester) will follow a somewhat different Time Wise scenario since the students are required to submit their own improvements and methodologies for waste elimination. Four 15-minute simulation rounds, spread over three class periods (each class period is approximately two hours in length), are used. This schedule seems to provide sufficient time for students to generate good lean solutions.

The Players

Ideally, the simulation involves 15-24 players. I cap my class at 18 students and generally all my classes have 17-18 students, an ideal number. Each class is also divided into three teams of 5 to 6 students for the purpose of completing assignments and preparing presentations. At the outset of the simulation, students must fill the following roles: assembler (face, back, clock and hand), inspector, reworker, sales representative, production scheduler, kitter, material handler, warehouse clerk, instruction crib attendant, tool crib attendant, industrial engineer, trucker and vendor.

Round 1

Before play begins, both the required roles and the product lines are introduced. Initially, two different clocks are produced with different lot sizes and different costs and sales prices. When students enter the classroom, the game is ready to play. There is WIP at various workcenters and there are finished goods in the warehouse. The facility layout has been established and the flow is jumbled. Roles have been assigned to each student. Students are told that the competitive priorities of Time Wise are quality and delivery. At the end of the 15 minute round, data

collected is placed in a spreadsheet that is made available to the students. The following data is provided: average manufacturing lead time, number of units in WIP (beginning and ending), number of units shipped (on-time and late), number of units in finished goods warehouse, number of employees, number of tables used, distance (in feet) travelled by a clock, and inspection results (number of units passed and failed). The overall dismal results are discussed. These results include lead times in excess of the 15 minutes of playing time with only beginning WIP making it to the finished goods warehouse. The WIP warehouse is overloaded with kits; the distance travelled is long. When the income statement is generated by the students, they find that Time Wise has lost money.

After the first round, the students discuss what happened and are asked to provide suggestions for improvement. The layout is the first item that is generally mentioned by the students. I talk about 5S, value and the value stream and discuss the seven types of waste. The assignment for the next class requires that each team prepare a presentation (1) identifying waste observed in round 1 (categorized by type), (2) providing an income statement, and (3) providing specific, actionable suggestions for improvement.

Round 2

Round 2 incorporates suggestions that have been generated by the student teams in each class. All teams provide similar layouts incorporating better flow for the second round and so I set-up play and provide WIP and finished goods before they enter the room. To date, at least one team in each class has recommended the elimination of the instruction and tool cribs with point of use storage (POUS) of the instructions and tools. Thus, POUS is incorporated in round 2 and students are required to 5S their work stations. The WIP storage area is also eliminated in this round.

Once again, there is discussion following the simulation. Students are encouraged to modify the layout and to suggest other improvements for round 3. Data is collected for round 2. The classes generally show some improvement in performance but few clocks are completed and quality is poor.

Round 3

Round 3 is played during the same class period as round 2. The students help to disassemble clocks and put all the materials back to their original placement. They modify the layout and make a variety of changes. If they do not suggest elimination of kitting, I probe until they realize that kitting is a waste and that kitters can be redeployed to value-adding activities. POUS is used for all parts that were formerly placed in the kits. With the elimination of the kitters, someone generally questions the lot sizes and so lot sizes for the two products are decreased from lots of 6 and 4 to lots of 3 and 2. Reduction in lot size, quality at the source and mistake-proofing are discussed. The use of a poka yoke device is introduced into the process and an engineering change order (ECO) introduces a new part that improves alignment of the face of the clocks. Performance improves. The main problem remaining at the end of the round is the inability of Time Wise to deliver products on time. The assignment for the next class involves (1) proposing layout changes, (2) identification and categorization of waste seen in rounds 2 and 3, (3) income statements for rounds 2 and 3, (4) specific, actionable suggestions for

improvement, (4) data and/or performance measures for quality, lead time, takt time, cycle time, capacity, distance travelled, and productivity.

Round 4

Round 4 utilizes a manufacturing cell with a well balanced flow. When the students enter the classroom to play the final round, they have analyzed all the data that has been collected including cycle times for each operation. They should know the takt time and should have determined the changes necessary to achieve takt time. They assign workers to jobs, and determine how many individuals are needed at each task. They determine the layout and set up the facility. The use of pull has been discussed in class and some students recognize that there is a disconnection between manufacturing to a forecast and shipping according to demand pull. They rarely understand or suggest the actual use of kanban cards. Using whatever layout they present, I introduce pull through the use of kanban cards. To date, all classes have been able to meet demand and have achieved nearly perfect quality with all orders shipped on-time. They are always amazed at what they are able to accomplish through the use of lean thinking. The final assignment/presentation requires that students show the change in performance measures, including income, from Round 1 to Round 4. They must provide an overview of the changes made and their impact. Finally, suggestions for continued improvement must be presented.

Interconnected Exercises

The Time Wise simulation does an excellent job of teaching lean thinking and forcing the students to think about value. Because there is a vendor in the simulation, various aspects of supply chain management are discussed during the simulation. The vendor, for example, begins with delivery to the warehouse according to a forecast provided. At the end, the vendor is a partner and is doing milk runs and delivering to point of use.

There are, however, additional topics that must be covered.   In an attempt to provide interconnections among some of the stand-alone topics found in a typical introductory operations management course, I created exercises centered on the use of the simulation. These include: 1) forecasting assignments that actually create the amount of clocks that are produced during the simulation rounds, 2) a quality improvement exercise using the PDCA cycle and data from the simulation, and 3) selection of performance measures to track improvement over the 4 simulation rounds.

LEARNING RESULTS

Both perceptions of learning and test results will be examined. The final examination included both multiple choice questions, problems in a service setting and an essay that required students to apply lean thinking to services.

Perceived Learning

At the end of the semester, students were asked to complete a survey rating their current proficiency in specific areas of lean knowledge. The following scale was used:

- o **Level 1: Have no exposure to or knowledge of**
  (have never heard of the topic, or only in casual conversation)
- o **Level 2: Have experienced or been exposed to**
  (have had some organized introduction to the topic, have had someone explain it to me)
- o **Level 3: Can participate in and contribute to**
  (can participate in and contribute to a discussion about the topic, or have participated in an event where the topic was used)
- o **Level 4: Can understand and explain**
  (have explained the topic to someone else, prepared a presentation about the topic, or written a paper about the topic)
- o **Level 5: Am skilled in the practice or implementation of**
  (have applied my knowledge in the topic by developing solutions to a case study or other academic exercise, solving a problem in an organization, or leading an activity)

Fifty-one of 69 students (74% response rate) returned the survey. The results (Table 1) indicated that the mean proficiency levels, as rated by the students, ranged from a low of 3.51 for Q14: problem-solving techniques (e.g., DMAIC, PDCA), to a high of 4.45 for Q5: the benefits of lean processes. Responses to Q14: problem-solving techniques also displayed the greatest variation

TABLE 1. Post Course Survey Results of Questions Related to Lean Knowledge

	Lean Knowledge Area	Proficiency Level		
		Mean	Std. Dev.	Coeff. Var.
Q1	Definition of lean	4.33	0.5538	12.8%
Q2	Fundamental principles of lean thinking	4.12	0.6826	16.6%
Q3	Examples that demonstrate lean thinking principles	4.02	0.6779	16.9%
Q4	Application of lean principles in different types of organizations	4.08	0.6586	16.2%
Q5	Benefits of lean processes	4.45	0.5409	12.2%
Q6	Lean terminology (e.g., takt time, kanban, value, flow, pull)	4.04	0.6917	17.1%
Q7	Process mapping, elements of a process	3.82	0.6843	17.9%
Q8	Process analysis, determining lead time, capacity, throughput	3.68	0.6528	17.7%
Q9	Relationship of lead time and WIP	3.88	0.6826	17.6%
Q10	Process improvement	4.12	0.5531	13.4%
Q11	Tools such as lot-size reduction, layout improvements, standard work	3.88	0.7388	19.0%
Q12	Kanban systems	4.14	0.7489	18.1%
Q13	Data analysis, and drawing conclusions from the analysis	4.04	0.6917	17.1%
Q14	Problem-solving techniques (e.g., DMAIC, Plan Do Check Act)	3.51	0.7582	21.6%

as measured by standard deviation and the largest relative dispersion as measured by the coefficient of variation. Q14 is the only question in the survey that received recorded responses lower than level 3 (can participate in and contribute to).

The ratings from Q14 (Figure 1) are somewhat surprising because all students participated in a PDCA Cycle related to Time Wise. The graded results of the exercise were not particularly good and perhaps the students recognized the need for additional exercises in analysis and application of problem-solving techniques. As with many team assignments, it may be that only some students actually worked on completion of the exercise while others attached their names to the graded input.

FIGURE 1. Ratings for Problem-solving Techniques
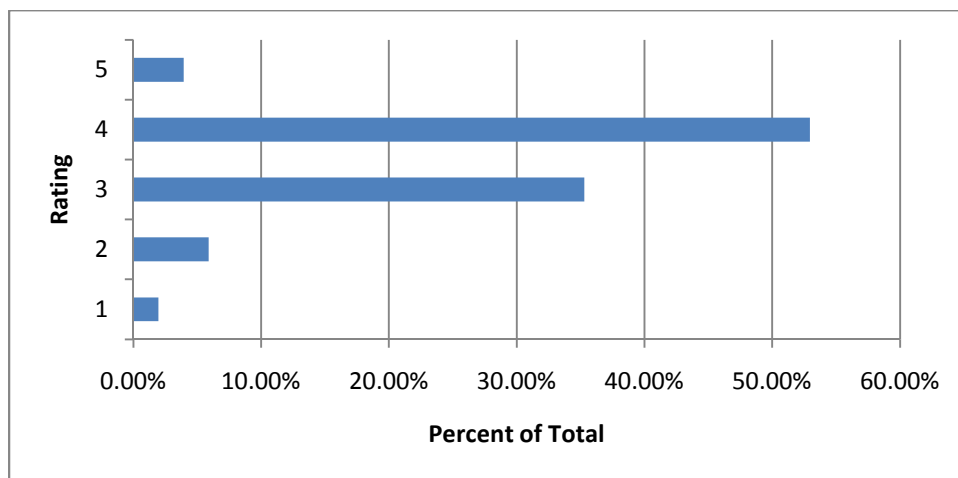
The ratings seen in Figure 2 provide evidence that students understand and can explain the benefits that can be attained through the implementation of lean processes in an organization. Ninety- eight percent of the respondents provided ratings of either 4 or 5 in response to this item.

FIGURE 2. Ratings for Benefits of Lean Processes
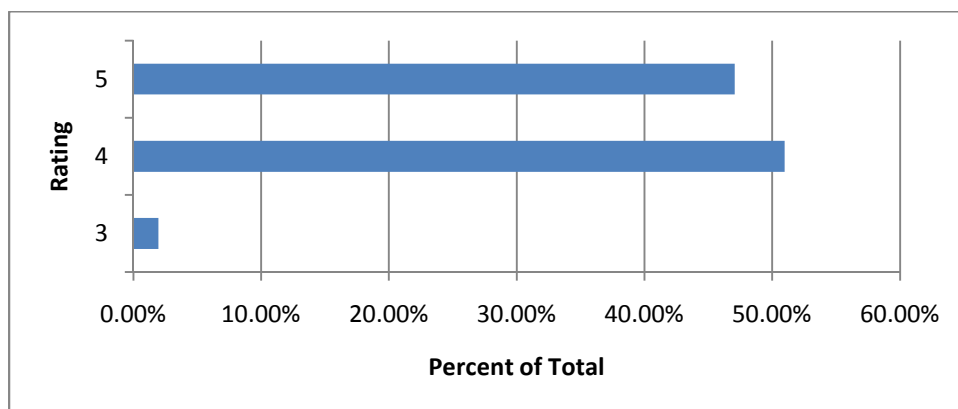
Figure 3 provides the ratings for the application of lean principles to different types of organizations. This provides a glimpse into student perceptions of their ability to transfer their lean knowledge from manufacturing to services. Eighty-two percent of the students responded that they can "understand and explain" (level 4) or they are "skilled in the practice or implementation of" (level 5).

FIGURE 3. Application of Lean Principles in Different Types of Organizations
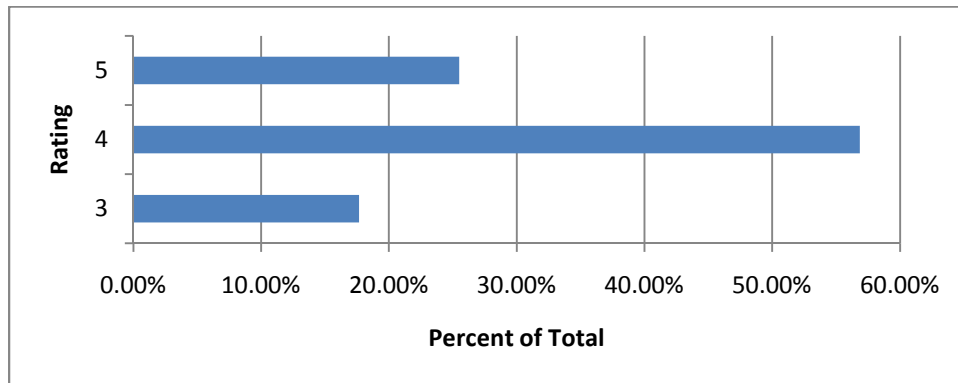
For all questions except questions 7, 8, 9, 11 and 14, at least 75% of the respondents rated their proficiency at a level 4 or 5. The mean ratings for these 5 questions were from 3.51 to 3.88, the only questions to not exceed 4.0. Each of these questions used language and terminology that seemed to require the ability to calculate and analyze something specific (e.g., process mapping, determining lead time, relationship between lead time and WIP, tools such as lot-size reduction, and DMAIC). It appears that the students have a generally positive perception of their general understanding of lean but are less comfortable with the specifics of the implementation process especially when it involves calculation using quantitative data. They rate their general knowledge of process improvement at a mean of 4.12 but are more reticent when it comes to specific improvement opportunities like lot-size reduction or use of the PDCA cycle.

Examination Results

The final examination covered several topics including lean and included multiple choice questions, problems and a take-home essay. There were 18 multiple choice questions that covered the topic of lean and the mean score on these 18 questions was 71%. There were short problems/essays that required students to specify capacity, identify a bottleneck, calculate takt time and suggest improvements that would allow a service enterprise to meet customer demand. The average scores on these problems were capacity (79%), bottleneck (84%), takt time (84%) and improvements (60%). These results suggest that students may understand the basics and may be able to apply them in a service environment but they have difficulty with completing a quantitative analysis and generating a suggestion that will lead to an improved process.

The take-home essay also required students to apply lean in a service environment. There was little discussion of lean services during the course. I did feel, however, that unless students could generalize the lean material learned that it might be a disservice to spend three class periods using the simulation. As we all know, few of our students will have a career working in a manufacturing enterprise. The great majority will be involved in a business that provides

services. Thus, it is imperative that we teach them concepts and demonstrate how they can be applied in services.

The students were given this essay as a take-home so that they would have the time to think about and reflect upon this type of application. The essay is found in Figure 4. Only 3 students (4.4%) failed to apply a category to the waste identified and 3 students (4.4%) incorrectly classified waste. After removing the essays that failed to categorize waste, 84.6% (fifty-five of 65) of the remaining essays described a corrective action that would lead to eliminating or reducing the waste identified. As Figure 5 shows, waiting and movement accounted for the majority (58.5%) of the waste identified. The top four wastes also included defects and inventory. This is not surprising since these wastes were visible in the Time Wise simulation and students identified them and found ways to reduce them. Although inventory waste played a big role in Time Wise, as expected, only 10.8% of the service waste identified was attributed to inventory. An interesting example of inventory waste was attributed to work as a lifeguard at a country club. One of the responsibilities of the job is to test the water and add needed chemicals. The student essay provides a wonderful example of the use of the 5S system (sort, set in order, shine, standardize, and sustain) to eliminate unnecessary and obsolete inventory while providing order so that lifeguards could see what chemicals were currently held in storage. This change is predicted to result in monetary savings for the club.

FIGURE 4: Take-Home Essay

Take Home Essay. This essay must be completed individually and without any assistance. It must be word processed and turned in at the final exam. (20 points)

Identify some activity with which you are familiar. It could be a service you have received, a process at Merrimack or perhaps some aspect of your everyday life. Consider how this service, process or activity could be modified to become leaner. In particular, think about waste that is part of the process as it is now done.

a) Describe this service or process. Provide a process flow diagram to help your analysis and my understanding of the process.

b) List, explain, and <u>categorize</u> one (1) example of waste in the service/activity/process.

c) Provide a potential modification to the service/activity/process that should eliminate the waste described in part (b) to make the activity leaner. Be sure to explain how this modification could be accomplished.

The Registry of Motor Vehicles (RMV) provided the setting for elimination of the waste of waiting. Customers waited for service and employees waited to use the single printer. In addition to waiting, the waste of unnecessary movement was observed as employees moved to access the single printer

FIGURE 5: Pareto Diagram for Waste Identified
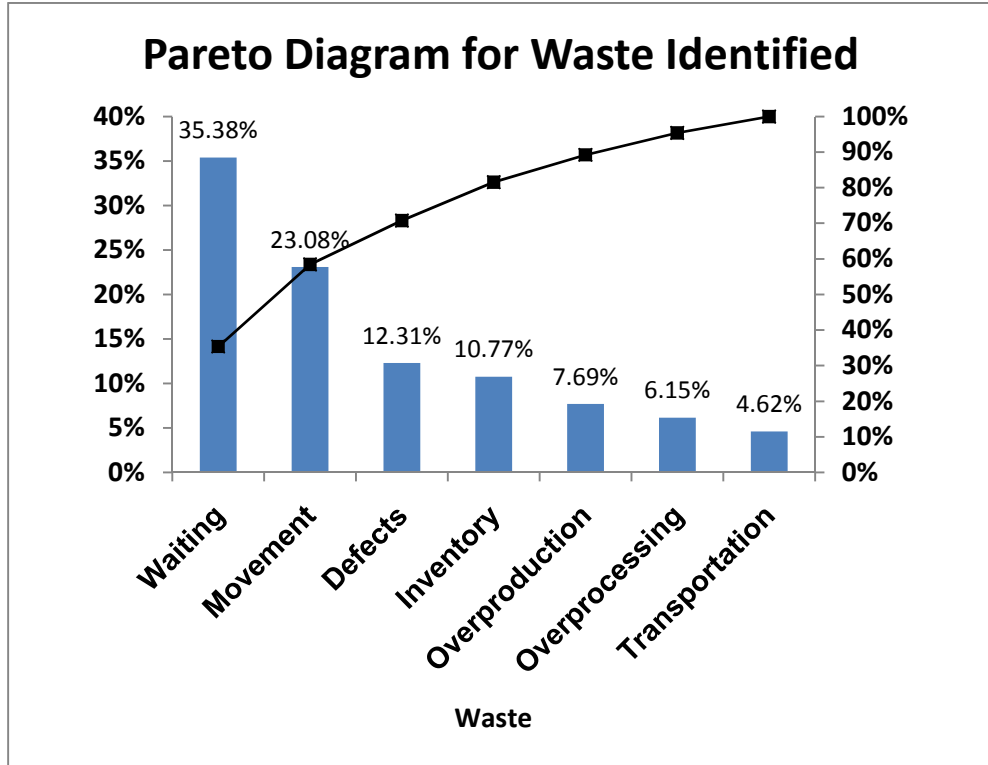
The themes of the essays varied from over-processing or non-value added processing in a college help center to over-production in various food operations to defects at a coffee shop to movement and waiting at a government repair operation, to movement in a hardscaping operation to transportation at an auto parts store. The majority of students were able to recognize waste and to provide solutions that would result in reduction of the waste. I would conclude that students were able to transfer the knowledge gained in the hands-on manufacturing simulation to the service sector. Things that students would have considered to be annoying are now seen as opportunities for improvement.

CONCLUSIONS

The Time Wise simulation allows students to see waste through direct observation and provides them an opportunity to change the manufacturing process in order to reduce and/or eliminate the waste observed. Performance measurements tracked throughout the various runs of the simulation allow students to see the results of waste elimination from an organizational perspective. Results from survey data show the students perceive they know the benefits that can be attributed to a lean implementation and that they can define lean.

Although class time was not spent discussing lean in services, the majority of students were able to transfer the knowledge gained in the hands-on lean manufacturing simulation to a service environment. The scores on the final examination problems that required the use of lean thinking in services were similar to or better than scores on the rest of the examination. Results from the final essay provide evidence that eighty-one percent of the students in the course (55/68

students) were capable of describing a service process, identifying waste, and describing actions that would reduce or eliminate the waste. Eighty-two percent of the survey respondents had high perceptions of their ability to apply lean in a different type of organization.

The overall results indicate that the application of lean thinking in services can be learned through a hands-on manufacturing simulation. The understanding and knowledge gained is not limited to manufacturing but can be transferred to a service environment.

ACKNOWLEDGEMENTS

Work in Progress: The State of Project Management in Information Systems Curriculum
Manouchehr Tabatabaei, mtabatab@georgiasouthern.edu

The history of Project Management (PM) is quite rich and extensive, but PM has gained a great deal of attention among Information Technology (IT) practitioners in recent years. PM has been with us since the beginning of civilization and finds its origins back as far as the building of the Egyptian Pyramids and the Great Wall of China. It has now found its way into various disciplines including Construction, Engineering, and Operations Management. PM is looked upon by these disciplines as one of their standard tools, and has found the use of PM as an integral part to their success.

The same importance is resonating throughout the technology community. The importance of PM can be seen in one of the newer disciplines, Information Systems (IS). IT is more pervasive in our society than ever and more organizations are using IT to add values and to gain a competitive advantage. A major need has risen in the IS community with this increased importance. IS professionals are relied on to explore the opportunities to apply technology in order to add value to business and gain a competitive edge. Thus, more systems are to be created and more projects to be managed.

PM skills are now just as important as programming skills. IT projects have a poor track record of success and many projects have failed over the years. IT projects often go over budget, past deadlines and often enough, the projects are cancelled at the loss of millions largely due to poor PM skills. According to the Standish Group's ongoing research on IT PM, their Chaos Reports states millions of dollars are frequently lost on IT projects. Their current statistics are troubling. The IT project's success rate is very low especially with large and complex project. This rate has improved in the past few years, but still is considered to be low.

Therefore, the importance of PM skills, techniques, methods, and tools has surfaced due to the enormous amount of capital invested in IT (Capell 2001, Reif & Mitri 2005), and the demand for capable team and project managers to lead information technology (IT) acquisition, development, and implementation projects is increasing. Effective project managers are in high demand and short supply. There is little question that organization demand for capable project managers has been and remains strong (Herzberg 2006, Schur 2001). In its recent round of CIO interviews, the Standish Group reported that CIOs are emphasizing the need for effective project managers. They often view IT project success directly attributable to the leadership provided by effective project managers.

As a result, Institutions of higher education are receiving steady pressure to better prepare students for PM positions and therefore the value being placed on PM courses is increasing in higher education. Many universities have made PM a required or elective course in computing and other disciplines. For example, PM courses are found in many curricula including computing. PM courses are project-based that require the completion of realistic group projects and most are designed to help students develop PM skills. Such skills contribute to decreasing problems associated with IT project failures, such as the inability to meet schedule deadlines or to stay within budget limits. The need for individuals with PM skills is unlikely to decrease with significant amount spent annually on IT projects in the U.S. (Capell 2001, Reif & Mitri 2005).

The Information Systems (IS) curricula models have recognized the need and specifically address PM courses and concepts. Therefore, PM is recognized as an important component of undergraduate and graduate IS degree programs, and has became an integral part of the IS Model Curriculum (Gorgone et al. 2005). ABET accreditation criteria also include the ability to assist in the creation of an effective project plan as one of the required skills for graduates from accredited programs (ABET 2004). In response to the inclusion of PM in IS Model Curriculum recommendations and ABET accreditation criteria, numerous programs now enable their students to take one or more PM courses. One of the earliest curricula oriented citations predicted that IS graduates would become project leaders and team members of systems development projects (Ashenhurst 1972, Couger 1973).

The short supply of project managers and the need for appropriately skilled project managers has also motivated industries to turn to the Project Management Institute (PMI) for project management training and certification programs. PMI now offers different levels of certifications based on work experience.

Further, the universities are now offering PM Graduate degrees. They are also offering Graduate Certificate program in PM. The MBA programs are also offering PM as a concentration. Many courses in these programs are offered online to make it convenience and accommodate participation of students.

PM is also tied with Enterprise Resource Planning (ERP) and Systems, Applications, and Products in Data Processing (SAP) as they are gaining worldwide popularity. ERP implementation is a challenge due to its complexity and effective implementation of an ERP system such as SAP requires an effective PM. SAP itself can guide and support PM activities.

There is an abundance of PM software available that can be used for support in planning and implementation of projects. Many also develop and used their own programs for different component of project planning and implementation. However, there is a need for tools to map the entire lifecycle of project completely and uniformly. This will help avoiding many unnecessary steps including documenting and entering data several times. Therefore an integration option will reduce time and cost which can be a great help to a successful project.

The companies that already use an SAP ERP system are now increasingly using SAP Project System (PS) to manage their projects and therefore benefit from the close integration of SAP PS with other areas including Accounting, Sales, HR, Production, and so on. SAP PS offers many functions for managing different types of projects. Most companies start by using only a small number of functions SAP PS has to offer and gradually use more functions of SAP PS as their experience using these functions increases.

The purpose of this research is to discuss the importance of PM and how it has evolved in higher education over the years specifically in IS curriculum. Further, the current status of PM in industry and higher education is reported along with its trend.

References

Ashenhurst, R., 1972, Curriculum recommendations for graduate professional programs in information systems, Communications of the ACM, 15(5), 365-398.

Cappel, J., 2001, Entry-level IS job skills: A survey of employers, The Journal of Computer Information Systems, 42(2), 76-82.

Couger, J., 1973, Curriculum recommendations for undergraduate programs in information systems, Communications of the ACM, 16(12), 727-749.

Gorgone, J., P. Gray, E. A. Stohr,, R. T. Wigand, J. S Valacich, 2005, MSIS 2006 Curriculum Preview, Communications of the Association for Information Systems, 15(30), 544-554

Herzberg, R., 2006, IT Project Managers will be in Demand in '07.
http://www.baselinemag.com/article2/0,1540,2072180,00.asp

Reif, H. and M. Mitri, 2005, Integration of Project Management Components in Undergraduate Information Systems Curricula, Journal of Computer Information Systems, 45(3), 24-31.

Schur, A., 2001, Project Managers in Peak Demand.
http://www.networkworld.com/newsletters/careers/2001/01139256.html

SUSTAINABLE EDUCATIONAL PRACTICES: THE CASE OF LIBERIA

Nekpo Brown, Bay Path College, 339 Nottingham Street, Springfield MA, 01104
nbrown75@verizon.net, (917) 279-0164
Louise Locario, Bay Path College, 4 Washington Street, Springfield MA, 01104
llocario@baypath.edu, (413) 575-1502
Dawn Notman, Bay Path College, 159 Brainard Road #37, Enfield CT, 06082
dnotman@gmail.com, (860) 394-7618

ABSTRACT

This research concentrates on development and implementation of an educational progressive academic program in Liberia. The focus of this program is not only to enhance the scholastic potentials of the Liberian youth, but also to train future industry and political leaders.

The study follows the settlement of freed slaves in 1820 to the formation of the Republic of Liberia in 1847. The research examines several factors along the social-cultural, economic and political environment of Liberia before and after the fourteen-year civil war that ended in 2003.

After such a long period of unrest, the economic and especially the education system in Liberia have been devastated. The current focus of the government of Liberia is to develop sustainable education programs that can teach the youth to hold books rather than guns.

This research proposes an education program specifically for economically disadvantaged youth with the potential of changing the mindset of these former child soldiers to overcome many obstacles and allow them to dream and realize their potentials. Through such a program, we cannot only change the future of this generation, but also create a foundation for sustainable economic prosperity by utilizing leadership training, self-worth and community service.

Keywords: Liberia, education, sustainable, leadership, community

THE TRANSFORMATIONAL EXPERIENCE OF INTERNATIONAL TRAVEL: MBA STUDENTS GO TO EUROPE

Richard W. Schrader
Bellarmine University
2001 Newburg Road
Louisville, KY 40205
rschrader@bellarmine.edu
502-272-8443

Michael R. Luthy
Bellarmine University
2001 Newburg Road
Louisville, KY 40205
mluthy@bellarmine.edu
502-272-8040

Alan B. Deck
Bellarmine University
2001 Newburg Road
Louisville, KY 40205
adeck@bellarmine.edu
502-272-8246

ABSTRACT

A multicultural experience has become a fundamental part of many MBA programs. These multicultural experiences may include specialized coursework without travel, exposure to foreign culture and business through speakers and media, or study in a foreign country. Research supports the importance of multicultural work and experiences. This paper extends the extant research by studying some of the differences in perceptions created by students' international travel. This study focuses on students who traveled to Europe (Paris, France or Madrid, Spain) to satisfy the international travel component of their MBA program. Pre-trip and post-trip surveys were administered to determine perceptions, interests and attitudes of students and to measure changes in those attributes.

BACKGROUND

An understanding of the ways in which language and customs of various countries affect that country's business practices is essential to successful business operations in that country. These

customs relate to how employees within a company interact and communicate with each other and to how customers in those countries relate to business practices. International books are replete with stories of well-intentioned but locally insensitive product launches, e.g., the Chevrolet Nova's failure in Spanish speaking countries (on the off chance that you haven't seen this story, *no va* in Spanish means "it doesn't go." This became a textbook error in branding). The best way for a student to gain an appreciation of these differences is to experience it for one's self by traveling and studying in a culture that differs significantly from their own.

This study examines an international travel experience in an American MBA program. The curriculum of this program requires an international component. Most students satisfy the international component by traveling (as a class) to a foreign country and experiencing the culture first hand. Those who choose this option may travel to Paris, France or Madrid, Spain. Some students cannot travel due to family or work commitments and satisfy the requirement by an on-campus course, but this options fails to place students out of their 'comfort zone' and is not considered to be as educational as the foreign travel experience.

For those who took the option of studying in Europe, surveys were administered immediately prior to and immediately after the travel. The survey was divided into three major sections; why the students chose the European option, their views on business practices in their destination country, and their views on the European educational experience. They were further asked if they would recommend foreign travel over the on-campus option in order to satisfy the international travel experience requirement.

The European Experience

The MBA trip to Europe covers a period of 11 days (including 1 ½ days for travel). In Spain, the students spend 5 ½ days in class at a Madrid University and 1 ½ days visiting Spanish businesses. Because of a national holiday in France, students there spend 5 days in class, one day visiting the European Union offices in Brussels, Belgium, and one day visiting French businesses. The remainder of the time can be used by students, individually or in groups, for exploring their city or visiting other cities. On this most recent trip, students visited such cities as Barcelona, Rome, and Stockholm.

When Franco was King of Spain, the use or teaching of English was forbidden. Although Spain is now a democracy, the anti-English language sentiment has persisted. As a result the students going to Spain hit strong language barriers. Students encountered these barriers in restaurants, in shopping areas and in their attempts to travel to other parts of Spain. This forced students to rely on limited Spanish vocabularies or resort to pointing. Students visiting France had less difficulty with language in Paris, as many Parisians in commercial establishments speak English, but those leaving Paris on various trips such as to Reims found that the provincial French tend to speak very little English, or at least to ignore attempts to speak in English. Attempts at high school French were, however, well received in the provinces.

The universities in Europe (ESCP/EAD) provide a series of European professors with different specialties. These professors coordinated their lectures which provided a lock-step presentation approach. Whereas the MBA program at home primarily uses a case-based approach, the professors at the European universities used a less participative, more European style lecture

approach. Although the professors indicated that student questions were welcomed, they allowed little time for questions and the answers provided did not allow for continued interaction.

LITERATURE REVIEW

Providing opportunities to understand international business practices is a fundamental responsibility of an MBA program. White & Griffith [8, p.104] suggest that critical thinking and analytic skills should be developed beyond one's "inherent biases." In addition, the internationalization of the MBA curriculum requires MBA programs to develop graduates' skills to manage business competently in the global environment [6]. The normal study abroad experience, however, often emphasizes the nuts-and-bolts of operating in a foreign country, however … how capital is acquired in Germany as opposed to the United States, for example. Varner [7, p.103] says of international experiences: "Culture expresses itself in politics, government policies, business regulations, educational systems, and business practices … One cannot separate culture from these issues and study culture in isolation".

Ortiz [5] noted that going to another country on an academic study program … is the best way to begin the process of understanding what it means to function in a global economy. Other authors [3] note that even short-term study abroad can help the student to a better understanding of another culture, as well as one's own culture in contrast. The Association to Advance Collegiate Schools of Business [1, p.6] has stated that "the length of the program must be balanced against intended results…Short term programs, when well- structured and value-added, are quite useful. While there cannot be a substitution for an extended period of work or study in a non-native culture, the awareness of cultural differences is critical for the MBA.

In light of this, many schools both in the United States and overseas have adopted some form of multicultural experience in their MBA. For example, the WHU-Otto Beisheim School of Management partners with the Kellogg school at Northwestern University and requires two weeks in the United States [2] while King, et.al. [4] describe the success of short term foreign travel MBA programs offered by St. Bonaventure University. Short term programs are particularly useful for MBA students in weekend or part time programs who are also employed full-time, and for whom lengthy foreign study is simple not possible due to family or job constraints. The university studied in the current paper.

THE SURVEY INSTRUMENT

A pre-trip survey was administered to the students at the airport prior to their departure for Europe. The survey focused on three areas; why the students chose the European option, their views on business practices in Spain/France, and their views on the European educational experience. A follow-up survey was administered at the end of the trip to determine how their pre-trip perceptions may have changed based upon their individual experiences.

Both surveys were voluntary. A total of 30 students were in each class. 30 students on the Spanish trip and 9 students on the French trip answered the pre-trip survey, and 22 students on the Spain trip and 21 students on the French trip completed the post-trip survey. Because the

purpose of the study was to examine changes in perceptions based on the trip, only the surveys of 28 students completing both surveys were used. It should be further noted that a few students may not have answered a specific question some of the questions.

The first set of questions examined which factors influenced their decision to go to Spain or France. The second set of questions examined their pre-trip expectations regarding elements of the Spanish or French travel experiences and how well these expectations were met by their actual experiences. The third set of questions (post-trip only) examined the students' comparison of the Spanish or French university educational system with that of their home University.

RESULTS

The students in the survey elected to travel to Europe rather than satisfying the international component of their MBA program through an alternative on-campus course cultural studies course. The pre-trip questions concerned the factors why they chose the Spain or France option. The responses ranged from 0 – 7, with zero being an isolated response being 'No factor at All,' 1 being 'Minimally Important and 7 being 'Highly Important.' The post-trip survey asked the respondents to how successful they the trip was related to each factor.

As shown in the following table, it appears that the international trip fell short of student expectations in the observance of business practices and in making business connections. These results are not surprising given the structure of the trip, with over half the days being spent in the classroom. Students had several days to choose between exploring cities in Europe and making business contacts. Several students did use this time to contact businessmen, but these were often prearranged by the students through current employers. The majority of the students used this time for travel in and around their host city.

The students were split on their evaluation of learning from faculty at another university. Although the mean response was .57 higher on the post-trip survey, there were nine who rated this experience higher versus 7 who rated it lower. Those who rated it higher indicated that they enjoyed the related flow of the lectures while those who rated the educational program lower generally focused on the fact that there was little open discussion between the faculty and the students.

The biggest difference noted was the practice of language skills. This was the lowest rated factor in the pre-trips survey; the rating increased the most of the four factors on the post-trip surveys. It appears that when the students were immersed into a culture where English was seldom spoken, the students were forced to expand their limited foreign language conversation skills.

EVALUATION OF FACTORS RELATED TO WHY STUDENTS CHOSE THE SPAIN OPTION				
Factor	Pre-Trip Mean	Post -Trip Mean	Number of Responses Decreasing	Number of Responses Increasing
Opportunity to Learn From Faculty at another University	4.85	5.57**	8	13
Practice My Language Skills	3.05	5.06*	6	11
See Business Practices in Europe	5.23**	4.21	16	11
Make Business Connections with Individuals	4.18	4.00	15	8

* - Paired t-test, significant at .01 level

** - Paired t-test, significant at .05 level

Views on European Business Practices

Prior to the trip, students were assigned readings on the current economic conditions in Spain or France. These conditions include an economy on the decline (more so in Spain than in France), high unemployment rates for young adults, and excellent benefits for the retired population (for example, until this fall, the retirement age in France was 60, with generous pensions). The students were also advised that very few Spanish citizens spoke English or even tried to converse in English. In Spain, they were not told that most of the citizenry working in the service and sales industries earned a fixed wage and that the level of service they provided meant nothing to them monetarily. These same topics were included as part of the curriculum at the Spanish University so that the students would understand the effects of these items on Spanish business practices. Students attending the French university were, however, advised in advance that tipping large percentages is not customary in France, and that a tip of a euro or so is considered ample in most environments. The French university did not cover such basic items in local customs.

These questions used a Likert scale ranging from 1 – 7. . For each question, the pre-trip mean and the post-trip mean are recorded and the number of students whose responses changed either in a positive or negative direction is noted. The following sections of this paper examine these responses. The tables below show the results of the pre-trip and post-trip surveys. The first table, relating to Spanish culture and business practices shows the 1, 4, and 7 points of the Likert scale along with the mean of the responses. The second table shows the comparative ranges and means from the question and indicates either shift in perception and any narrowing or expanding of the perceptions.

QUESTIONS RELATING TO FOREIGN CULTURE AND BUSINESS PRACTICES					
#	1	4	7	Pre-trip	Post-trip
1	Spanish/French employees work *considerably fewer hours per year* than their U.S. counterparts	Spanish ?French employees work *about the same hours per year* as their U.S. counterparts	Spanish/French employees work *considerably more hours per year* than their U. S. .counterparts	N = 28 Mean = 2.55	N = 28 Mean = 3.32
2	Spanish/French employees receive *considerably fewer benefits* than their U.S. counterparts	Spanish/French employees receive *about the same benefits* as their U.S. counterparts	Spanish/French employees receive *considerably more benefits* than their U.S.counterparts	N = 28 Mean = 5.00	N = 28 Mean = 4.95
3	Government regulation of business in Spain/France is *considerably less intrusive* than in the U.S.	Government regulation of business in Spain/France is *about as intrusive as it is* in the U.S.	Government regulation of business in Spain/France is *considerably more intrusive* than in the U.S.	N = 28 Mean = 4.74	N = 28 Mean = 5.02
4	Spanish/French businesses face *a lower tax burden* than businesses in the U.S.	Spanish/French and U.S. businesses *face about the same tax burden*	Spanish/French businesses face *a higher tax burden* than businesses in the U.S.	N = 28 Mean = 5.02	N = 28 Mean = 4.71
5	Spanish/French business practices are *less advanced* than in the United States	Spanish/French and U.S. business practices are *equally advanced*	Spanish/French business practices are *more advanced* than in the United States	N = 28 Mean = 3.47	N = 28 Mean = 3.04
6	The Spanish/French employees I meet will be *less customer oriented* than their U.S. counterparts	Spanish/French and U.S. business employees are *equally customer oriented*	The Spanish/French employees I meet will be *more customer oriented* than their U.S. counterparts	N = 28 Mean = 3.49	N = 28 Mean = 1.46
7	Spanish/French people I will meet in non-work settings will be *less friendly* than those in the U.S. employees	Spanish/French and U.S. people I meet in non-work settings are *equally friendly*	Spanish/French people I will meet in non-work settings will be *more friendly* than those in the U.S. employees	N = 28 Mean = 4.41	N = 28 Mean = 3.07

The questions can be further broken down by the type exposure the students had regarding the question subject. The number of hours worked (question 1) and the level of business practices (question 5) were discussed in the classroom and experienced in person as the students explored their European city and outlying areas. Employee benefits (question 2), government regulation (question 3), and tax burden (question 4) were items one does not normally observe in a tourist capacity. The students' responses to questions concerning meeting Spanish or French employees (question 6) and Spanish or French people in non-work settings (question 7) were most likely influenced by their casual observations as they ate, shopped and traveled in the area.

For questions 2 - 4, concerning employee benefits, government regulation, and the tax burden on business, there were fairly small changes in the mean responses, with the highest change being -.31 (question 4). For the pre-trip survey, they had only the suggested readings about the economic situation in Spain and France, and for the post-trip survey, they had the benefit of the

Spanish or French classroom instruction and company visits. This result provides an indication that students prepared properly for the trip by finishing the suggested readings and that these readings were fairly consistent with the information provided in the classroom.

For questions 1 and 5, concerning number of hours worked and Spanish or French business practices respectively, the change in the mean response was higher. Concerning the number of hours worked by the Spanish or French work force (question1) as compared to that in the United States, the students, in both surveys rated the Spanish or French people as working less. The difference in the mean response (.77) indicates that the perceived number of hours worked by Spanish or French employees, once given exposure to the actual activities of the employees as well as classroom discussion, were greater than what the students initially expected.   Just the opposite is true on the question concerning how business practices compare to that of the U.S. (question 5).   The mean response decreased by .43 on the two surveys which indicates that the gap concerning business practices was much greater than anticipated.

The largest changes in means from the pre-trip to the post-trip survey concerned the orientation of Spanish or French employees towards customers (question 6) and the friendliness of the Spanish or French people (question 7).   Before exposure to the foreign culture, the students thought that the Spanish people were slightly friendlier than Americans, (although the 6 in France suggested the opposite) but after exposure in everyday life, that perception turned decidedly negative (with a decrease in mean of 1.34).   The most dramatic mean change concerned the perception of Spanish business practices toward customers. The mean decreased by 1.83 indicating that the students found Spanish and French business to be far ruder to them than they originally thought.   The large drop in means for these two questions are likely tied to the fact that, in Spain, many Spanish people do not speak English and a large majority of the students did not speak Spanish. In France, there were several incidences of French speakers who obviously understood English insisting on transacting affairs in French. The American students, who are taught to be customer oriented in their business programs, may have expected Europeans to at least attempt English, especially in a business setting.   Of course, this is very dependent on the location in which students found themselves, as this was less prevalent in tourist areas in Paris

Comparison of the Spanish or French University with That of Their Home University

The third set of questions asked the students to compare their experience at the Spanish University to their home university.   Due to their nature, these questions were asked only on the post-trip Survey. Again, a 7 point Likert scale was used.

The results of the survey are shown in the table below. In general, the students appear to have been disappointed with their overall University experience in Spain or France.   Questions 8 and 10 refer to the instructional style and the overall level of education of the two schools (Spain or France vs. home).

	QUESTIONS RELATING TO THE EUROPEAN EDUCATIONAL SYSTEM			
#	1	4	7	Post-trip only
8	The class sessions held in Madrid/Paris were *less interactive* than those at my home university	The class sessions held in Madrid/Paris were *as interactive* as those at my home university	The class sessions held in Madrid/Paris were *more interactive* than those at my home university	N = 28 Mean = 2.47
9	The physical school facilities in Madrid/Paris had *fewer amenities* than those at my home university	The physical school facilities in Madrid/Paris had *about the same amenities* as those at my home university	The physical school facilities in Madrid/Paris had *more amenities* than those at my home university	N = 28 Mean = 2.96
10	The overall level of educational quality at the school in Madrid /Paris was *worse* than at my home university	The overall level of education quality at the school in Madrid/Paris was *about the same* as at my home university	The overall level of educational quality at the school in Madrid/Paris was *better* than at my home university	N = 28 Mean = 3.38

As noted previously, the professors at both the Spanish and French Universities left little time in their presentations for questions or discussion.   In contrast, questions and discussions are welcomed and encouraged at their home university (which embraces a case-based approach to teaching and learning). This major difference in pedagogical styles was evident as shown by the extremely low mean for question 8. The students rated the overall education level of the Spanish or French university lower than that of their home university (question 10), but were not as negative in this rating as they were in question 8.

The students also felt that the amenities encountered at the Spanish University (question 9) were inferior to those of their home university. This was not, however, true at the French university, which had an excellent cafeteria. From the professors' points of view, the amenities appeared to be about the same as those of the home university, other than rather improvised climate control in both locations, so it is unclear why the student evaluations in this regard were negative.

Recommendation of Spain/France to Other Students

The students were asked one further question "Would you recommend the international trip to other students over the on-campus MBA seminar? Again, a Likert scale ranging from 1 (Very Strongly NO) to 7 (Very Strongly YES) was used.   The modal response was a 7 and the mean response was 6.05 for Spain and 6.5 for France, with only one response being below 4, indicating that the international trip was perceived by the students taking the trip as far more valuable than the on-campus seminar for satisfying the international experience requirement of the MBA program. Note, however, that these students had not experienced the on-campus seminar or talked to students taking that seminar.

SUMMARY AND CONCLUSIONS

Overall, the students had mixed feelings concerning their experiences in Spain and France. For the first seven questions related to Spanish or French business and cultural behaviors, some means increased and some means decreased. A change in one direction or the other is not of major importance. What is important is that there were changes, indicating that their experience changed their views. For those questions (2 – 4) where the students were given preliminary readings or lectures regarding the Spanish or French economy, the means changed only slightly, indicating that the students were relatively well-prepared for what they learned in the classroom. For the remaining seven questions (1 and 5-10), the change in means was more dramatic. Each of these questions related to actual events encountered on the trip.

Despite the fact that students a) encountered language barriers and differences in customs, and b) did not look favorably upon the teaching style and accommodations of the Spanish or French university setting, they strongly indicated that the travel experience was superior to an on-campus seminar. Such a strong recommendation provides evidence that the international trip is successful in introducing students to different culture and customs and demonstrating how those differences influence the conduct of business in foreign countries.

Future research will explore how students with previous foreign travel respond to these questions, compared to students for whom this is the first foreign travel experience, and further compare the perceptions of students staying on campus to those traveling for their multicultural experience.

REFERENCES

[1] Association to Advance Collegiate Schools of Business International, Report of the Michigan State University Center for International Business Education and Research, 2001 roundtable. *Study abroad programs in business schools: Issues and recommendations by leading educators.* 2002, East Lansing, Michigan.

[2] Bryant, C., (2008). *Partnership with Kellogg provides International View*. Financial Times, 2008, 10, 7-12.

[3] Davis, B. J., & Redmann, D. H. *Methods of Internationalizing Business Communication Courses*. National Business Education Yearbook, 1991, 29, 53-64.

[4] King, D.L., C.J. Case, K.M. Premo, and E.D. Kallenbach. *Costs and Benefits of Foreign Educational Travel Programs*, Allied Academies International conference, Academy for Studies in International Business Proceedings; Cullowhee. 2009, 9(2), 15.

[5] Ortiz, J. *International Business Education in a Global Environment: A Conceptual Approach.* International Education Journal, 2004, 5, 255-265.

[6] Tuleja, E. A. *Aspects of Intercultural Awareness through an MBA Study Abroad Program: Going "Backstage".* Business Communications Quarterly, 2008, 71, 3, 314-337.

[7] Varner, I. *Teaching Intercultural Management Communication: Where are we? Where do we go?* Business Communication Quarterly, 2001, 64(1), 99-111.

[8] White, D.S., & Griffith, D. *Graduate International Business Education in the United States-Comparisons and Suggestions*. Journal of Education for Business, 1998, 74, 103-115.

Preparing Future Managers for Better Cultural Communication

Nawal Bonomo
Bloomsburg University
nbonomo@bloomu.edu

ABSTRACT

As borders dissolve and we are becoming more and more interconnected, it becomes crucial for managers to be able to understand the cultural differences and become effective communicators. In order to do so, facing one's stereotyping and perception of other cultures is crucial. A manager may be business savvy, but it is those cultural nuances that can make or break a deal.

This study discusses how one can teach Arabic language embedded with cultural activities that can shed light on students' stereotyping and perceptions of the Middle Eastern cultures. This paper discusses how the learning does not stop when the class ends. Teaching a language is used to open the door to more than words. The culture, music, dance, food, politics, business and history are entrenched in this style of teaching as a total experience that enable students to comprehend and appreciate other cultures while becoming aware of their own.

Computer Information Literacy Project

Nina Dini
Springfield College
ndini@spfldcol.edu

The goal of this research was to examine how university-level students process information obtained from computers and search engines. Different exercises were developed to test several aspects of information obtained through the search engines. These exercises explored the students' capability to evaluate the accuracy, objectivity, currency and the comprehensiveness of the information obtained via different search engines. Students in twelve sections of upper level computer science courses were used as the study subjects.

The results of this study indicated that students with a higher level of technological skill lacked the appropriate knowledge to sort and interpret the obtained information as for its accuracy, objectivity, currency and comprehensiveness. To address this issue, we developed and incorporated the subject of "Information Literacy" in the freshman level computer information technology courses. We developed cases that allowed students at the freshman level to evaluate, interpret and cross check the accuracy and validity of the information obtained through the search engines.

For the past two year since we started the "Information Literacy" cases, we have been able to enhance not only the capability of our students to interpret, sort and utilize the obtained information in a much more accurate and appropriate way, but also we have seen a major improvement in our students' communication and written skills.

This study proposes that the developed cases can be used across different courses to improve and enhance our students' skills in evaluating, and interpreting the information obtained via different databases which in turn can improve their communication and written skills.

CRITICAL THINKING SKILLS AS A PREDICTOR IN SELECTING ACCOUNTING AS A MAJOR IN ACCOUNTING AND OTHER BUSINESS DISCIPLINES

Nathan Slavin, Department of Accounting, Taxation, and Legal Studies,
Hofstra University, Hempstead, NY, (516) 463-5690, Nathan.Slavin@Hofstra.edu
Deb Sledgianowski, Department of Accounting, Taxation, and Legal Studies,
Hofstra University, Hempstead, NY, (516) 463-4759, Deb.Sledgianowski@Hofstra.edu
William E. Bealing, Jr., Department of Accounting,
Bloomsburg University, 570-389-4386, WBealing@bloomu.edu

ABSTRACT

The purpose of this research-in-progress is to determine whether there is a relationship between critical thinking skills, personality traits, and the selection of accounting as a college major and other areas of study within business such as finance, information technology, management, and marketing.

Approximately 170 students enrolled in the introductory course in Managerial Accounting at an AACSB accredited university will be assessed on their performance on examination materials that specifically examine critical thinking skills. Additionally, the students will complete the Keirsey and Bates [6] personality preference test which classifies individuals into 16 personality types. Univariate tests and multinomial logistic regression will be used to test our hypothesis that critical-thinking skills are significantly associated with major selection.

Keywords: Accounting

LITERATURE REVIEW

There are several studies that have attempted to examine the determinants and predictors of selecting accounting as a field of specialization and ultimately entering the accounting profession. Specifically, Bealing et.al. [3] have examined various personality traits as a predictor in selecting accounting as a major. The researchers used the Keirsey and Bates [6] personality preference test as a predictor of selecting accounting as a field of study. In subsequent studies presented at the NEDSI conferences, Baker et.al. [2] expanded their research where they postulated that personality traits as well as grades received in a student's first accounting course would serve as a predictor of performance in other business core courses. In their most recent paper, the issue of critical thinking skills was suggested as a possible predictor of performance in future accounting courses. It was suggested that critical thinking skills can readily be assessed using examination materials provided by publishers of accounting textbooks. This purpose of this research is to determine whether there is a relationship between critical thinking skills and selection of accounting and other areas of study within business such as finance, information technology, management, and marketing.

The AICPA Core Competency Framework for Entry into the Accounting Profession [EAP] [1] promotes a list of core competencies required for students to successfully enter the accounting profession. Critical thinking is one of the broad business perspective category items. The following description of critical thinking is provided by the AICPA EAP framework database:

Critical thinking encompasses the ability to link data, knowledge, and insight together from various disciplines to provide information for decision-making. Being in tune with the "big picture" perspective is a necessary component for success. Individuals entering the accounting profession should be able to communicate to others the vision, strategy, goals, and culture of organizations.

Given the AICPA's emphasis on developing critical thinkers, surprisingly there has been little research to-date on critical thinking in accounting education. Research has focused on (a) assessing the degree of critical thinking skills in students who were already immersed in the accounting program as accounting majors and on (b) predicting students who may do well in an accounting principles class based on their level of critical thinking skills. For example, Jenkins [4] found a significant relationship between critical thinking skills and performance in an advanced auditing course and Kealey et al [5] found that after controlling for academic aptitude, critical thinking skills significantly explain variation in student performance in an accounting principles course. Kealey et al suggest that understanding the relationship between critical thinking and success in accounting may promote critical thinking awareness in preparing students to become accounting professionals.

Our research-in-progress attempts to combine these three research streams [3, 4, 5] to identify a model (see Exhibit 1) to determine the strength of the relationship between personality traits and degree of critical thinking skills to predict students who will major in accounting.
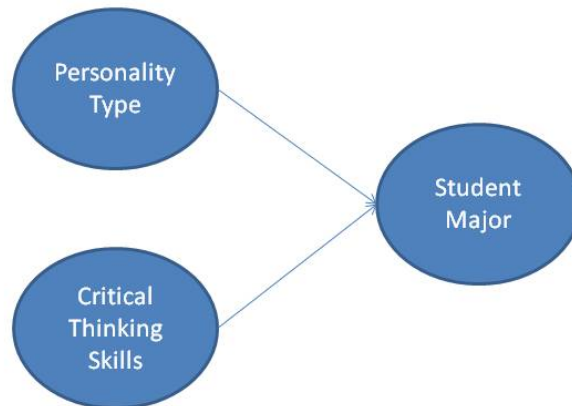
Exhibit 1. Predictors of Selection of Accounting as a Major

Our research questions are as follows:

Research Question #1: What is the strength of the relationship between personality type and critical thinking skills in predicting student major?
Research Question #2: Does the degree of critical thinking skills of Accounting majors differ from those of other business majors?

RESEARCH METHODOLOGY

Students enrolled in the introductory course in Managerial Accounting at an AACSB accredited university will complete the Keirsey and Barber personality preference test known as the Keirsey Temperament Sorter (KTS) which classifies individuals into 16 personality types. In addition, students will also be examined on their performance on examination materials that specifically examine critical thinking skills. There are currently three sections of Managerial Accounting consisting of approximately 170 students assigned to three instructors. The course is specifically designed as a service course for business majors other than accounting. Accounting students are required to register for the course. During the course, the students will be given three examinations, first exam, midterm and a final. All of the examination will contain objective multiple choice questions that will be coded for critical skills as well as for other skills such as conceptual reasoning. It is suggested that controlling for personality traits, performance on critical thinking skills will influence the selection of a student's selection of a business major. Specifically, it is anticipated that there will be a strong relationship between critical thinking skills and the selection of accounting and finance as a major field of study. Univariate tests and multinomial logistic regression will be used to test our hypothesis that critical-thinking skills are significantly associated with major selection. Multinomial logistic regression is a type of logistic regression used when dependent variables are nominal with multiple response levels having no specific order

PROGRESS OF RESEARCH

We anticipate to have the data collected and a preliminary analysis to report at the 2011 NEDSI conference. A more complete literature review will be available to be published in the proceed should this research-in-progress manuscript be accepted for presentation at the Montreal conference.

ACKNOWLEDGEMENTS

REFERENCES

[1] American Institute of Certified Public Accountants (AICPA). "The AICPA Core Competency Framework for Entry into the Accounting Profession", last accessed November 22, 2010 at: http://www.aicpa-eca.org/library/ccf/default.asp?

[2] Baker, R. L., Bealing, W. E., and Pitingolo, E.. "Does Performance in a Student's First Accounting Course Provide Insight Into Performance in Other Business Core Courses?" *Proceedings of the Northeast Decision Sciences Institute 2010 Annual Meeting and Conference,* Alexandria, VA.

[3] Bealing, W. E., Baker, R. L., and Russo, C. J. "Personality: What it takes to be an accountant", The Accounting Educators' Journal, 2006, 16, pp. 119-128.

[4] Jenkins, E. "The significant role of critical thinking in predicting auditing students. performance. Journal of Education for Business, 1998, (May/June), pp. 274-280.

[5] Kealey, B. T., Holland, J., and Watson, M. "Preliminary Evidence on the Association between Critical Thinking and Performance in Principles of Accounting", Issues in Accounting Education, 2005, 20(1), pp. 33-49.

[6] Keirsey, D and Bates, M. "Please Understand Me: Character & Temperament Types", Del Mar, CA: Gnosology Books, Ltd., 1978.

ASSESSING DIFFERENCES IN FACULTY COMPLIANCE WITH B-SCHOOL LEARNING GOALS IN UNDERGRADUATE CORE-REQUIRED COURSES

Mark L. Berenson
Department of Management & Information Systems
Montclair State University, Montclair, NJ 07043, 973-655-6857
berensonm@mail.montclair.edu

Ore Fasehun
Department of Management & Information Systems
Montclair State University, Montclair, NJ 07043, 973-655-4335
talktoore@gmail.com

ABSTRACT

In April 2003 AACSB International revised its accreditation standards and stipulated that all institutions preparing for reaccreditation would need to address these new guidelines. For the School of Business at Montclair State University, this meant that it would be necessary to assess how the fundamental core-required undergraduate courses offered by the various departments met the educational philosophy and learning goals articulated in its strategic charter. To that end, the School of Business' standing Learning Goals and Assessment Committee developed a questionnaire that would enable core course faculty to self-assess their compliance with the educational philosophy and these stipulated learning goals. This study is an inferential assessment of faculty-reported compliance in addressing the 24 learning goals comprising five educational constructs stipulated in the strategic charter.

INTRODUCTION

In April 2003 AACSB International revised its accreditation standards and stipulated that all institutions preparing for reaccreditation would need to address these new guidelines [1] [2]. For the School of Business at Montclair State University, this meant that it would be necessary to assess how the fundamental core-required undergraduate courses offered by the various departments met the educational philosophy and learning goals articulated in its strategic charter. To that end, the School of Business' standing Learning Goals and Assessment Committee developed a questionnaire that would enable core course faculty to self-assess their compliance with the educational philosophy and these stipulated learning goals. In addition to facilitating the eventual development of the appropriate AACSB International self-study reports required for the reaccreditation process, a more immediate intent of the study findings would be to monitor compliance and enable both department coordinators of the multi-section core course offerings and department chairs to assist individual faculty in making improvements in their pedagogical assessments of learning.

This paper provides a thorough analysis of the questionnaire responses obtained by the Learning Goals and Assessment Committee from the full-time core-required course faculty (at the end of the fall 2003 semester in order to demonstrate how the findings could lead to a reassessment of

curriculum offerings and topic alignment. While shown here as an enumerative study or "snapshot" conducted at one moment in time for a particular institution of higher education, the demonstrated process methodology is much broader and could be employed at other colleges and universities as an analytic study or "motion picture" over time with program reevaluations provided semester-after-semester to monitor trends and to take corrective actions to "close the loop" of continuously enhancing curriculum and improving learning.[1]

This paper contains three additional sections. The Methodology section discusses the questionnaire developed for faculty self-assessment of compliance. The Findings section provides the results and analyses at the aforementioned institution. The section on Discussion and Implications, Conclusions and Recommendations highlights key points and offers suggestions that enable the process to be adapted at other institutions of higher education.

METHODOLOGY

Introduction and Goals

As part of its strategic charter, the *vision statement* from the School of Business at Montclair State University was to graduate students who are immediately effective in cutting-edge business organizations. Moreover, the *mission* was to prepare students for entry and mid-level managerial positions, entry-level professional specializations, and entrepreneurial roles, and to equip them with the capacity to identify their own goals and to proactively manage their life-long career prospects. In addition, the educational philosophy of the strategic charter was to collectively prepare the students for the aforementioned roles prior to their graduation by assuring their mastery of five educational constructs relating to desired learning goals:

- Thinking Skills – 8 learning goals
- Discipline-Specific Knowledge and Competencies – 3 learning goals
- Communication Skills – 5 learning goals
- Change Management – 5 learning goals
- Self-Development – 3 learning goals

Although these learning goals could be assessed, more or less, in every core-required course, some pertain more directly to the discipline-specific program concentration courses.

As the School of Business began its preliminary deliberations in preparation for AACSB International reaccreditation its Learning Goals and Assessment Committee was tasked with preparing an initial outcomes assessment analysis of the fundamental core-required courses offered by the various departmental disciplines.

---

[1] It is interesting to point out that at the time this study was initially conducted the administration was interested in a preliminary, descriptive analysis of results for planning purposes. Subsequently, in 2009, based on several other studies and assessments, the School of Business thoroughly revised its Educational Philosophy and Learning Goals articulated in an updated Strategic Charter. The School of Business was reaccredited by AACSB International in October 2009.

Development

To obtain the needed information for this specific study, a 24-question Undergraduate Core-Required Course Assessment Questionnaire, each with two parts (see Appendix A1), was developed. This questionnaire was designed to obtain information as to whether – or how – faculty cover any, some, or all of the learning goals affiliated with these five educational constructs in the fundamental core courses. In addition, the questionnaire was designed to ascertain how faculty assess student mastery of the various learning goals. The educational construct Thinking Skills was addressed in eight learning goal questions, Discipline Specific Knowledge and Competencies was addressed in three learning goal questions, the educational constructs Communication Skills and Change Management were each addressed in five learning goal questions, and Self Development was addressed in three learning goal questions. Each of these 24 learning goal questions provided a potential for fifteen different faculty-reported methods of assessment – student response to questions, debate, oral exam, oral report, quiz, written exam, written report, essay, memo, Email correspondence, student portfolio, case analysis, problem sets, team project and "other" – for each core course section offered.

Scope: Outcome Assessment Method Employed

There are several ways to study the self-reported faculty compliance with the specific learning goals. The Learning Goals and Assessment Committee chose to measure proficiency on an individual core-course section basis, although class size and number of sections offered per course differed somewhat over the thirteen core-required courses. This approach was taken to provide a unit of measurement in consistent format for all analyses.

The Study: A Three-Phase Outcome Assessment Analysis

This outcomes assessment study, using the data obtained in December 2003 from the survey of full-time faculty teaching undergraduate core-required courses at that point in time, is conducted in three distinct phases. Phase III was presented at NBEA in October 2010 [3] and Phase I was presented at DSI in November 2010 [4]. This paper discusses Phase II of the overall study.

Phase II is an inferential examination of ratings on a 0 to 5 scale indicating how well faculty self-assess compliance with the 24 learning goals that comprise the five educational constructs in the strategic charter.

The following research hypotheses are examined:

II A: There is evidence of a significant difference among the five educational constructs with respect to faculty self-assessed mean ratings of compliance with learning goals over the thirteen core required courses.

II B: There is evidence of a significant difference among the five educational constructs with respect to the coefficients of variation or relative variability in the compliance ratings of learning goals over the thirteen core-required courses

The analyses provided herein are intended to enable an assessment by School of Business administrators that will result in improvements in overall core-required course delivery and improvements in compliance with the 24 learning goals.

FINDINGS: PHASE II RESULTS

Phase II of the current study examines *how successful* the faculty claim they are in teaching to the five educational constructs. This will be investigated two ways – by comparing their mean ratings and by comparing the coefficients of variation in the ratings. The latter comparisons enable an assessment of variability in the ratings relative to the size of the respective mean ratings. Higher mean ratings are deemed better and lower coefficients of variation are deemed better.

How Core-Course Faculty Self-Assess Compliance with the Learning Goals Comprising the Five Educational Constructs in the School of Business Charter – The Mean Ratings
For this enumerative study or "snapshot in time," conducted at the end of the fall 2003 semester, Table 1 contains the 13 by 5 matrix of the mean faculty ratings (and their ranks) over each of the five educational constructs for each of the thirteen core required courses. From this table, the first research hypothesis (II A) is addressed.

Table 1: Data Comparing Mean Ratings (and Ranks) of Five Educational Constructs Based on Thirteen Core Required Courses

The rankings of mean ratings by core course are tabularized as follows (1 = highest mean rating of educational construct, 5 = lowest mean rating of educational construct).

Course\EdContruct	Thinking Skills		DiscSpec K&C		Commun Skills		Change Mgmt		Self Developmt		Course Means	
ECON 101	3.13	(3)	4.07	(1)	2.44	(5)	2.76	(4)	3.20	(2)	3.03	(11)
ECON 102	3.40	(2)	4.11	(1)	3.20	(3)	2.60	(5)	3.00	(4)	3.23	(7)
ACCT 201	3.32	(2)	4.33	(1)	3.11	(3)	2.20	(5)	2.67	(4)	3.09	(9)
ACCT 202	2.89	(2)	3.81	(1)	2.69	(3)	1.97	(5)	2.38	(4)	2.71	(13)
BSLW 264	3.96	(1)	2.67	(3)	2.53	(4)	3.47	(2)	2.00	(5)	3.15	(8)
INFO 270	3.56	(2)	4.28	(1)	2.53	(3)	0.97	(5)	2.11	(4)	2.72	(12)
MGMT 231	3.63	(3)	3.11	(4)	4.03	(1)	3.87	(2)	0.00	(5)	3.24	(6)
MKTG 240	3.08	(2)	3.07	(3)	2.60	(5)	3.56	(1)	2.87	(4)	3.05	(10)
FINC 321	4.38	(1)	4.00	(2)	2.80	(4)	3.20	(3)	1.33	(5)	3.38	(3)
INFO 371	3.50	(3)	3.73	(2)	3.48	(4)	3.84	(1)	3.07	(5)	3.54	(2)
INFO 375	4.00	(1)	3.92	(2)	3.65	(3)	2.55	(4)	2.00	(5)	3.36	(4)
MKTG 346	3.94	(3)	3.50	(5)	3.85	(4)	4.35	(1)	4.08	(2)	3.97	(1)
MGMT 439	4.08	(1)	3.40	(4)	3.56	(3)	3.64	(2)	0.13	(5)	3.30	(5)
Construct Means	3.50	(2)	3.73	(1)	3.13	(3)	2.88	(4)	2.25	(5)		
Rank Totals		26		30		45		40		54		

The following results are noted:

- **Thinking Skills**: Faculty indicated that the eight learning goals in this educational construct were most successfully complied with in FINC 321, MGMT 439, INFO 375 and in BSLW 264. Lower compliance with these learning goals occurred in ECON 101, MKTG 240 and in ACCT 202.
- **Discipline Specific Knowledge and Competencies:** Faculty indicated that the three learning goals in this educational construct were most successfully complied with in

ACCT 201 and in INFO 270. Lower compliance with these learning goals occurred in MGMT 231, MKTG 240 and in BSLW 264.

- **Communication Skills:** Faculty indicated that the five learning goals in this educational construct were most successfully complied with in MGMT 231 and in MKTG 346. Lower compliance with these learning goals occurred in INFO 270, BSLW 264 and in ECON 101.
- **Change Management:** Faculty indicated that the five learning goals in this educational construct were most successfully complied with in MKTG 346, MGMT 231 and in INFO 371. Lower compliance with these learning goals occurred in ACCT 201, ACCT 202 and in INFO 270.
- **Self Development:** Faculty indicated that the three learning goals in this educational construct were most successfully complied with in MKTG 346. Lower compliance with these learning goals occurred in FINC 321 and in MGMT 439 and not at all in MGMT 231. Interestingly, none of the faculty teaching the six sections of MGMT 231 believed it applicable to assess the three self development learning goals in that course.

To test whether there is evidence of a difference among the five educational constructs with respect to faculty self-assessed compliance with learning goals over the thirteen core required courses, the Friedman rank test with Nemenyi comparisons as well as an alternative Kendall coefficient of concordance are used [5]. These permit an assessment of significance of differences in faculty-reported compliance as well as patterns of association among the five educational constructs over the thirteen core-required courses.

Using Friedman's rank test with a 0.05 level of significance, there is a highly statistically significant difference in faculty self-assessment of compliance with the learning goals comprising the five educational constructs (Fr = 15.75 > χ^2 (4 *d.f.*) = 9.49 with *p*-value < 0.005).

Employing Nemenyi's *a posteriori* comparisons [6] with an experimentwise error rate of 0.05 it is concluded that faculty self-assessment of compliance with the learning goals for the educational constructs Thinking Skills and Discipline Specific Knowledge and Competencies were significantly higher than that for the learning goals associated with Self Development. All other pairwise differences in self-assessed compliance between educational constructs were due to chance.

Viewing this data set from another lens, Kendall's coefficient of concordance, W, represents the degree of association on a scale from 0 to 1 (the higher the better) among the rankings given to the mean self- assessed compliance ratings of the five educational constructs for the thirteen core courses. From the above table, W = 0.303, and this indicates a highly significant degree of association between the core courses and the construct rankings (*p*-value < 0.005).

How Core-Course Faculty Self-Assess Compliance with the Learning Goals Comprising the Five Educational Constructs in the School of Business Charter – Coefficient of Variation in Ratings
Table 4 contains the 13 by 5 matrix of the coefficient of variation in faculty ratings (and their ranks) over each of the five educational constructs for each of the thirteen core required courses. From this table, the second research hypothesis (II B) is addressed.

Table 2: Data Comparing Coefficient of Variation in Percentage Ratings (and Ranks) of Five Educational Constructs Based on Thirteen Core Required Courses

The rankings of the coefficients of variation in percentage ratings by core course are tabularized as follows (1 = lowest percentage rating of educational construct, 5 = highest percentage rating of educational construct).

Course\EdContruct	Thinking Skills		DiscSpec K&C		Commun Skills		Change Mgmt		Self Developmt	
ECON 101	49.2%	(2)	25.3%	(1)	73.0%	(5)	56.5%	(3)	58.1%	(4)
ECON 102	46.2%	(2)	34.3%	(1)	47.5%	(3)	72.7%	(5)	69.7%	(4)
ACCT 201	50.0%	(2)	27.7%	(1)	61.7%	(3)	87.3%	(5)	75.3%	(4)
ACCT 202	58.5%	(2)	39.4%	(1)	71.0%	(3)	83.8%	(5)	79.0%	(4)
BSLW 264	36.9%	(1)	74.9%	(3)	89.3%	(4)	46.1%	(2)	97.0%	(5)
INFO 270	43.3%	(2)	25.0%	(1)	67.2%	(3)	119.6%	(5)	87.2%	(4)
MGMT 231	48.2%	(4)	58.2%	(5)	31.5%	(2)	39.5%	(3)	Undefined	(1)
MKTG 240	49.7%	(2)	60.9%	(3)	76.9%	(4)	38.2%	(1)	77.7%	(5)
FINC 321	11.9%	(1)	25.0%	(2)	92.5%	(4)	55.9%	(3)	173.7%	(5)
INFO 371	28.3%	(2)	55.0%	(5)	38.2%	(4)	26.8%	(1)	28.7%	(3)
INFO 375	37.0%	(2)	31.6%	(1)	48.2%	(3)	61.6%	(4)	77.0%	(5)
MKTG 346	27.4%	(1)	55.1%	(5)	46.5%	(3)	35.2%	(2)	47.3%	(4)
MGMT 439	31.1%	(1)	42.6%	(2)	44.4%	(3)	51.4%	(4)	400.0%	(5)
Rank Totals		24		31		44		43		53

The following results are noted:

- **Thinking Skills**: The least amount of variability relative to the mean faculty compliance ratings for the eight learning goals in this educational construct was observed in FINC 321, MKTG 346 and in INFO 371. The most relative variability in the ratings of compliance with the learning goals occurred in MGMT 231, ECON 101, MKTG 240, ACCT 201 and in ACCT 202.

- **Discipline Specific Knowledge and Competencies:** The least amount of variability relative to the mean faculty compliance ratings for the three learning goals in this educational construct was observed in INFO 270, FINC 321, ECON 101 and in ACCT 201. The most relative variability in the ratings of compliance with the learning goals occurred in INFO 371, MKTG 346, MGMT 231, MKTG 240 and in BSLW 264.

- **Communication Skills:** The least amount of variability relative to the mean faculty compliance ratings for the five learning goals in this educational construct was observed in MGMT 231 and in INFO 371. The most relative variability in the ratings of compliance with the learning goals occurred in BSLW 264 and in FINC 321.

- **Change Management:** The least amount of variability relative to the mean faculty compliance ratings for the five learning goals in this educational construct was observed in INFO 371, MKTG 346, MKTG 240 and in MGMT 231. The most relative variability in the ratings of compliance with the learning goals occurred in ACCT 201, ACCT 202 and in INFO 270.

- **Self Development:** The least amount of variability relative to the mean faculty compliance ratings for the three learning goals in this educational construct was observed in INFO 371. The most relative variability in the ratings of compliance with the learning goals occurred in BSLW 264, FINC 321 and in MGMT 439. Interestingly, none of the faculty teaching the six sections of MGMT 231 believed it applicable to assess the three self development learning goals in that course.

To test whether there is evidence of a difference among the five educational constructs with respect to the coefficients of variation or relative variability in the compliance ratings of learning goals over the thirteen core-required courses, the Friedman rank test with Nemenyi comparisons as well as an alternative Kendall coefficient of concordance is again used. These permit an assessment of significance of differences in the relative variability in the faculty-reported compliance as well as patterns of association among the five educational constructs over the thirteen core-required courses.

Using Friedman's rank test with a 0.05 level of significance, there is a highly statistically significant difference in relative variability in faculty self-assessment of compliance with the learning goals comprising the five educational constructs (Fr = 17.11 > χ^2 (4 *d.f.*) = 9.49 with *p*-value < 0.005).

Employing Nemenyi's *a posteriori* comparisons with an experimentwise error rate of 0.05 it is concluded that the relative variability in the mean ratings of faculty self-assessment of compliance with the learning goals for the educational constructs Thinking Skills and Discipline Specific Knowledge and Competencies were significantly lower than that for the learning goals associated with Self Development. All other pairwise differences in self-assessed compliance between educational constructs were due to chance.

Viewing this data set from another lens, Kendall's coefficient of concordance, W, represents the degree of association on a scale from 0 to 1 (the higher the better) among the rankings given to the relative variability in the self- assessed compliance ratings of the five educational constructs for the thirteen core courses. From the above table, W = 0.329, and this indicates a highly significant degree of association between the core courses and the construct rankings (*p*-value < 0.005).

The above inferential analyses are intended to enable an assessment by School of Business administrators that will result in overall core-required course delivery and improvements in compliance with the 24 learning goals.

DISCUSSION AND IMPLICATIONS, CONCLUSIONS AND RECOMMENDATIONS

Given that the thirteen core-required courses are offered across all four levels (100 – freshman to 400 – senior) and stress different interests and skills (ACCT 201, ACCT 202, INFO 270, INFO 375 and FINC 321 are quantitatively oriented, INFO 371 is technologically oriented, and ECON 101, ECON 102, BSLW 264, MGMT 231, MGMT 439, MKTG 240 and MKTG 346 are qualitatively oriented) it should not be surprising that the above findings demonstrate some significant differences in assessment of compliance with the 24 learning goals. In addition, and aside from "academic freedom," given that faculty have different teaching styles and have different thoughts on topic focus, it should not be surprising to find some differences in self-reported assessments among faculty teaching various sections of the same core-required course. Nevertheless, it is the contention here that in order to assure a successful, cohesive B-school program, overall curriculum goals must dominate individual course goals, and it becomes absolutely essential that the latter dominate individual goals and interests. When lower-level core-required courses are intended to be prerequisite to higher-level core-required offerings, a

successful, cohesive B-school program can only be had if the individual core-course faculty members agree that responsibility to overall program supersedes "academic freedom" in the overall selection of course topics and level/depth/breadth of topic coverage. In short, department chairs and/or multi-section course coordinators must be able to monitor course delivery and compliance with learning goals in order to create more uniformity in section offerings and thereby "close the loop" in the assurance of learning process.

The process of assessing the achievement of learning goals comprising the five educational constructs proscribed in the strategic charter could have been accomplished by the thirteen core-course coordinators. Such an approach, however, would have several disadvantages. Without responses from the individual faculty it would be impossible to articulate the communalities and variations in the individual 64 course sections. Another drawback to such an approach is that the coordinator would be assessing what is *supposed to be taught* rather than a fuller analysis of what faculty state *is actually taught*. There is a difference in what is supposed to be and what actually is.

The major advantage to asking individual faculty for self-assessments of compliance is that it provides the opportunity for the chair and/or course coordinator to try to apply peer pressure on some maverick faculty and enable truer course coordination among core courses that are taught in multi-sections. This will foster the articulation of a more cohesive, integrated body of knowledge through the formulation of more connections among courses and an integration of tools and learning ideas which result in the students' more complete business education where they can be expected to have the best set of skills, abilities, backgrounds and understanding to leave the university best prepared to take their places in the business world.

REFERENCES

[1]   AACSB. *Assurance of Learning Standards: An Interpretation – An AACSB White Paper*, AACSB International Accreditation Coordinating Committee and AACSB International Accreditation Quality Committee, November 20, 2007. http://www.aacsb.edu/publications/papers/accreditation/assurance-of-learning.pdf

[2]   AACSB. *Overview of Assessment*. Assessment Resource Center, November 20, 2007. http://www.aacsb.edu/resource_centers/assessment/overview-expectations.asp

[3]   Berenson, M. L. & Fasehun, O. *Assessing How Faculty Assess Compliance in Addressing Mandated B-School Learning Goals: A Correlation Analysis.* Proceedings of the Northeast Business and Economics Association (NBEA) 37th Annual Conference, Morristown, NJ, October 2010.

[4]   Berenson, M. L. & Fasehun, O. *Descriptive and Inferential Assessment of Faculty Compliance in Addressing Mandated B-School Learning Goals*. Proceedings, 41st Annual Meeting of the Decision Sciences Institute, San Diego, CA, November 2010.

[5]   Marascuilo, L. A. & McSweeney, M. Nonparametric and Distribution-Free Methods for the Social Sciences. Monterey, CA: Brooks/Cole, 1977.

[6]   Nemenyi, P. *Distribution-Free Multiple Comparisons*, Unpublished Ph.D. Dissertation, Princeton University, June 1963.

Appendix A1:   Undergraduate Core-Required Course Assessment Questionnaire Development

<u>EDUCATIONAL CONSTRUCT and Learning Goals</u>

- **THINKING SKILLS**
    1.   **The capacity to exercise logical, critical, and integrated analysis.**
    2.   **The ability to use good judgment.**
    3.   **The ability to be creative, non-traditional problem solvers.**
    4.   **Proficiency in ethical reasoning.**
    5.   **The ability to identify sources of relevant information and discern good and useful information from bad and irrelevant details.**
    6.   **The ability to identify problems that need resolution, design alternative solutions to the problem, establish solution selection criteria, and identify ways to implement the solution.**
    7.   **The ability to challenge presuppositions, explore alternative perspectives, transform old ways of thinking, and act on new perspectives.**
    8.   **The ability to use thinking skills in a collaborative team setting.**
- **DISCIPLINE-SPECIFIC KNOWLEDGE AND COMPETENCIES**
    9.   **Knowledge and comprehension of the history and current practices in the specific discipline.**
    10.   **Utilization of information technology.**
    11.   **Enhancement of quantitative skills for problem-solving.**
- **COMMUNICATION SKILLS**
    12.   **An understanding of the importance of the role of communication in the business world.**
    13.   **A familiarity with the wide variety of communications modalities available and the appropriate conditions for their use.**
    14.   **Enhancement of oral presentation skills.**
    15.   **Enhancement of written presentation skills.**
    16.   **Enhancement of distance communication skills.**
- **CHANGE MANAGEMENT**
    17.   **An understanding of trends – global, demographic, psychographic, economic, technological, etc. -- that shape our society.**
    18.   **Utilization of numerous sources to understand the impact of societal trends on business practices, business opportunities, and/or employment opportunities.**
    19.   **An understanding and shaping forces of change.**
    20.   **A focus on globalization.**
    21.   **Ability to formulate and evaluate strategies for competitive advantage.**
- **SELF-DEVELOPMENT**
    22.   **Capacity to engage in self-management of lifelong learning.**
    23.   **Capacity to achieve continuous professional growth.**
    24.   **Capacity to achieve continuous personal growth.**

Full-time faculty teaching core-required undergraduate courses were asked to fill out a questionnaire containing 24 questions, each with two parts, that pertain to the above learning goals. Each question was worded the same. Below is the two-part question for Learning Goal 1.

1a.   On a scale from 0 to 5 *("none"* to *"very much"*), how would you rate your focus on providing the student with **the capacity to exercise logical, critical, and integrated analysis** in this course?

                              0   1   2   3   4   5

Please elaborate how you do this_____

1b.   In this course how do you assess student competency in **the capacity to exercise logical, critical, and integrated analysis**?

(*Check all that apply*)

N/A__   No Assessment__   Response to Questions__   Debate__   Oral Exam__   Oral Report__
Quiz__   Written Exam__   Written Report__   Essay__   Memo__   Email Correspondence__
Student Portfolio__   Case Analysis__   Problem Sets__   Team Project__
Other_____(*Specify*)

COMPARING TRADITIONAL TEACHING METHODS WITH VARIABLE CERTIFICATION REQUIREMENTS IN COMPUTER-ASSISTED INSTRUCTION- A TWO INSTRUCTOR STUDY

Lawrence V. Fulton
Assistant Professor, QM, McCoy School of Business
Texas State University, 601 University Drive, San Marcos, TX 78666
210-837-9977 / 512-245-3237
Lf25@txstate.edu

Lana V. Ivanitskaya
Doctor of Health Administration Program, School of Health Sciences
Central Michigan University, HPB 1205, Mount Pleasant, MI 48859
Tel. 989.774.1639; FAX 989.774.2888
ivani1sv@cmich.edu

Rasim Musal
Assistant Professor, QM, McCoy School of Business
Texas State University, 601 University Drive, San Marcos, TX 78666
512-245-2295
Rm84@txstate.edu

Dmitry Erofeev
Adjunct Professor
Central Michigan University, Mount Pleasant, MI 48859
Tel. 989-832-6994; FAX 989.774.2888
erofe1da@cmich.edu

ABSTRACT

The purpose of this research was to evaluate variable computer-assisted (CA) certification standards on student performance in a required undergraduate business statistics course. Two instructors from a single university provided the data for this study. Based on theory, we posited that CA-learners exposed to higher homework certification standards (e.g., 80% correct versus 70% correct as a precursor for receiving full assignment credit) would score higher on a standardized final examination. To conduct this study, we established four different groups: a control group that did not use CA technology, a CA group requiring 70% correct for homework certification, a CA group requiring 80% correct, and a CA group requiring 90% correct. We discovered that two variables accounted for 20% of the adjusted variance: grade point average and time spent in certification. The effect of group membership and certification standard disappeared when including these controls. The study is expanding to include results from another university.

Key words: computer assisted instruction, CA learning, Hawkes Learning System, certification, statistics courses, learning goals

INTRODUCTION

The purpose of this research was to evaluate the effect of computer-assisted (CA) learning standards on student performance in undergraduate business statistics courses. In previous CA instruction studies, authors demonstrated tangible benefits from the use of CA instruction in classrooms [1] [2], investigated rigid versus flexible deadlines along with time management variables as predictors of performance for students [3], and identified that grade point average and time spent practicing in the CA environment affected performance on final examinations [4]. Given the demonstrated value of CA instruction, we were particularly interested in the optimal settings for one particular software: Hawkes Learning System (HLS) [5].

HLS provides interactive education using audio and video instruction, practice problems, and certification. The certification process requires students to answer a certain percentage of questions correctly in a particular topic area and therefore is likely to facilitate goal setting. The certification questions themselves contain randomly generated values. For example, a binomial question would have randomly generated values (within a range) for the total number of trials, for the total number of successes, and for the range to investigate (e.g., <, >, ..). While the framework (but not order) of a particular question is identical for everyone, the answer may be largely different. Once a student has answered a certain percentage of the assigned problems correctly, he or she receives full credit for that assignment, which serves as evidence that the student has the baseline knowledge or mastery of the content. The HLS also provides testing capability with this same randomization function.

From a practical standpoint, we sought to understand which level of certification would result in performance as measured by a standardized final examination. From a theoretical standpoint, we anticipated that students with higher certification requirements would perform better than those with lower certification standards. This proposition derived from Brehm's motivational theory [6] [7] which suggests that higher certification standards should equate to improved performance. This is likely to be the case if certification standards (which are essentially externally assigned learning goals) are understood, accepted, and adopted by the learners. According to the goal setting theory, challenging, specific and attainable goals are more likely to lead to improved performance than no goals or goals that are easy, general or unattainable [8]. Locke and Latham proposed that learners may develop stronger commitment to challenging goals and act purposefully, which will lead to improved performance. Goal setting can improve performance if several conditions are met. First, it is important that learners are aware of the actions that need to be taken to attain the goal and have sufficient ability to meet it. Second, feedback is valuable because it offers cues about how hard one should work to attain the goal.

The HLS design creates conditions for internalization (adoption) of goals set externally via certification standards. The system brings these goals to learner's awareness by displaying the maximum number of strikes allowed for earning a certificate. It also gives feedback on the number of strikes or incorrect answers. The goal of error-free performance is challenging, somewhat specific, and attainable by most learners. However, it is possible that some learners

will fail to accept externally imposed goals or view them as too easy or too hard. Goal specificity and knowledge of what to do to achieve the goal may also be insufficient, especially for low-ability learners, although HLS encourages practice as a path to successful certification. It re-directs everyone who attempts but does not achieve a certification back into problem practice. The computerized practice function incorporates detailed feedback and an interactive tutor that explains solutions step by step. Given these possibilities, research on certification levels is warranted because it can inform the practice of the CA instructional design.

Literature

Recent literature confirms that both performance outcomes and student satisfaction benefit from the inclusion of CA methods in traditional classroom settings. Fulton, Mangelsdorff, and Bewley (2009) demonstrated that in-class use of technology normally associated with distance learning improved student outcomes and satisfaction in statistics courses [1]. Sosa, Berger, Saw and Mary (2010) also identified a performance advantage in using computer-assisted (CA) instruction in statistics classes [2]. These two studies represent recent examples of how CA has benefit traditional classroom. Others studies have demonstrated the same type of CA benefit specifically for HLS.

In a previous HLS study, Fulton, Ivanitskaya, and Erofeev (2010) confirmed that implementation of more frequent deadlines for homework content certification correlated with improved performance of graduate students in statistics classes after isolating students' locus of control [3]. This single-university study did not include a performance comparison of traditional written homework versus interactive certifications, nor did it evaluate the efficacy of different certification standards (e.g., 70% versus 80% certification standards). In an undergraduate setting that did include certification standards, Fulton, Musal, Ivanitskaya, and Haidar (2011) identified that grade point average and time spent practicing homework in the HLS CA system were the most important predictors of performance in undergraduate statistics learning. While homework certification standards were not found to be statistically significant, the measurements for this study were based solely on a single instructor's classrooms and omitted an evaluation of traditional classroom instruction sans CA intervention [4]. This study resolves these issues.

The research question for this study follows. What are the effects of varying homework certification standards in HLS on student performance in business statistics undergraduate programs in a large, public university located in the south-central region of the United States? Answering this question should help faculty resolve the appropriate certification levels to optimize performance. Given existing theory [6] [7], we expected that increased certification requirements would increase performance.

METHOD

The setting for this study was a single university in the south-central region of the United States. Two instructors teaching the same undergraduate business statistics course provided data for the study. One instructor used CA technology (HLS) and another used paper and pencil testing. The study was designated exempt by the host university Institutional Review Board.

The design for this research was a multi-group pre-post with a "control" group but no randomization. Students were able to self-select into different classes, and for internal consistency, random assignment of a specific individual to a specific group was not feasible. The "control" group reflected a traditional statistics class with textbook coupled with lecture as the primary means of education and traditional paper examinations. For CA learning groups, HLS served as the primary source of homework, and certifications for each homework area were set at 70%, 80%, and 90%. The content areas for HLS included those generally found in basic business statistics courses. Such content included levels of measurement, descriptive statistics, probability, discrete random variables (including binomial, Poisson, and hypergeometric distributions), continuous random variables (including the uniform, normal, and t-distribution), sampling distributions, estimation and confidence intervals, hypothesis testing (one and two-sample), regression analysis (simple and multiple), and Chi-square (association and goodness of fit). The groups in the study follow and sample sizes follow:

1. Standard pencil and paper homework and in-class instruction, $n_1=104$
2. HLS with certification set at 70%, $n_2=35$
3. HLS with certification set at 80%, $n_3=35$
4. HLS with certification set at 90%, $n_4=43$

Variables

Dependent variable. For this study, the measure of performance derives from students' scores on a standardized final examination (ratio-level data). Due to instructor differences, a common set of final questions were selected, and student performance on these questions was measured. The "control" group questions derived from HLS but were provided on paper. The treatment groups took the examination using computers.

Manipulated variable. The manipulated variable was group membership. Students were assigned to one of the nominal groups, {1, 2, 3, 4}, described previously. The baseline group for this study was the control group. All other groups were recoded into dichotomous variables with 0 indicating non-membership and 1 indicating membership.

Categorical independent variables. Gender and Hispanic status were investigated as possible demographic controls. Hispanic status was important as the university itself serves a large Hispanic population. Other than Hispanic categorization, the students themselves represented a fairly homogenous group. The base category for gender was "female" coded 0 with male coded as 1, while the base category for "Hispanic" was "non-Hispanic" coded 0. No other demographic information was available for the study.

Quantitative independent variables (covariates). Age as of 1 January 2011, grade point average (GPA) on a four-point scale, and undergraduate hours completed were investigated as reasonable covariates. Additionally, we evaluated time spent in the three phases of HLS (instruct, practice, and certify) along with the number of attempts an individual took to certify.

We investigated a multiple regression model to forecast performance on the final examination. A priori, we estimated greater than 95% power with $\alpha=.05$ using G*Power 3.1[8]. We used a

stepwise rather than enter method in order to determine those variables most highly associated with performance on the final examination.

RESULTS

Descriptive statistics

Table 1 provides the descriptive statistics for the study variables applicable to all groups. The average final grade was .644, while the median was .682. The percentage of males was 64.1%, and 20.7% of the sample was Hispanic. The average age was 22.139 years. The average and median GPA of students was comparable (2.999 and 2.980). Most students were classified as juniors with 68.793 mean hours. Omitted from Table 1 are the instruct, practice, and certify times as well as the number of attempts to certify. These data apply only to the HLS groups. Table 2 breaks out the mean times for the time variables. The average HLS student spent 9.330 hours per week in instruct, practice, and certify.

Table 1. The descriptive statistics for the variables in the study follow.

Variable	N	Mean	Median	Mode	s	Min	Max
Grade on Final	217	.644	.682	.682	.208	.091	1.000
Control? 1= True	217	.479	.000	.000	.501	.000	1.000
70% Group? 1=True	217	.161	.000	.000	.369	.000	1.000
80% Group? 1= True	217	.161	.000	.000	.369	.000	1.000
90% Group? 1= True	217	.198	.000	.000	.400	.000	1.000
Male? 1= True	217	.641	1.000	1.000	.481	.000	1.000
Hispanic? 1= True	217	.207	.000	.000	.406	.000	1.000
Age as of 1/1/11	217	22.139	21.233	21.180	3.317	18.437	53.213
GPA	217	2.999	2.980	2.750	.397	2.060	3.880
Credit Hours Earned	217	68.793	67.000	60.000	16.831	26.000	137.000

Table 2. The mean time spent in all phases is depicted here.

	Time in Instruct	Time in Practice	Time in Certify	Time in all Phases
70% HLS Group	320.743	678.514	1398.771	2398.029
80% HLS Group	322.200	719.143	1356.857	2398.200
90% HLS Group	411.907	910.116	1718.721	3040.744
Total	185.323	405.774	785.032	7836.973
Total Hours	3.089	6.763	13.084	130.616
Total Hours / Week	0.221	0.483	0.935	9.330

Inferential statistics

Because we were largely interested in test scores for different groups, we ran a simple comparison of means using Analysis of Variance as a starting point. Figure 1 shows the comparison of means plot. The results were statistically significant (p<.001), and post-hoc analysis (Dunnet's C) revealed differences between the non-HLS and 70% HLS group as well as the non-HLS and the 80% HLS group. The mean grades for the non-HLS group, the 70% HLS group, the 80% HLS group, and the 90% HLS group were .583, .721, .725, and .663 respectively.
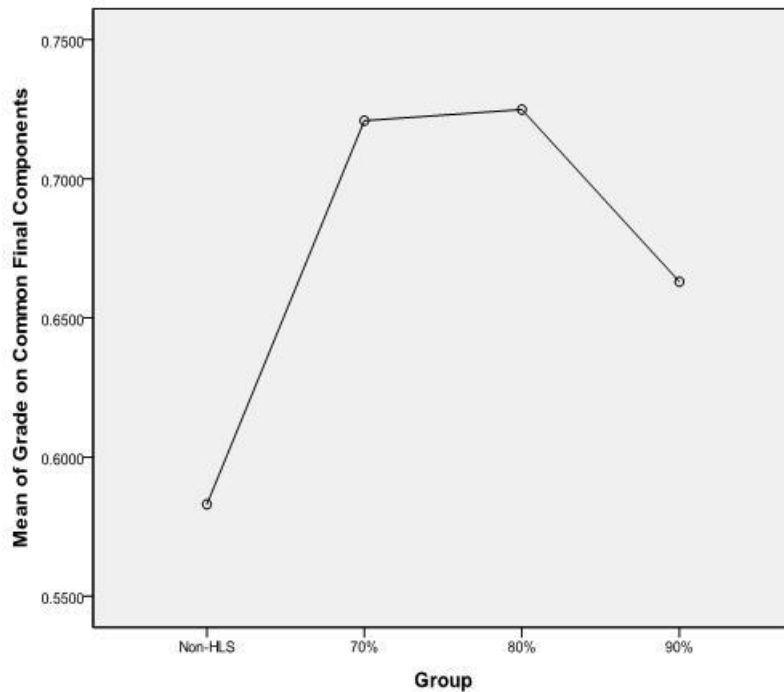
Figure 1. A comparison of means plot for the Analysis of Variance

Stepwise regression analysis of the independent variables on the comprehensive final revealed a statistically significant model, {F(2,214)=29.876, p<.001, adj. R^2=.211}, which accounted for 21.1% of the variance with only two variables, grade point average and time spent in certify. Collinearity diagnostics indicated that independent variables were not related, and the residuals (Figure 2) appeared normally distributed. Table 3 is the coefficient matrix. The standardized coefficients indicate that the two surviving variables, GPA and minutes spent in certification, are of approximately the same effect size.
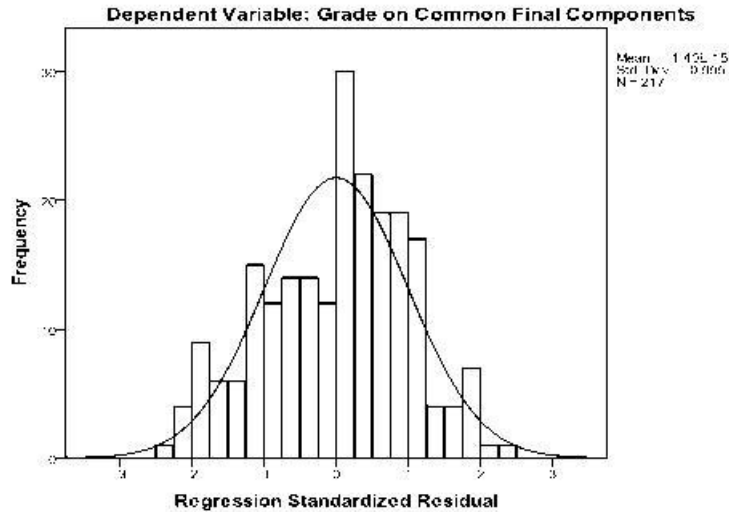
Figure 2. The histogram of the residuals suggests that the regression model was a reasonable choice.

Table 3. The coefficient matrix for the regression model follows.

Variable	B	SE	Beta	t	P	VIF
(Constant)	-.004737	.098626		-0.048	.961	
GPA	.196002	.031956	.373	6.133	<.001	1.015
Minutes in Certify	.000078	.000014	.329	5.405	<.001	1.015

DISCUSSION

When accounting for GPA and minutes to certify, we found no effect of group membership or any other control variable. This finding was not expected. Based on theory, we had assumed that increased performance standards (certification requirements that are similar to assigned goals) would have resulted in increased performance. What we do may be seeing in the original ANOVA is retreating marginal returns. Students may be discouraged with performance standards set exceedingly high and view these are unattainable. Given the amount of time the "average" HLS student spent in instruct, practice, and certify, those assigned to the more difficult groups might have perceived this level of effort to be unsustainable. Some students, such as those with low initial ability, may have rejected an externally assigned goal of 90% error-free performance in favor of their own goal ("I just want to pass this class") that was less than optimal. Consequently, they may have expected to do poorly on the examination and put insufficient effort into earning high exam grades. As compared to students from 70% and 80% HLS groups, students in the 90% HLS group spent more time getting CA instruction and practicing problems but this additional time on task did not translate into improved course performance. The pattern identified in the ANOVA becomes statistically irrelevant once ability (as proxied by GPA) and application of that ability (as proxied by minutes spent attempting to certify) are included in a more complete model.

The findings of this study are similar to those discovered in a single-instructor study without control group [4] with one major exception. This model indicates that for this population, certification time is the second-most important consideration (other than pre-existing ability / dedication as proxied by GPA).

Some of the limitations of this work are the inability to randomize and the isolation to a single university. Further, we assume that a comprehensive final examination is a reasonable proxy for performance. In fact, we will be including the results of this work in a two-university study once secondary IRB approval is obtained from our partner. We will also be including multiple measures of standardized performance in a multivariate analysis. Finally, we will be applying Bayesian methods to provide estimates of group membership probability.

REFERENCES

[1] Fulton, L., Mangelsdorff, D. & Bewley, L. *An Analysis of Statistics Education Technologies in Graduate Health and Business Administration*. International Conference of Education, Research and Innovation 2009, Barcelona, Spain. Abstract ISBN: 978-84-613-5536-5, Paper ISBN: 978-84-613-5538-9.

[2] Sosa, G., Berger, D., Saw, A., & Mary C. *Effectiveness of Computer-Assisted Instruction in Statistics: a Meta-Analysis*. Review of Educational Research, 2010, Online First. doi: 10.3102/0034654310378174.

[3] Fulton, L., Ivanitskaya, L., & Erofeev, D. *Frequent Deadlines in Statistics Courses: If Health Executives Learn Better When They are Enforced, Why Won't Undergraduate Students?* Proceedings of the 2010 Joint Statistics Meeting, Vancouver, BC, 2010.

[4] Fulton, L., Musal, R., Ivanitskaya, L., & Haidar, S. *The Effects of Computer-Assisted Instruction and Variable Certification Requirements on Student Performance in Undergraduate Statistics Classes*, To appear in Proceedings, International Technology, Education and Development Conference 2011.

[5] Hawkes Learning System. *Hawkes Learning System Statistics Software* [software]. Available from http://fileburst.hawkeslearning.com/Setups/STATDIS-Student-Setup-Full.exe, 2010.

[6] Brehm, J. W., & Self, E. *The Intensity of Motivation*. Annual Review of Psychology, 1989, 40, 109–131.

[7] Brehm, J. W., Wright, R. A., Solomon, S., Silka, L., & Greenberg, J. *Perceived Difficulty, Energization, and The Magnitude of Goal Valence*. Journal of Experimental Social Psychology, 1983, 19, 21–4

[8] Locke, E. A., & Latham, G. P. *A Theory of Goal Setting and Task Performance*. Englewood Cliffs, NJ: Prentice Hall, 2009.

[8] Faul, F., Erdfelder, E., Buchner, A., & Lang, A. *Statistical Power Analyses Using G*Power 3.1: Tests for Correlation and Regression Analyses*. Behavior Research Methods, 2009, 41, 1149-1160.

WEBEDQUAL: DEVELOPING A SCALE TO MEASURE THE QUALITY OF ONLINE MBA COURSES

Rose Sebastianelli, sebastianer1@scranton.edu, (570) 941-4287
Nabil Tamimi, tamimin1@scranton.edu, (570) 941-4288
Kingsley Gnanendran, skg355@scranton.edu, (570) 941-4190
Kania School of Management, University of Scranton, Scranton, PA 18510

ABSTRACT

This paper presents preliminary work toward the development of a valid and reliable scale (WebEdQual) to measure and ultimately improve the quality of online MBA courses. Scale items reflect the various aspects of online education presumed to affect the quality of e-learning experiences. Based on data collected from MBA students who have completed online courses, we use factor analysis and derive empirically the following nine underlying e-learning quality dimensions: *Professor-Student Interaction, Course Content-Structure, Content Rigor, Technology, Student-Student Interaction, Assessment, Flexibility-Convenience, Team-Based Learning,* and *Delivery Method.* Our results indicate that several of these dimensions are correlated significantly with student outcome measures.

Keywords: Online MBA, Quality, Satisfaction, Perceived Learning.

INTRODUCTION

The shift toward online business education is well underway as evidenced by the increasing number of business courses and business degree programs offered entirely online. In the first of a series of three papers exploring online education, Dykman and Davis [2] provide an excellent discussion of what is fueling this shift toward Web-based instruction addressing issues such as educational access, changing paradigms for teaching and learning, competition among universities, a more global perspective, improved technologies, and economic considerations. Moreover, Web-based education is rapidly gaining acceptance from both the student perspective (e.g., convenience, flexibility) as well as that of the educational institution (e.g., cost effectiveness, new markets). With some predicting that the Internet will be the primary channel for delivering MBA programs in the future, it is hardly surprising that business educators have directed attention toward improving the quality of e-learning experiences.

Our study continues in this line of research as we work toward the development of a scale to measure the quality of online MBA courses. We include items that relate to various aspects of online instruction presumed to affect the quality of e-learning experiences, and based on data collected from online MBA students, we use factor analysis to derive a set of underlying e-learning quality dimensions. While the scale needs to be further refined, preliminary results do indicate that many of the extracted factors correlate significantly with outcome measures, such as perceived quality of, and student satisfaction with, online MBA courses. An empirically validated scale to measure the quality of online MBA courses (WebEdQual) would be a useful tool for both monitoring and improving the online delivery of MBA courses and programs.

LITERATURE REVIEW

Best Practices

A number of articles suggest approaches for assuring a quality online educational experience. Grandzol and Grandzol [5], in a review of the literature on best practices for online business education, provide guidance on course design and delivery, student services, and administration. Consistent themes include standardizing course structure, modularizing course content, giving prompt and constructive feedback, providing technical support for students, creating a learner-centered environment, providing training for faculty, and limiting class sizes. Some of these components have been shown to increase student satisfaction. For instance, students report higher levels of satisfaction and increased learning in courses with greater consistency in structure and format compared to those with less consistency [10]. Similarly, Dykman and Davis [3] reiterate many of these same best practices (e.g., a standardized approach to course design), but also stress the importance of clearly defined learning objectives and planning in an online environment where courses must be prepared almost entirely in advance. Unlike in a traditional classroom, learning objectives cannot evolve as an online course progresses. Consequently, they advocate articulating overall learning objectives for the course as well as specific learning objectives for each unit to drive course design and better define student expectations.

In a study that included not only business but several other disciplines (education, liberal arts, etc.), Gaytan and McEwen [4] surveyed both faculty and students about their perceptions regarding online education. Specifically, they were interested in the types of instructional strategies and techniques for assessing student performance that are perceived to enhance the quality of online learning. The major recommendations from their findings are that (1) faculty should use a variety of techniques for evaluating student work such as exams, quizzes, projects, portfolios, discussion boards, and (2) students need meaningful and timely feedback to facilitate online learning. Given the obvious concerns faculty have about using online exams to assess student performance (e.g., the inability to verify the identity of the test taker in an unproctored online environment), Khare and Lam [6] provide a critical evaluation of online exams both in terms of pedagogical and technical issues. They suggest that online exams should focus on mastery of knowledge and competence of skills rather than on factual recall. Furthermore, based on data from an online MBA program, they conclude that a comprehensive online exam is one of several important assessment tools for measuring effectiveness and providing feedback.

Online MBA Outcomes

Several empirical studies have focused specifically on factors that affect outcome measures in online MBA courses and programs. In an early study, Bocchi, Eastman and Swift [1] profiled students in an online MBA program and identified strategies that were successful in addressing potential attrition issues. These students, who tended to be older with significant professional business experience, cited accessibility, convenience, fit with career and personal growth plans as their primary reasons for joining the online MBA program. They found the following retention strategies to be successful: a cohort and team-based approach to learning, extensive faculty feedback and interaction, relevant content and activities, and faculty who are both interested and competent in teaching online.

Exploring the possibility that more complex relationships exist among the variables affecting perceived quality, perceived learning and satisfaction in online courses, several studies employed structural equation modeling (SEM) to examine proposed interdependencies. Marks, Sibley and Arbaugh [7] used SEM to test their theoretical model that shows online instructional factors (instructor-student interaction, student-student interaction and student-content interaction), online educational advantages (e.g., convenience) and student characteristics (e.g., gender) affecting satisfaction indirectly through their impact on perceived learning. Using data collected from students taking online courses within a traditional MBA program, they found that instructor-student interaction is the most important factor affecting perceived learning/satisfaction in an online environment. Student-student interaction, few (but not all) student-content interaction variables, and online educational advantages were also found to be significant. Online educational advantages, although significant, were found to be the least important. Finally, student characteristics were not found to significantly impact perceived learning/satisfaction in online courses.

Another study, carried out by Peltier, Schibrowsky and Drago [9], also used SEM but to test a much more complex model of interrelationships among various dimensions and the perceived quality of online learning. Their study included the following six dimensions of online teaching quality: (1) student-to-student interactions, (2) student-to-instructor interactions, (3) instructor support / mentoring, (4) lecture delivery quality, (5) course content and (6) course structure. Not only were interrelationships among these dimensions proposed in the model, but each dimension was hypothesized to have a direct positive relationship on the perceived quality of an online learning experience. Based on data collected from students enrolled in an online MBA program at a Midwestern university, they found, among several significant interrelationships, that three dimensions (instructor support / mentoring, course content, and course structure) have a significant positive relationship with the perceived quality of online learning. No student characteristics were considered in this study.

METHOD

Research Setting

This study was carried out in conjunction with the MBA program at a private comprehensive university. Both traditional and online MBA programs are offered, with the online program having been initiated in the spring of 2008. Traditional MBA students may opt to take up to six credits online; however, students entering either program may be required to take up to twelve one-credit foundation modules which are offered solely online. Consequently, most traditional MBA students take at least one course online.

The online MBA program is offered via partnership with a provider of online education that helps academic institutions with program and course development, marketing and publicity, student recruitment and retention, and technical support services. As such, all online courses in the program follow many best practices cited in the literature such as consistent course structure, modularized content, 24/7 help desk for students, explicit student learning objectives, extensive training for faculty on ANGEL (the Web platform used for course delivery), and limited class sizes. Multiple options are used to encourage a high level of professor-student and student-

student interaction. These include a discussion forum in which students are required to make a minimum of two posts per week in response to questions posed by the instructor, an "ask the instructor" forum for students to raise questions about problems they encounter with course material, a student lounge designed for more social, non course-related discourse, and e-mail.

Questionnaire

The questionnaire consists of two sections. The first section includes a few questions that gather identifying information (i.e., student ID number) and some background data (e.g., employment status). The second section consists of a series of statements for which students are asked to indicate their level of agreement on a seven point Likert-type scale (1 = strongly disagree, 2 = disagree, 3 = somewhat disagree, 4 = neither agree nor disagree, 5 = somewhat agree, 6 = agree, 7 = strongly agree). These scale items were developed to capture the major factors thought to impact students' perceptions of, and satisfaction with, online learning that have been cited in the literature. We include items that represent all six dimensions (course content, course structure, lecture delivery quality, professor-student interaction, student-student interaction, and instructor support/mentoring) used in previous research [9]. We also include items to measure online educational advantages, such as convenience, that have also been cited [7]. Finally, we develop items to capture two new additional factors, assessment of student performance and technology issues. The number of items totaled 59. We also include three statements to capture student outcome measures, namely students' perceptions of the quality of their online MBA courses, student satisfaction with their online MBA courses, and the extent to which students believe that they have met the learning objectives in their online MBA courses.

We pre-tested the questionnaire using a small sample of traditional MBA students who had taken at least one online course. Based on the pre-test results, some items were eliminated or reworded to improve clarity. In accordance with standard practice, some questions were "negatively" worded and the order of the items was randomized (similar questions relating to the same aspect of online instruction were not presented together). This was done to reduce the potential of any "halo effects." The complete list of items is shown in the Appendix.

Data Collection

Students in the online MBA program and those in the traditional MBA program who had taken at least one course online were contacted via e-mail to participate in the study. As this is part of an ongoing longitudinal study, the data for this paper were collected at two different points in time, in mid-August 2009 and then again in mid-May 2010. The e-mail providing a link to the online survey was sent to a total of 280 online MBA students and 100 traditional MBA students. An incentive lottery was used to increase participation.

RESULTS

Respondent Profile

A total of 110 students completed our online survey resulting in a 29% response rate. Of those responding, 75 % are enrolled in the online MBA program, 85% are employed full-time, and

55% are male. They report an average of 11 years professional experience, ranging from none to 35 years. It should be noted that students applying to the online MBA program must have a minimum of 3 years of professional (or supervisory) experience; however there is no such admission requirement for students in the traditional on-campus MBA program.

Factor Analysis

In order to exploit the intercorrelations among the 59 items representing various aspects of online instruction, factor analysis was performed to identify the underlying e-learning quality dimensions. Prior to analysis, the mean item rating was used to replace any missing values for individual items and the responses to "negatively" worded questions were reversed. Based on an examination of the scree plot, nine factors were extracted to account for about 54% of the total variation in the observed ratings. Figure 1 shows the items that loaded strongly on each of the nine factors. In developing this factor solution, items with loadings less than 0.35 (after varimax rotation) were dropped. Under each factor the items are listed in descending order according to loading magnitude.

The nine dimensions extracted are labeled based on the specific items that load most heavily on each factor. Therefore, the nine underlying dimensions are *Professor-Student Interaction, Course Content-Structure, Content Rigor, Technology, Student-Student Interaction, Assessment, Flexibility-Convenience, Team-Based Learning,* and *Delivery Method.* In order to assess the internal consistency of these derived factors, *Cronbach's alpha* was computed as a measure of reliability. These are also reported in Figure 1.

FIGURE 1: Factor Solution: Items Loading on Each Factor
_____

Factor I – "Professor-Student Interaction" *(Cronbach's alpha = .891)*
- My online professors are very responsive to students' concerns.
- Most of my online professors respond to questions in a timely manner.
- Most of my online professors actively facilitate discussion in forums.
- Most of my online professors participate in the Student Lounge on a regular basis.
- My online professors often identify key points to facilitate learning.
- My online professors show interest in students' progress.
- I find the professor's responses to questions posted to the "Ask the Instructor Forum" helpful.
- In general, there is little interaction between the professors and students in my online courses.
- My online professors adjust their methods of instruction based on student feedback.
- Course content is communicated effectively in my online courses.
- My online professors encourage the expression of different viewpoints.
- My online professors design course content to stress important concepts.
- Feedback from professors is meaningful in my online courses.

Factor II – "Course Content - Structure" *(Cronbach's alpha = .801)*
- My online MBA courses are current and up to date.
- "Tasks for the week" helps me to meet course requirement deadlines.
- The content in my online courses is applicable and useful to professional work situations.
- My online courses are organized in a way that is easy to navigate.
- The consistent format for each course makes it easy for me to access materials I need for the week.
- The content in my online courses add value to my MBA educational experience
- Individual assignments helped me to understand content through application of concepts and skills covered in the course.
- In general, the exams in my online courses are fair.
- The weekly overview and objectives clearly identify learning goals to be achieved in the upcoming week.

Factor III – "Content Rigor" *(Cronbach's alpha = .495)*
- The content in my online courses is less rigorous than I expected.
- The content in my online courses is challenging.
- In general, my online courses cover fewer topics than I expected.
- I don't spend much time studying for online exams.
- In general, my online courses cover fewer topics than I expected.

Factor IV – "Technology" *(Cronbach's alpha = .692)*
- Technology problems interfere with my online learning.
- I am often frustrated with technology in my online courses.
- I sometimes encounter problems accessing materials in my online courses.
- I am sometimes overwhelmed by the volume of posts to the discussion forum.
- Technical support is available when I need it.
- More often than not I felt intimidated asking my online professor questions.
- I would generally classify myself as being tech-savvy.

Factor V – "Student-Student Interaction" *(Cronbach's alpha = .671)*
- Most students participate more than required in Discussion Forums.
- Other students' posts to the Discussion Forum are helpful in understanding different viewpoints.
- I learn more from my fellow students in online courses than I have in traditional classroom settings.
- I don't feel comfortable asking my online professors for advice.
- My online courses offer me ample opportunities for problem solving experiences.
- Other students' posts to the Discussion Forum are not useful in learning course content.

Factor VI – "Assessment" *(Cronbach's alpha = .999)*
- The grades I receive in my online courses accurately reflect my performance.
- I like having access to course materials during online exams.

Factor VII – "Flexibility-Convenience" *(Cronbach's alpha = .574)*
- Taking courses online makes it easy for me to balance my education with other responsibilities.
- I consider not having to commute to class a major advantage of online courses.
- Enrolling in online courses allows me to complete my MBA in less time than expected.
- PowerPoint slides are an effective way to deliver course content online.
- I like that course content becomes available weekly rather than all at once at the beginning of my online courses.

Factor VIII -"Team-Based Learning" *(Cronbach's alpha = .146)*
- Student evaluation is more likely based on individual rather than group work in my online courses.
- Group work is encouraged by professors in my online courses.
- In general, my online professors are supportive of my personal circumstances.
- I have enough opportunities for collaborative team work in my online courses.

Factor IX – "Delivery Method" *(Cronbach's alpha = .280)*

- I would like to see more video—audio content available in my online courses.
- Video—audio presentations facilitate learning the key topics in my online courses.
- I prefer online courses that use a variety of methods (e.g., assignments, discussion posts, exams) to evaluate student performance.
- The online environment is not favorable for learning certain topics typically taught in MBA programs.

_____

Correlation Analysis

In order to determine how these dimensions relate to the outcome measures of perceived quality, satisfaction, and student achievement of learning objectives, correlations were computed between the average ratings of the items comprising each factor and responses to the three outcome items. These correlations appear in Table 1.

TABLE 1: Correlations of Factors with Outcomes

	Quality	*Satisfaction*	*Learning Objectives*
I. Professor-Student	0.449	0.678	0.168
II. Content-Structure	0.629	0.551	0.407
III. Content Rigor	0.322	0.249	-0.007
IV. Technology	0.222	0.039	0.344
V. Student-Student	0.631	0.455	0.203
VI. Assessment	0.257	0.108	0.196
VII. Convenience	0.202	0.463	0.151
VIII. Team-Based	0.297	0.241	0.344
IX. Delivery Method	0.326	0.201	0.208

DISCUSSION AND IMPLICATIONS

While a number of studies have considered how dimensions of online instruction relate to perceived quality, perceived learning and satisfaction in online courses, to the best of our knowledge these dimensions were determined *a priori*. In other words, researchers measured these dimensions using items that they grouped together prior to gathering student responses. Consequently, our study adds to the literature by attempting to derive these underlying dimensions empirically using factor analysis on student responses.

That said, our study is exploratory in nature and more work is needed to refine our scale. *Cronbach's alpha* values for a few of our derived factors are too low. Although the generally acceptable minimum alpha is usually 0.70, Nunnally [8] suggests a somewhat lower threshold, such 0.50, for exploratory work involving the use of newly developed scales. The final two factors extracted in our analysis, *Team-based Learning* and *Delivery Methods*, fail to meet even this lower threshold.

Nonetheless, we are encouraged that a number of our better defined and more reliable factors are congruent with those cited in the literature and used in previous research studies. The first two factors extracted in our analysis, *Professor-Student Interaction* and *Course Content-Structure*, have a sufficient number of strongly loading items as well as suitable values for *Cronbach's alpha* indicating a high level of internal consistency. Moreover, these two dimensions have been found to be significant in predicting student outcome measures [7] [9]. Our correlations are consistent with these previous findings (see Table 1). Student responses to the outcome item "My online courses are of high quality" are correlated fairly strongly (0.629) with the average ratings across the items comprising Factor II *(Course Content-Structure)*. In addition, student responses to the item "I am very satisfied with the online courses in my MBA program" are correlated fairly strongly with average ratings across the items comprising Factor I *(Professor-Student Interaction)*.

Our preliminary results also suggest that perceived quality is correlated fairly strongly with *Student-Student Interaction* and that satisfaction is correlated moderately with *Course Content-Structure* and *Flexibility-Convenience*. With respect to students' perceptions about meeting learning objectives in their online MBA courses, we find this to be the most correlated with *Course Content-Structure, Technology,* and *Team-Based Learning*; while significant at $\alpha = 0.05$, these correlations are not very strong.

An obvious limitation of our study is the sample size. In addition to enabling more definitive results regarding the extracted dimensions, more observations would have allowed us to test a complex, and perhaps more realistic, model of interrelationships among the factors and outcome measures using SEM. The online MBA program being studied is growing rapidly, and as noted above, we plan to continue surveying students as part of an ongoing longitudinal approach.

Further work is needed to refine our scale items. Some items may have to be discarded (those that did not load on any factors), some reworded, and some new items added. The goal is to develop a reliable and valid scale that can be used to measure and ultimately improve the quality of online MBA education.

APPENDIX: Items on Questionnaire

1. The online environment is not favorable for learning certain topics typically taught in MBA programs.
2. The content in my online courses add value to my MBA educational experience
3. The content in my online courses is applicable and useful to professional work situations.
4. My online courses offer me ample opportunities for problem solving experiences.
5. The content in my online courses is challenging.
6. The content in my online courses is less rigorous than I expected.
7. In general, my online courses cover fewer topics than I expected.
8. Topics covered in my online courses are easy to learn.
9. My online MBA courses are current and up to date.
10. Individual assignments helped me to understand content through application of concepts and skills covered in the course.
11. The weekly overview and objectives clearly identify learning goals to be achieved in the upcoming week.
12. The consistent format for each course makes it easy for me to access materials I need for the week.
13. "Tasks for the week" helps me to meet course requirement deadlines.
14. My online courses are organized in a way that is easy to navigate.
15. I like that course content becomes available weekly rather than all at once at the beginning of my online courses.
16. PowerPoint slides are an effective way to deliver course content online.
17. I would like to see more video—audio content available in my online courses.
18. Online materials contain information not covered in the textbook.
19. Course content is communicated effectively in my online courses.
20. Video—audio presentations facilitate learning the key topics in my online courses.
21. My online professors design course content to stress important concepts.
22. I regularly post questions to the "Ask the Instructor Forum."
23. I find the professor's responses to questions posted to the "Ask the Instructor Forum" helpful.
24. Most of my online professors actively facilitate discussion in forums.
25. In general, there is little interaction between the professors and students in my online courses.
26. Most of my online professors respond to questions in a timely manner.
27. Most of my online professors participate in the Student Lounge on a regular basis.
28. Feedback from professors is meaningful in my online courses.
29. More often than not I felt intimidated asking my online professor questions.
30. Most students participate more than required in Discussion Forums.
31. Other students' posts to the Discussion Forum are helpful in understanding different viewpoints.
32. Other students' posts to the Discussion Forum are not useful in learning course content.
33. I have enough opportunities for collaborative team work in my online courses.
34. I learn more from my fellow students in online courses than I have in traditional classroom settings.
35. Group work is encouraged by professors in my online courses.

36. Student evaluation is more likely based on individual rather than group work in my online courses.
37. I am sometimes overwhelmed by the volume of posts to the discussion forum.
38. In general, my online professors are supportive of my personal circumstances.
39. I don't feel comfortable asking my online professors for advice.
40. My online professors are very responsive to students' concerns.
41. My online professors encourage the expression of different viewpoints.
42. My online professors show interest in students' progress.
43. My online professors adjust their methods of instruction based on student feedback.
44. My online professors often identify key points to facilitate learning.
45. I consider not having to commute to class a major advantage of online courses.
46. Taking courses online makes it easy for me to balance my education with other responsibilities.
47. Enrolling in online courses allows me to complete my MBA in less time than expected.
48. I like the ability to access my online courses at any time.
49. Online exams should be included as one way of assessing student performance in every course.
50. In general, the exams in my online courses are fair.
51. The grades I receive in my online courses accurately reflect my performance.
52. I like having access to course materials during online exams.
53. I don't spend much time studying for online exams.
54. I prefer online courses that use a variety of methods (e.g., assignments, discussion posts, exams) to evaluate student performance.
55. I would generally classify myself as being tech-savvy.
56. Technical support is available when I need it.
57. Technology problems interfere with my online learning.
58. I am often frustrated with technology in my online courses.
59. I sometimes encounter problems accessing materials in my online courses.

Outcome Measures

- My online courses are of high quality.
- I am very satisfied with the online courses in my MBA program.
- I have met the learning objectives in most of my online MBA courses.

REFERENCES

[1] Bocchi, J., Eastman, J.K., & Swift, C.O. *Retaining The Online Learner: Profile of Students In An Online MBA Program and Implications for Teaching Them.* Journal of Education for Business, 2004, 79 (4), 245-253.

[2] Dykman, C.A. & Davis, C.K. *Online Education Forum Part One – The Shift Toward Online Education.* Journal of Information Systems Education, 2008, 19 (1), 11-16.

[3] Dykman, C.A. & Davis, C.K. *Online Education Forum Part Three – A Quality Online Educational Experience*. Journal of Information Systems Education, 2008, 19 (3), 281-289.

[4] Gayton, J., & McEwen, B. C. *Effective Online Instructional and Assessment Strategies*. The American Journal of Distance Education, 2007, 2(3), 117-132.

[5] Grandzol, J. R., & Grandzol, C. J. *Best Practices for Online Business Education*. International Review of Research in Open and Distance Learning, 2006, 7(1), 1-18.

[6] Khare, A., & Lam, H. *Assessing Student Achievement and Progress with Online Examinations: Some Pedagogical and Technical Issues.* International Journal on E-Learning, 2008, 7(3), 383-402.

[7] Marks, R.B., Sibley, S.D., & Arbaugh, J.B. *A Structural Equation Model of Predictors for Effective Online Learning*. Journal of Management Education, 2005, 29 (4), 531-563.

[8] Nunnally, J.C. Psychometric Theory, 2nd ed., New York: McGraw-Hill, 1978.

[9] Peltier, J.W., Schibrowsky, J.A., & Drago, W. *The Interdependence of the Factors Influencing the Perceived Quality of the Online Learning Experience: A Causal Model*. Journal of Marketing Education, 2007, 29 (2), 140-153.

[10] Shea, P.J., Fredericksen, E.E., Pickett, A.M., Pelz, W. & Swan, K. *Measures of Learning Effectiveness in The SUNY Learning Network.* In J. Bourne and J.C. Moore (Eds.) *Online Education: Learning Effectiveness, Faculty Satisfaction and Cost Effectiveness,* 2001, 31-54.Needham, MA: Sloan – C.

Use of Mastery Modules in a Blended Learning Environment

Bhupinder S. Sran, Ph.D., John Weber, D.B.A., Devinder Sud, M.Sc., *DeVry University, North Brunswick, New Jersey*

Before the emergence of the World-Wide Web, most college courses were offered in a traditional format where the faculty delivered all content in a classroom setting. Reinforcement activities, such as homework, were done outside of the classroom. With the widespread use of the World Wide Web, non-traditional delivery modes, such as online and blended, have emerged. Blended courses, also known as hybrid courses, require the integration of online and onsite activities. While faculty are well-versed in onsite delivery, developing meaningful and effective online activities can be a challenge. This is especially true for courses that are quantitative in nature.

The Keller Graduate School of Management at DeVry University has adopted the online and delivery formats for courses in its academic programs. To help the students become proficient in quantitative topics, online mastery modules have been incorporated into several courses. The mastery modules engage the students by requiring them to solve several exercises each week in an interactive learning environment. When students are not able to derive the correct solution for an exercise, online guidance is provided through step-by-step tutorials. Once the student has completed all exercises, he/she is given up to two opportunities to pass an online mastery test, which contains problems similar to the exercises.

This study explores how mastery modules were used in three graduate level courses in the areas of accounting and statistics during the Summer and Fall semesters in 2010 at the North Brunswick, NJ campus of DeVry University. The courses were taught by the researchers.

A NEW APPROACH TO AN OLD PROBLEM:
USING COGNITIVE MAPPING TO FOSTER INTER-FUNCTIONAL INTEGRATION IN CORE BUSINESS COURSES

Katia Passerini
New Jersey Institute of Technology
University Heights, Newark, NJ 07102
Email: pkatia@njit.edu Phone: 973-642 7328

Jose C. Casal
New Jersey Institute of Technology
University Heights, Newark, NJ 07102
Email: jose.c.casal@njit.edu Phone: 973-596 3254

Annaleena Parhankangas
New Jersey Institute of Technology
University Heights, Newark, NJ 07102
Email: annaleena.parhankangas@njit.edu Phone: 973-596 4281

Mark J. Somers
New Jersey Institute of Technology
University Heights, Newark, NJ 07102
Email: mark.somers@njit.edu Phone: 973-596 3279

ABSTRACT

Criticism of management education spans an array of issues from the relevance of content to methodology and pedagogy. One of the vexing problems associated with business school pedagogy, for example, is an ongoing struggle with inter-functional integration; that is, the tendency for both business school faculty and students to think of management problems in terms of discrete functional areas without seeing connections between them. This study uses cognitive mapping to assess the lack of inter-functional integration and to suggest remedial interventions. It proposes using mapping tools to guide and assess student learning at the course level, presenting examples from management information system courses. Finally, it suggests extending concept mapping tools to interventions at the broader program and curriculum-levels.

Keywords: mind maps, cognitive maps, assurance of learning, student learning outcomes

INTRODUCTION

An ongoing and spirited debate about the relevance, efficacy and future direction of management education is evident in the recent literature. The debate is multi-faceted and is focused on several critical issues including the purpose and relevance of research in business schools [cf., 2], the skill base of business school graduates [cf., 16] [20], the relevance of business school curricula to management practice [1] [16] and the intellectual underpinnings of management education [10] [24].

Although criticism of management education spans an array of issues, an important, underlying theme is evident. It is centered on Ackoff's [1] observation that we are teaching students the wrong things by using an educational model that emphasizes linear, deterministic thinking [10] [24] by building technical skills that have little relevance to management practice [cf., 2] [16] [20] [22].

One of the vexing problems associated with business school pedagogy is an ongoing struggle with inter-functional integration; that is, the tendency for both business school faculty and students to think of management problems in terms of discrete functional areas without seeing connections between them. For example, students tend to think of "loss of market share" in terms of marketing related issues such as advertising or competitive pressures without ever considering the influence of other areas such as poor customer service stemming from dissatisfied employees (which is a human resources management problem).

Indeed, Porter & McKibbion [21] found this to be a problem with management education over twenty years ago, and it remains one today as curricula and courses are sliced into small pieces that are focused on technical knowledge [10]. The problem tends to be magnified in business courses that have a large technical component such as finance, information systems and accounting because mastery of basic concepts in these areas is often seen as the primary objective of such courses [21].

Attempts to address the problem of inter-functional integration in business school curricula have taken the form curriculum changes (most notably addition of capstone courses or capstone sequences) or "infusion" of certain concepts across the curriculum (e.g., infusion of ethical reasoning). Although such efforts are admirable, they are limited in effectiveness in two ways. First, such efforts are typically not based on data capturing student learning so that it is not clear exactly where or how deficiencies in the desired level of integration across functions are present. Second, they are almost always large scale programs designed to address problems throughout the curriculum when a smaller, more focused approach based on a few courses - or one course - might be a useful prototype to drive subsequent larger scale efforts.

This paper addresses both of these potential problems. It begins with a broad assessment of the degree of inter-functional integration in an MBA program using mind maps. Based on those findings, changes were made to the core MBA course in information systems by using concept maps. The intention was to diagnose problems with inter-functional integration and then use those results to reinforce connections between the management of information systems and other management functions.

USING MIND MAPS TO ASSESS STUDENT LEARNING

Mind mapping is a technique in which the thinking process is visually represented by connecting concepts and ideas to a central issue or problem [5]. It provides insights into critical thinking through visual representation of the manner in which people deploy and organize concepts around a central issue [9]. Mind maps, thus, provide insights into the concepts that were deemed relevant to a particular problem or process by graphical representation of how knowledge is structured and integrated.

Mind maps have been used widely in studying student learning in academia. Specifically, they have been used to both assess and facilitate student learning in degree programs in several disciplines including the social sciences [4], nursing [9], engineering [26], and business [15]. Research in engineering education indicates that mind maps enhance student creativity [26] while the use of mind mapping in EMBA programs helps students to integrate diverse higher-order constructs and to develop metaphorical thinking [15]. Mind maps can also be used to assess changes in students' knowledge structures over time [8] [11] and to improve students' skills in reaching accurate diagnoses in nursing programs [17].

Mind Maps as Representations of Inter-Functional Integration

Mind mapping begins by placing a central thought or focus area in center of the maps. This represents the problem or issue to be addressed. Branching from the central focus are groups of related concepts represented by keywords associated with them. These concepts are then linked with arrows that demonstrate associations among them. Mind mapping allows the process of solving a problem to be viewed holistically and there is evidence that using mind maps as a learning tool encourages both left and right brained thinking [25]; that is, analytical and creative thinking.

By modeling the cognitive processes underlying critical reasoning and problem solving, mind maps capture the manner in which students frame and interpret complex problems by identifying the knowledge they deem relevant and by providing a visual representation of how that knowledge is organized. As such, they are ideally suited to capturing the degree to which students integrate business functions as they provide a visual representation of how students organize business concepts in relation to a specific

problem. Further, they capture the degree to which students make connections between concepts from the various business disciplines.

Therefore, mind maps were chosen as a suitable tool to investigate the question on whether business school students have a weak grasp of inter-functional integration among various business components. In order to identify which areas of the curriculum should be modified and how, without making assumptions of the underlying inter-functional knowledge gained by our students in the course of the MBA program, we started our study with a broad and open research question:

RQ1: *Do students in MBA programs show a strong understanding of how the key elements of a business are integrated in order to make such business successful?*

STUDY METHODS AND RESULTS

Study Participants

The sample was comprised of 24 advanced graduate students enrolled in the capstone course in strategic management at an AACSB accredited university located in the Northeastern region of the US. All of the students were in their last year of study and had completed most of the MBA curriculum. Participation was voluntary and students did not receive extra credit or any other incentives for their involvement in the study.

Procedures

Approximately midway through a fifteen-week semester, students were introduced to the concept of mind mapping and were trained on how to develop mind maps. A practice exercise in which students mapped the process of finding a job was administered to ensure that students were familiar with the technique and that they were able to apply it effectively. Results from this pilot exercise indicated that students understood the concept of mind mapping and were able to apply it in a meaningful fashion. Students were then asked to develop a mind map for the problem of defining a successful company. More specifically, before engaging in the activity, students looked at a sample map relevant to the process of preparing for an examination. They were then asked to develop (on paper) a mind map that answered the question: "*What do you need to make your business successful? What are the elements of a successful business?*" The concept "successful business" was placed in the center of the map and students were instructed to identify and link the concepts that they thought were associated with the operation of a well-managed, successful company.

Our intention was to choose a broad problem area to assess the depth and the breadth of students' knowledge base. Further, examining a firm in its totality, provides insights into

the degree to which students think in terms of processes, structures, or functional areas as well as the degree to which knowledge is isolated (e.g., atomized) or integrated.

Semantic Analysis

Semantic analysis of mind maps [12] was conducted by coding the content of the mind maps and then counting the number of times each category was present across the entire sample of 24 mind maps. We were interested in determining the extent to which functional areas of business were present in the mind maps as reflected by their inclusion in the maps. By counting the number of concepts related to various functional areas, the results indicated that some areas were better represented than others. Marketing was represented in 21% of all concepts; business strategy was represented in 21% of all concepts; HRM/OB was represented in 19% of all concepts; finance was represented in 11% of all concepts; management was represented in 9% of all concepts; and other concepts from the areas of Operations Management, MIS, and general concepts such as ethics and economics were represented in the remainder of the concepts included on the maps (see Figure 1).
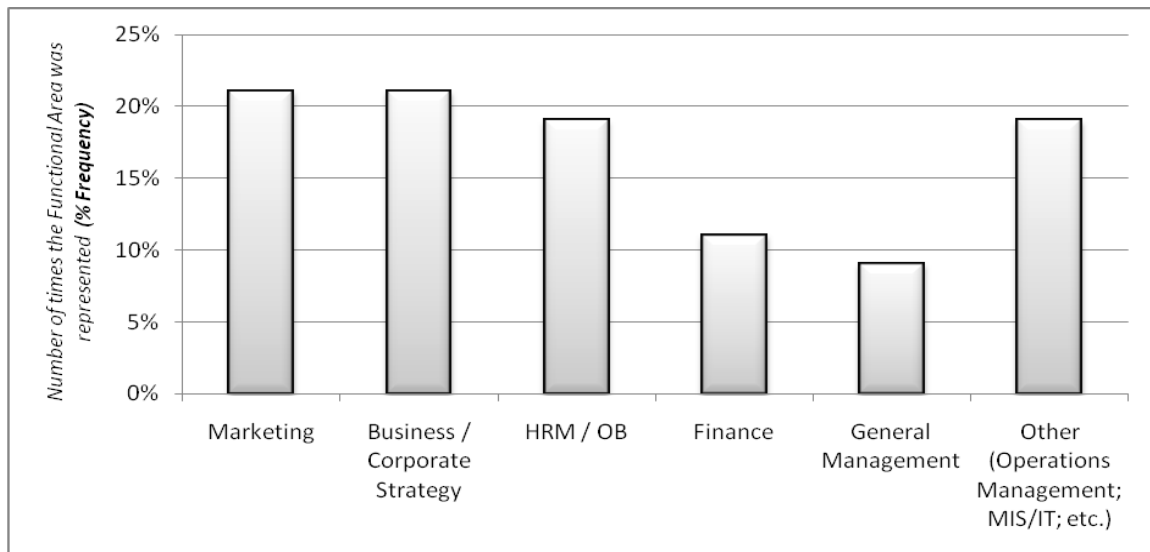
Figure 1: Functional Areas Represented in Mind Maps

It is noteworthy that in the majority of cases, students simply listed the functional areas by name (e.g., marketing) and then included secondary concepts to describe them more fully (e.g., advertising effectiveness, product development etc. in relation to marketing). At this level, the semantic analysis focused on the identification of the primary cluster concepts, as they are sufficient to identify the overall functional orientation of the map elements.

Analysis of Configurations

Conceptual maps including mind maps can also be interpreted in terms of their structure or configuration. Three basic configurations have been identified: chains, spokes and networks [7]. Chains represent knowledge organized sequentially; while spokes are indicative a single level of hierarchy represented by a primary concept and associated satellite or secondary concepts. Networks represent the deepest level of understanding and are characterized by inter-relationships among concepts that usually take the form of a circle in that a closed loop among concepts is present [cf., 12]. Figure 2 shows an example (from [7]) of map structures.

The observed frequency of each configuration was tallied by the researchers, and inter-coder reliability across semantic and configuration groupings was computed to validate the results. Of the three possible configurations of mind maps, only one was observed. Specifically, all of the students' mind maps took the form of a spoke structure so that there was not a single instance of linkage between concepts representing different management functions. That is, there were no observed network structures and no observed circles or loops.
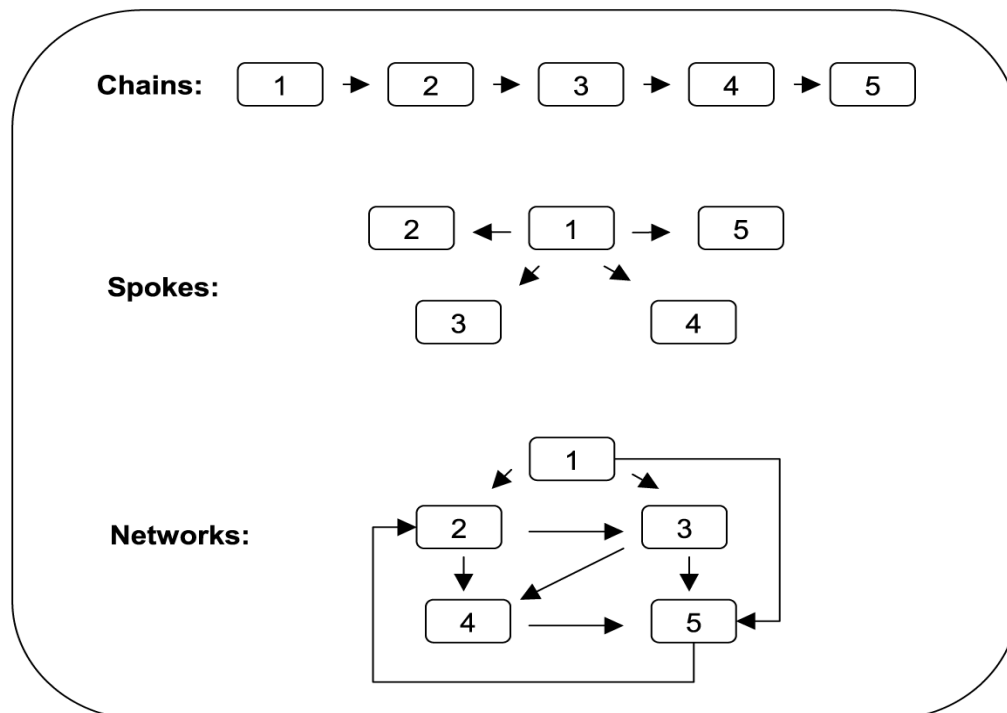
Figure 2: Map Structures Classifications (from [7, p.131])

The only observed linkages were between primary and satellite concepts in spoke structures. This observation shows that while mind maps per se are not necessarily hierarchical, students tended to cluster their representations around hierarchical structures (primary, secondary or tertiary concepts). They focused more on displaying how business

concepts are linearly organized rather than how they are connected (as in a network structure). Network structures integrate new knowledge into rich webs characterized by loops and interconnections so that the addition of new concepts and new loops has the potential to change the qualitative nature of the cognitive structure leading to new insights and deeper understanding.

Findings

Although preliminary and based on a limited sample which is being augmented by additional data collection efforts, our findings suggest that knowledge is atomized such that students have deconstructed a business into discrete, disconnected pieces. The exclusive presence of the spokes structure for the mind maps reinforces this point.

A cognitive structure based on hubs and spokes is designed to add new knowledge by expanding the existing configuration. That is, as students move through the curriculum, new knowledge is likely to be integrated by augmenting the number of concepts that they are familiar with by adding them to a "list" rather than by integrating new concepts into complex knowledge structures [7].

This key finding naturally links our results to another visual mapping structure similar to mind maps: that of concept maps [18]. Novak described concept maps as useful representations of knowledge scaffolding. Concept maps approximate the initial way in which humans learn new concepts by organizing them into clusters, that is by adding them to a "list. In terms of Bloom's learning taxonomies, concept mapping can be useful to display learning at the lowest levels of the pyramid (knowledge capture and comprehension). Network structures represent higher level learning such as analysis, synthesis and evaluation [3].

Since our study showed that students spontaneously chose a hierarchical concept map representation, we tailored our interventions based on the awareness that we needed to start with a focus on lower learning objectives, and slowly guide the students to achieve inter-functional integration-level thinking through analysis, synthesis and evaluation first at the course level. After an on-going evaluation of the results of this phased approach, we hope to be able to extend our reach to curriculum-wide program changes which will be built over-time.

USING CONCEPT MAPS TO FOSTER INTER-FUNCTIONAL INTEGRATION

Course-level focus: Management Information Systems

In order to address the issue of lack of integrative business knowledge emerged from the analysis of the data, and to take into account the evident need for focused knowledge

scaffolding, we decided to target our pilot interventions to graduate level management information systems courses.

The introductory Information Systems Principles course is a pre-requisite course, generally taken in the first year of study by all MBA students. While focused specifically on an overview of information systems (IS) and information technology (IT) in organizations, this course is highly suitable to expose learners to issues of functional and inter-functional integration. The course spans from an initial review of systems used in specific business areas (such as accounting IS, marketing IS, HRIS, etc.) to a broad overview of enterprise systems that are integrated to support and optimize information flows across the entire organization. The course also introduces notions of IT strategy and business-IT alignment as sustainable drivers for a company's competitive advantage. It details how such drivers need to be aligned to and support corporate strategy decisions.

Course Redesign

Earlier editions of this course stressed the importance of understanding the role of IS as a support function to various functional areas, assuming that notions of inter-functional integration would easily emerge from the discussions around enterprise applications and enterprise level performance management systems. The evaluation of such integration was conducted through traditional assessment exercises (such as quizzes or short-essays) that tested students' understanding of key concepts.

In addition to placing more emphasis on cases highlighting the limitations of "silos" approaches to interconnected topics such as business performance management, the course was redesigned to leverage concept mapping both as an assessment as well as a teaching tool to further promote and emphasize the relationships between IS and other areas of business. Specifically, a take-home visualization exercise replaced the final exam. After being introduced to the notion of concept mapping, students are asked to draw a diagram (using Microsoft Visio, PowerPoint Drawing tools or other visualization software) to show how cross-functional information systems contribute to successful business performance (i.e. impact the bottom-line). More specifically, they have to sketch a visual map and answer, through a short write-up, the following question: "*What types of information systems activities (tools, processes, organization or management) are critical to the long-term successful business performance of an organization)*?"

Students are alerted that many solutions are possible but they need to identify and link key IS activities that are critical to business performance by identifying at least four (or more) macro-concepts as well as the sub-concepts related to the macro-concepts.

Preliminary Outcomes

A total of 32 students were exposed to the redesigned approach during Fall 2009. Two sample concept maps submitted by the students are displayed in the appendix. The maps

and the related write ups were evaluated based on the five criteria listed in Table 1, an evaluation scheme based on Novak & Gowin [18] and focused on analyzing concept integration and the establishment of connections among various systems.

#	Criterion	Novak & Gowin, [18] Concept Maps Evaluation Levels
1	The Map clearly differentiates among major concepts and sub-concepts	**Hierarchy level** (representing multiple concepts: main and sub-level)
2	The Map clearly links items that should be related	**Cross-link level** (represented by multiple arrows connecting concepts)
3	The Narrative provides a conceptual definition of the topics selected	**Concept level** (represented through labels)
4	The Narrative describes the connections between various systems	**Proposition-level** (represented by linked meaningful statements)
5	Completeness and Visual Presentation (including grammar/edits, clarity)	N/A in Novak & Gowin (but useful in course assessments)

Table 1: Review Criteria for Concept Maps in MIS Courses

An additional 50 students are currently completing another iteration of the refocused course. By Spring 2010, the authors will have gathered a significant number of maps from multiple sections that, in addition to the criteria presented in Table 1, will be further analyzed based both on semantics and configuration characteristics. Based on the outcomes, additional interventions for longitudinal data collection can be institutionalized. For example, MBA students that were exposed to the MIS course during their first semester may be re-assessed in capstone courses to evaluate progress made during the MBA program or, at minimum, re-evaluate retention (i.e. "*Has the notion of inter-functional integration solidified or eroded throughout the two-year program?*")

CURRICULUM-LEVEL OPPORTUNITIES

We believe that interventions based on the use of cognitive maps may eventually drive curriculum-level change opportunities. For example, they may help the faculty to plan, develop and implement their courses and programs [14]. Concept mapping has been frequently used in curriculum development, especially in science teaching - see for instance Starr & Krajcik [23]. By using concept mapping, faculty can more comprehensively articulate the intended learning outcomes, design new courses, identify linkages between different courses, assess student learning outcomes, and revise their courses, if needed [14].

Prior research associates several advantages to the use of concept maps in curriculum development. For instance, Edmondson [6] used concept mapping to develop a problem-based veterinary education. Edmondson found that using concept mapping principles in curriculum development makes course content more accessible and more easily integrated by students. Others suggest that using concept maps in planning a curriculum or a course in science helps to make the instruction "conceptually transparent" to students [19].

Martin [13] suggests that concept mapping also provides educators a better understanding of what students need to learn and helps in sequencing the courses in an optimal way. Edmondson [6] also argues that concept maps have helped interdisciplinary groups of faculty to move beyond their disciplinary boundaries, and thus assist in the development cross-disciplinary courses and programs.

LIMITATIONS AND FUTURE WORK

Several limitations emerge from this preliminary study. A major limitation is related to the assumption that it may be possible to measure the impact of various educational interventions (i.e. the outcomes from teaching a specific integrative topic) outside of a laboratory environment. Multiple environmental, social and personal development factors play a role in shaping individual cognitive representations of knowledge. Educational outcomes research has always been challenged by the difficulty of achieving laboratory level conditions in holistic learning outcomes assessments, and cognitive mapping clearly suffers from the same limitations. However, because we are interested in learner progressions across a timeline, pre-post experimental settings may nevertheless not be suitable to our study. We expect to be able to build a more cohesive and integrative outcome by continuing to work on a variety of interventions both at the course and at the curriculum level.

Another limitation is related to the small sample size (which may not be representative of the entire MBA population) and small geographical focus (a single AACSB institution in the Northeast of the United States) that spurred some of our interventions. Current and future work is expanding our sample size across course-levels and institutions, and also beyond the boundaries of the US business education environment. To date, we have collected mind maps in business schools in over five countries in Europe, Asia and North Africa with the objective to understand whether problems of inter-functional integration span across national approaches to business education.

We cannot ignore limitations that may be related to our data collection procedures and instruments. With regard to procedures, while we are not comparing data collected in the strategic management course with the data collected in the management information systems course directly (as the former focused on initial benchmarking and the latter focused on assessing the results of a specific treatment), we acknowledge that data collection in the strategic management course was voluntary. In the MIS course, it was

integrated as part of the course deliverables (and thus assigned a specific score). This aspect may have increased the motivation of the students, and thus led to an improved outcome which is not necessarily a reflection of a successful intervention. Another issue could be that of instrumental validity. Given that mapping abilities may vary based on cognitive skills and learner's preferences that are often independent from the instructional experience, we may need to further investigate whether we can effectively use mind maps as representation of student learning in a specific course (rather than student learning in general).

Despite the above limitations, the picture that emerges from our pilot data collection efforts and remedial initiatives confirms that outcomes-related problems exist. It also shows that cognitive mapping may indeed be a useful tool to solve such problems by eliciting connections and integration at the individual-level (student learning), enabling actionable transformations at the course-level, and finally promoting sustainable interconnection at the program-level (curriculum).

REFERENCES

[1]   Ackoff, R. (2002.) Russel L. Ackoff: interview by Glenn Detrick. *Academy of Management Learning & Education*, *1*, 56-63.

[2]   Bennis, W. & O'Toole, J. (2005). How business schools lost their way. *Harvard Business Review*, *83*, 96-104.

[3]   Bloom, B. S. (1956). *Taxonomy of educational objectives: cognitive domains*. New York: David McKay Company.

[4]   Budd, J. (2004). Mind maps as classroom exercises. *Journal of Economic Education*, *35*, 35-46.

[5]   Buzan, T. (1995). *The mindmap book*. London: BBC Worldwide Publishing.

[6]   Edmondson, K. M. (1995). Concept mapping for the development of medical curricula. *Journal of Research in Science Teaching, 32*(7), 777-793.

[7]   Hay, D. & Kinchin, I. (2006). Using concept maps to reveal conceptual typologies. *Education and Training*, *48*, 127-142.

[8]   Hay, D. & Kinchin, I. (2008) Using concept mapping to measure learning quality. *Education and Training*, *50*, 167-182.

[9]   Kern, CS, Bush, KL, McCleish, JM (2006). Mind-mapped care plans: integrating an innovative educational tool as an alternative to traditional care plans. *Journal of Nursing Education*, *45*, 112-119.

[10]   Kilpatrick, J., Dean, K. & Kilpatrick, P. (2008). Philosophical concerns about interpreting AACSB assurance of learning standards. *Journal of Management Inquiry*, *17*, 200-212.

[11]   Kinchin, I.M. (2000). Concept mapping activities to help students understand photosynthesis − and teachers understand their students. *School Science Review*, *298* (82), 11-14.

[12]   Lourdel, N., Gondran, N., LaForest, V., Debray, B. & Brodhag, C. (2007). Sustainable development cognitive map: a new method of evaluating student understanding. *Journal of Sustainability in Higher Education*, *8*, 170-182.

[13]   Martin, D. J. (1994). Concept mapping as an aid to lesson planning: a longitudinal study. *Journal of Elementary Science Education, 6*(2), 11-30.

[14]   McDaniel, E., Roth, B., & Miller, M. (2005). Concept mapping as a tool for curriculum design. *Issues in Informing Science and Information Technology*, 505-513.

[15]   Mento, A., Martinelli, P. & Jones, R. (1999). Mind mapping in executive education: applications and outcomes. *Journal of Management Development*, *18*, 390-407.

[16]   Mintzberg, H. (2004). *Managers not MBA's: A hard look at the soft practice of managing and management development*. San Francisco, CA: Berrett-Koehler.

[17]   Mueller, A., Johnston, M. & Bligh, D.   (2001). Mind-mapped care plans: A remarkable alternative to traditional nursing care plans. *Nurse Educator*, *26*, 75-80.

[18]   Novak, J. & Gowin, D (1984). *Learning how to learn*. NY: Cambridge University Press.

[19]   Novak, J. (1998). *Learning, creating, and using knowledge: concept maps as facilitative tools in school and corporations*. Mahwah, NJ: Lawrence Erlbaum Associates.

[20]   Pfeffer, J. & Fong, C. (2004). The business school "business": some lessons from the US experience. *Journal of Management Studies*, *41*, 1501-1520.

[21]   Porter, L. & McKibbion, L. (1988). *Management education and development: drift or thrust into the 21st century*. New York: McGraw-Hill.

[22]   Rubin, R. & Dierdorff, E. (2008).   How relevant is the MBA? Assessing the alignment of required curricula and the required managerial competencies. *Academy of Management Learning and Education*, *6*, 84-101.

[23]   Starr, M. L. & Krajcik, J. S. (1990). Concept maps as a heuristic for science curriculum development: toward improvement in process and product. *Journal of Research in Science Teaching*, *27*(10), 987-1000.

[24]   Starkey, K. & Tempest, S. (2009). The winter of our discontent: The design challenge for business schools. *Academy of Management Learning & Education*, *8,* 576-586.

[25]   Wycoff, J. (1991) *Mindmapping: Your personal guide to exploring creativity and problem-solving.* NY: Berkley Books.

[26]   Zampetakis, L. & Tsironis, L. (2006). Creativity development in engineering education: the case of mind mapping. *Journal of Management Development*, *26*, 370-380.

APPENDIX
SAMPLE CONCEPT MAPS (MIS COURSES)

What types of information systems activities (tools, processes, organization or management) are critical to the long-term successful business performance of an organization?

Chotikorn Buranarachada
MIS645 - Final Exam
Fall 2009 - Dec 10th, 2009
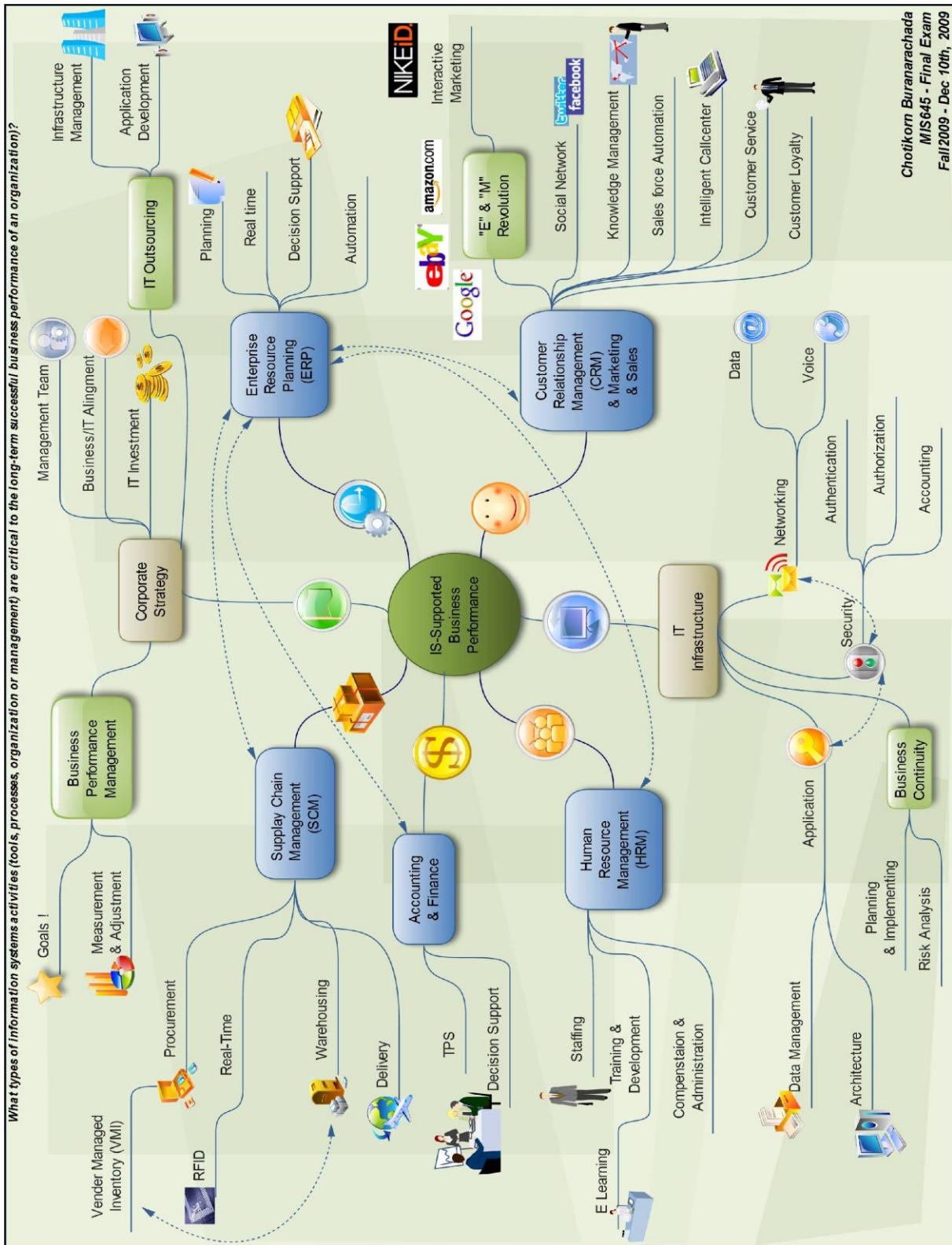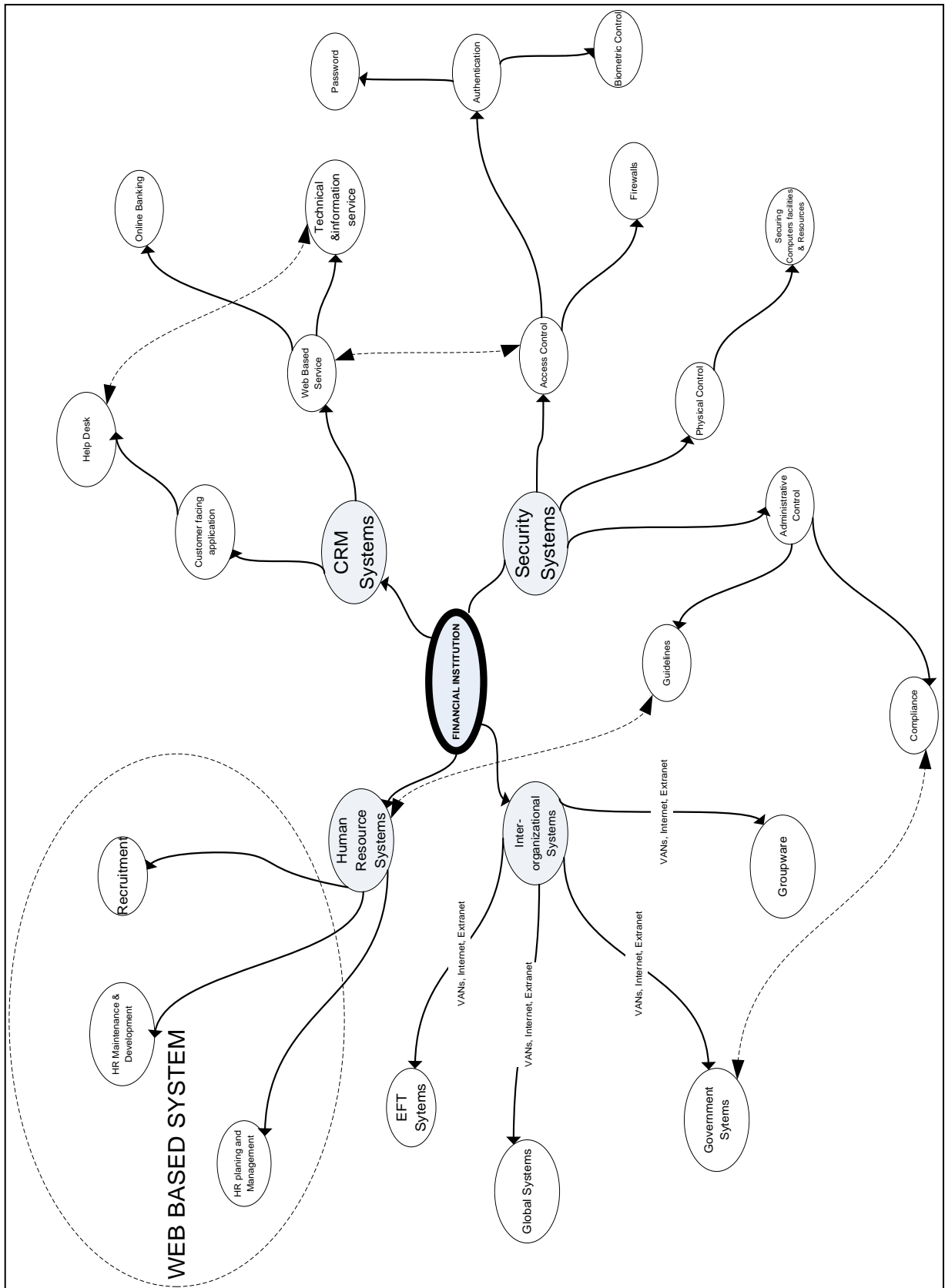
A GROUNDED THEORY APPROACH TO INVESTIGATING CHANGE MANAGEMENT SKILL ACQUISITION BY MBA STUDENTS: A RESEARCH PROPOSAL FOR DISCUSSION

Nicole Jean Christian, MPA (Walden University Doctoral Student)
Dowling College
150 Idle Hour Blvd.
Oakdale, NY 11769
Email: Nicolejchristian@gmail.com, christin@dowling.edu
Ph: 631-879-1588

ABSTRACT

This paper is an overview for discussion of the author's proposed doctoral research. The author's plan for the dissertation is to explore the literature, perceptions of students and MBA curricula as it relates to change management to discover and understand more about what students learn about change management and change agency specifically. The goal of this study is to generate and understand theory on student's preparedness as change agents and change leaders through completion of typical MBA programs.

Keywords: MBA curriculum, change management, organizational change, grounded theory.

INTRODUCTION

Leaders inspire, innovate, develop others, challenge the status quo and have a long-term view. They basically look to change the way things are. Organizational systems help align people's behavior to achieve organizational goals and these systems must be agile in order to adapt to changes in the environment [3] [4] [5] [7]. Unfortunately, organizational systems are resistant to change; we fall in love with the systems we create. Therefore, mastering adaptation and transformation are at the cornerstone of a new kind of leadership: leadership that can and does embrace and support change [4] [5]. It takes trained and developed leaders to be able to successfully implement and manage change processes that could take months or even years. Certainly, a very necessary skill for leaders of tomorrow is the ability to create and implement organizational change. MBA programs are at the forefront of change management preparation for leaders since MBA programs are where future business leaders incubate, develop and are primed for eventual organizational responsibility.

Building skill competency is an important function of MBA programs, but those skills can change almost as soon as they are learned [2]. Student preparedness for the marketplace and the world at large is a cornerstone of most MBA programs. The goal of MBA programs is to prepare students to identify, address and solve conflicts, build teams and to be a stabilizing force in the ever changing workplace. A 2009 study on IT academic programs and student preparedness

found that hiring managers sought graduates with MIS degrees that also had soft skills, high level interpersonal skills and business acumen [2]. The study found that students needed to have competence in skill areas outside of the curriculum such as budgeting, interpersonal skills and communication [2].

Recently, even the use of textbooks and student interaction with textbooks has been challenged. One study found that textbook interaction among students varied based on individual motivation; not all students were found to use course textbooks optimally, which meant that building competency in areas not covered in or reviewed by students in textbooks was left to the responsibility of the instructor [8].

Instructors sometimes do provide students with skills and insight that may or may not be available or illuminated in the traditional course material. Instructors often augment classroom learning with relevant contemporary skill building exercises in competencies not outlined in traditional materials. One such example is in soft skills such as communication, interpersonal and inter-team dynamics and change management. It is critically important for students to be given the skills to help them be flexible, navigate change and facilitate change processes within a team and organization-wide. Currently, these skills may not be critical components of MBA curriculum. The aim of this study will be to present a student-experience centered perspective on change management skill acquisition through traditional MBA programs.

GROUNDED THEORY APPROACH

Grounded theory is one of the more "validated and tested" [10, p.340] qualitative research methods. The purpose of grounded theory is to derive theory from data collected in a "natural setting" [6, p. 144]. Grounded theory is an inductive method of theory construction whereby through observations of certain aspects of life a theory is drawn from the words, action, imagery and ideas introduced by the individuals under study [1] [6] [9] [10]. These words, imagery and ideas introduced by those studied form 'emergent themes' which are then transcribed, coded and then compared to prior research (i.e. literature). Then the researcher "iterates frequently between data analysis and theory development" [8, p. 23] to help explain or understand more about the process under study. Once no more additional information is found in the data the iteration ends. Grounded theory studies are particularly useful when there is a lack of theory and understanding about particular phenomena [6] [8]. This is the case for MBA curriculum as it relates to change management and change agency preparation.

RESEARCH QUESTION(S)

What is the affect, if any, of MBA programs on student's preparedness as change agents/leaders? Or said another way, do typical MBA programs help strengthen a student's ability to become a better change agent? And most importantly, what aspects of MBA programs are critical in this regard and how can we strengthen those areas?

RATIONALE FOR STUDY

Organizations, public and private, are influenced by outside forces that make it necessary to alter procedures and systems in order to stay successful. Understanding the patterns of and different types of organizational change allows people in organizations to shift from being 'reactive' to becoming 'active' participants in change processes. Once able to be 'active' in the change process, leaders, workers and executives can guide the organization through change to ensure greater success. In the 21st century and beyond, MBA students, as future business leaders will need the knowledge, skill and ability to facilitate, embrace and implement change in the workplace.

DATA COLLECTION METHODS

The target population for this study will be MBA students and graduates in two colleges on Long Island. The population will be divided between those nearing graduation (one semester away from graduation) and those newly graduated (one to two years with MBA). Data will be generated through a demographic questionnaire, interviews (small group and individual) and focus groups with current MBA students and recent MBA graduates. The sessions will be transcribed, coded and then analyzed to examine emergent theories. Keywords will be identified from contextual literature, interviews and focus group discussion so when data from any of the focus groups or interviews illustrate a particular theme that interview/discussion will be transcribed, coded and analyzed in full.

RESULTS/DISCUSSION

The goal of this study is to generate and understand theory on student's change management preparedness through completion of typical MBA programs. Once the theory is identified, further research will need to be completed to determine how, if necessary, MBA curriculum can be augmented to include change management as a core competency.

The author's goal is to have completed for discussion and presentation an initial research design and data collection methodology by the April 2011 NEDSI conference.

Nicole Jean Christian holds her Master of Public Administration with a concentration in Nonprofit Management from George Mason University (2001). Currently, she is a Senior Adjunct Professor of Management and Leadership at Dowling College in Oakdale Long Island. She is pursuing a PhD in Applied Management and Decision Sciences from Walden University with a specialization in knowledge and learning management. She is slated to defend in 2011.

Professor of Business, Walden University; my husband Majai Q. Goodine, my mother Barbara Jean Christian, MBA, alumnus of Dowling College; the Christian family, and in memory of my father, Philip A. Christian.

REFERENCES

[1] Babbie, E. (1995). The practice of social research. 7th Ed. California: Wadsworth Publishing.

[2] Havelka, D., Merhout, J. (2009). Toward a theory of iformation technology professional competence. Journal of Computer Information Systems, Winter 2009, pp. 106-116.

[3] Hughs, R. (2009). Leadership: Enhancing the lessons of experience. New York: McGraw-Hill Irwin.

[4] Kotter, J. Cohen, D. (2002). The heart of change: Real-life stories of how people change their organizations. Massachusetts: Harvard Business School Press.

[5] Kotter, J. (1990). A force for change: How leadership differs from management. New York: The Free Press.

[6] Leedy, P., Ormrod, J. (2005). Practical research: Planning and design. 8th Ed. New Jersey: Pearson Education, Inc.

[7] Lussier, R. (2008). Human relations in organizations: Applications and skill building approach. Boston, Ma: McGraw-Hill Irwin.

[8] Phillips, B., Phillips, F. (2007). Sink or swim: Textbook reading behaviors of introductory accounting students. Issues in Accounting Education, 22(1), pp. 21-44.

[9] Thomas, G. (2006). Reinventing grounded theory: Some questions about theory, ground and discovery. British Educational Research Journal 32(6), p. 767-795.

[10] Timlin-Scalera, R., Ponterotto, J., Blumberg, F., Jackson, M. (2003). A grounded theory study of help-seeking behaviors among white male high school students. Journal of Counseling Psychology, 50(3), pp. 339-350.

THE ABILITY OF WESLEY COLLEGE STUDENTS TO APPLY LEARNING OBJECTIVE FROM THEIR MATH COURSE TO A FOLLOW-UP COURSE

Derald E. Wentzien, Ph.D, Wesley College
Mary Jo Benson, Wesley College
Melissa Earley, Wesley College

ABSTRACT

The ability of Wesley College Students to learn and retain concepts from an Intermediate Algebra and a Math Concepts and Operations course was analyzed. Two specific learning objectives, the use of proportional reasoning and the use of formulas, were assessed using scoring rubrics. Histograms and a paired t-test were used to assess and compare the rubric score for the learning objective in the math course taken in Fall 2009 to the rubric score for the same learning objective for the same student in a course taken in spring 2010. The output was then used to determine if the students retained the concept and were able to solve similar problems in a course taken the following semester. The results indicated that the students were able to correctly apply the use of proportional reasoning, but they were not able to correctly apply the use of formulas in a follow-up class.

BACKGROUND

Wesley College is a small liberal arts college located in Dover, DE with an enrollment of approximately 1,500 undergraduate students. All students are required to take at least two analysis courses. Wesley College is accredited by Middle States and the assessment of student learning outcomes is one of the standards Middle States evaluates during the 10-year re-evaluation process. The faculty in the math department at Wesley College decided to use scoring rubrics to assess student learning outcomes in the math classes. The assessment occurs on examinations administered during the math class, and on a survey administered during a follow-up course for which the math class is a prerequisite. Not all students take a follow-up course, but the multiple assessments make it possible to study whether or not the students are mastering and retaining the learning objective.

Scoring rubrics are frequently used to assess student learning. After decades of using multiple choice items for assessments, educators realized the need to change current assessment practice and started moving towards more authentic performance assessments. These kinds of assessments present students with tasks that simulate real world challenges, and can be answered in more than one way. Consequently, more detailed scoring rules are needed to assess the variety of problems solving techniques used by students.

Scoring rubrics are descriptive scoring schemes that are developed by teachers and other evaluators to guide the analysis of the products or process of students' efforts [1]. By developing

a pre-defined scheme for the evaluation process, the subjectivity becomes more objective. The purpose of scoring rubrics is to provide evaluators an effective tool that can help them focus on a target and understand what attributes an item or task truly measures.

Scoring rubrics have three main functions to teachers, test developers, decision makers, and students [2]:

- Scoring rubrics provide uniform, objective criteria for judging a performance assessment. Having clear, well-established scoring criteria will improve the agreement between scorers and will reduce bias.
- Scoring rubrics provide established expectations for teachers and students that help them identify the relationship among instruction, learning, and assessment. This practice reduces test anxiety and provides more feedback for both teacher and student.
- Scoring rubrics reinforce a focus on content, providing performance standards of student work. This practice minimizes or eliminates distractions caused by peripheral variables.

The purpose of the evaluation should determine whether to use a holistic or analytic scoring rubric. Holistic scoring produces a single score based on an established scale. Holistic scoring is preferred when an overall judgment is desired and when skills being assessed are highly interrelated. In analytic scoring, each critical dimension of the performance criteria is judged independently and awarded an individual score, and then an overall score is given to the examinee for the item. Analytic scoring provides more detailed information that can be used for diagnostic purposes for both teachers and evaluators. A holistic scoring rubric was used in this assessment.

METHODOLOGY

The assessment of the use of proportional reasoning and the use of formulas were assessed in five sections of Intermediate Algebra and six sections of Math Concepts and Operations in Fall 2009 and courses for which the math courses were a prerequisite in Spring 2010. The follow-up courses included a nursing dosage calculation course, an economics course, and a chemistry course. A scoring rubric was designed for a question involving the use of proportional reasoning and a question involving the use of formulas. The faculty in the mathematics department selected a standardized question which was used by all of the instructors in both courses on an examination. The same scoring rubric and standardized questions were used in Spring 2010.

The analysis consists of histograms for the scores received on the examinations in the math courses and a paired t-test between scores received on the examination and the scores received on the question administered during the follow-up course. The paired t-test was used to determine if the mean score of the question given in the Spring 2010 class was lower than the mean score of the question from the examination from the Fall 2009 class for each of the learning objectives.

RESULTS

A total of 116 scores were recorded for the problem involving the use of proportional reasoning on an examination in the Intermediate Algebra and Math Concepts and Operations courses. Ninety-two students received a 3 or 4 on the question representing 79.31% of the 116 students who participated in the assessment. The faculty in the math department established a goal of 70% of students scoring a 3 or 4 on the question and was satisfied that the students were learning how to use proportional reasoning to solve a problem. Figure 1 presents the distribution of rubric scores for the use of proportional reasoning.

Figure 1
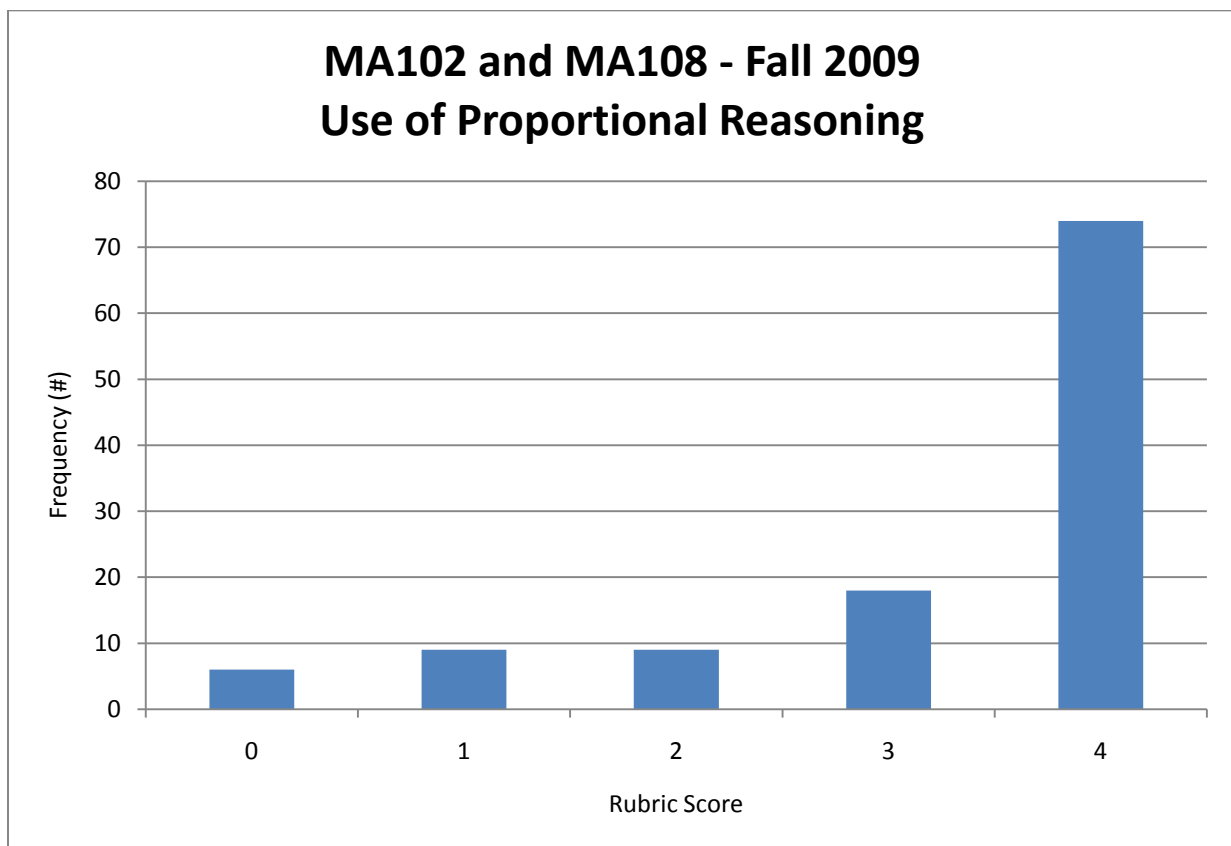
A total of 127 scores were recorded for the problem involving the use of formulas on an examination in the Intermediate Algebra and Math Concepts and Operations courses. Ninety-two students received a 3 or 4 on the question representing 72.44% of the 127 students who participated in the assessment. The faculty in the math department established a goal of 70% receiving a 3 or 4 on the question and was satisfied that the students were learning how to use proportional reasoning to solve a problem. Figure 2 presents the distribution of rubric scores for the use of formulas.

Figure 2

MA102 and MA108 - Fall 2009
Use of Formulas
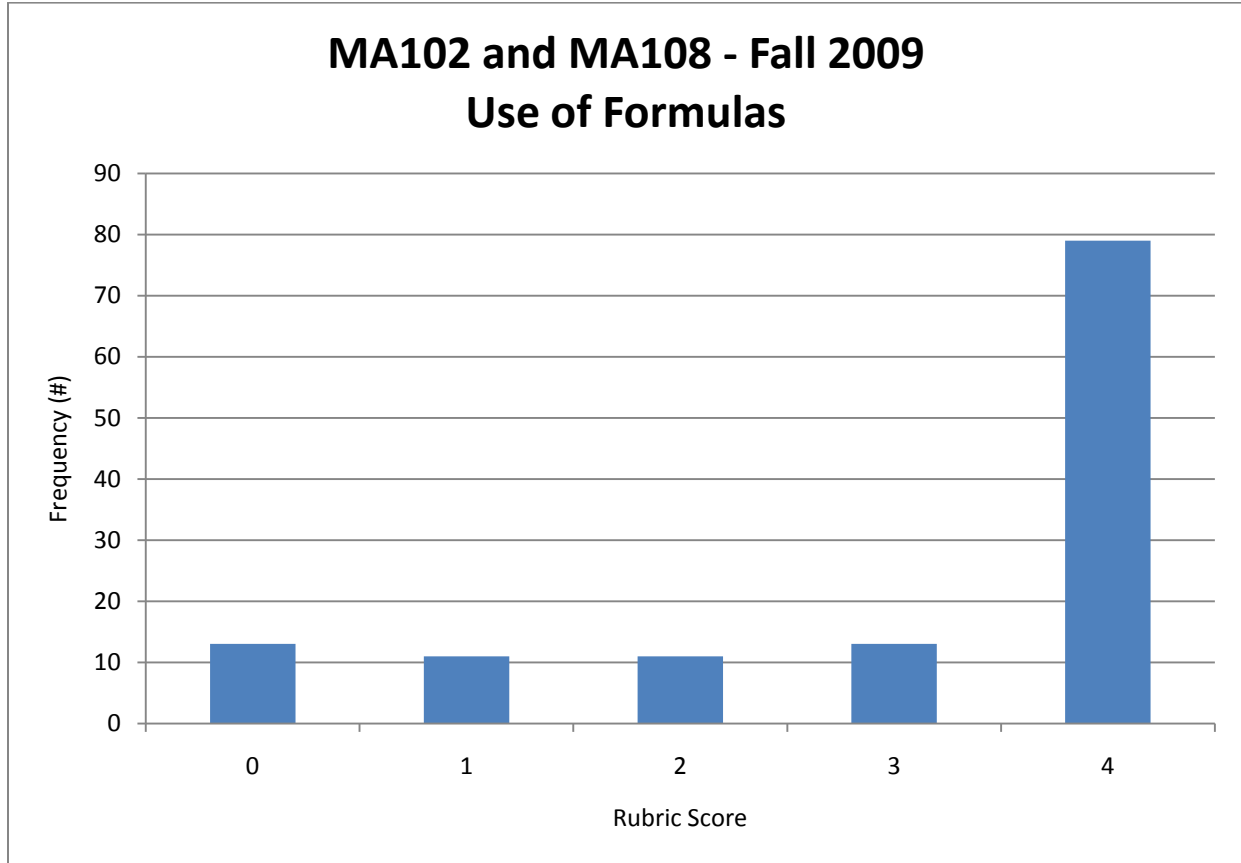
In Spring 2010, students in a nursing dosage calculation course, an economics course, and a chemistry course were given problems involving the use of proportional reasoning and the use of formulas. Their performance was scored on the same five-point rubric used in Fall 2009. The scores for the students who also answered the questions in their math class from the Fall 2009 were compiled. The two scores were recorded for a total of eighteen students in the use of proportional reasoning and a total of sixteen students in the use of formulas. Two paired t-tests were then performed to determine if the mean score was lower in the courses taken in Spring 2010 for the use of proportional reasoning and the use of formulas.

The t-test for the mean rubric score measuring understanding of the use of formulas was 2.366 with a p-value of 0.0159. Therefore, there is significant evidence in the sample to support the claim that the mean score on the use of formulas was lower in the follow-up course. It appears that the students were not able to retain the knowledge they learned about the use of formulas in a class taken the following semester.

The t-test for the mean rubric score measuring understanding of the use of proportional reasoning was 0, indicating that the mean scores in the math class and a course taken the following

semester were identical. Therefore, it appears that the students were able to retain the knowledge they learned about the use of proportional reasoning in their next class.

CONCLUSION

Since the results indicated that the students were not retaining the information for the use of formulas, the math department at Wesley College is going to make adjustments to the way they teach formulas this fall to the new freshman. A greater emphasis will be placed on specific applications and the use of formulas in future classes. Hopefully, the students will realize that they will need to use formulas in the future and that retention of this knowledge is critical to their success in future classes. The test supports the claim that the students at Wesley College are retaining the information for the use of proportional reasoning, so the math department has no need to change the way proportional reasoning is being taught at this time.

REFERENCES

[1] Brookhart, S.M. (1999). *The Art and Science of Classroom Assessment: The Missing Part of Pedagogy*. ASHE-ERIC Higher Education Report (Vol 27, No. 1) Washington, DC: The George Washington University, Graduate School of Education and Human Development.

[2] Westat (2001). *ABC on Scoring Rubrics Development for Large Scale Performance Assessment in Mathematics and Science*. Posted online at http://lsc-net.terc.edu.

STUDENT ATTITUDES TOWARD BUSINESS STATISTICS: COMPARISON OF DATA ANALYSIS APPROACH VS. TRADITIONAL APPROACH

Borga Deniz
Christopher Newport University, Luter School of Business, 1 University Place, Newport News, VA 23606
Phone: (757) 594 8915, e-mail: borga.deniz@cnu.edu

Robert B. Hasbrouck
Christopher Newport University, Luter School of Business, 1 University Place, Newport News, VA 23606
Phone: (757) 594 7174, e-mail: rhasbro@cnu.edu

Harland Hodges
College of Charleston, School of Business, 66 George Street Charleston, SC 29424
Phone: (843) 9534810, e-mail: hodgesh@cofc.edu

ABSTRACT

In this study we compare two business statistics teaching approaches based on their effect on student attitudes. The first approach is the *traditional approach* which focuses on mathematical calculations and procedures; the second approach is the *data analysis approach* which focuses on reading, understanding and interpreting statistical data with use of a statistics software package. In order to do the comparison we measure students' attitudes towards statistics using a validated survey, both before and after a 15-week course for each approach. Our results indicate that the data analysis approach provides significantly more positive attitudes toward statistics than the traditional method.

INTRODUCTION AND PROCEDURE

Student attitude toward business statistics is often not very positive, and this could affect student learning. The literature related to student attitudes toward statistics includes [1], [2], [4] and references therein. In [4] the authors compared online (E-learning) and traditional methods and they found that significant differences exist in student attitudes towards to learning business statistics between the different modes of learning. In [1] the author investigates whether a student-designed data collection project yields more positive attitude toward statistics and finds that inclusion of a project may not significantly impact students' attitudes toward statistics. There are also studies that compare computer assisted instruction and the traditional method of teaching such as [2] and [3]. In [2] it was observed that students who take statistics with computer assisted instruction do better in the course. However it is concluded in [3] that technology availability in classroom does not affect learning if statistics software is used for teaching statistics.

In this study we investigate if using a data analysis approach makes a difference in student attitude toward statistics compared to traditional approach. Traditional approach is a commonly used approach and it focuses on mathematical calculations and procedures. It is grounded in the mathematical process of business statistics. An alternative approach is data analysis approach which extensively allows students conduct data analysis. In our case data analysis focuses on reading, understanding and interpreting statistical data with use of a Microsoft Excel based statistics software package called VISA. Students who are taught via the data analysis approach take their business statistics class in a computer lab and conduct data analysis while solving numerous examples together with the professor during class; they also take their tests using computer. In this study we compare two groups of students' attitudes toward statistics, one studying using the traditional approach, and the other studying using the data analysis approach. The comparison is based on students' attitudes towards statistics measured using a validated survey, both before and after the 15-week course for each approach. Students provide Likert-type responses to statements in the survey. The survey is called *Attitudes Toward Statistics* and it is developed by Wise in [5]. Students take the survey at the beginning and at the end of their business statistics course. Students who learned statistics via data analysis approach and traditional approach were in different sections of the course. The survey contains has two subscales: Subscale-1 is on *Attitude toward field* and Subscale-2 is on *Attitude toward course*. Comparisons are made with students who are in a sophomore-level business statistics course. Sample sizes were 40 for the data analysis approach and 11 for the traditional approach in this study.

ANALYSIS AND RESULTS

In order to check if there were statistically significant differences in mean responses at the beginning of the semester for the two groups with respect to the two subscales we used the following null hypotheses:

$$\mu_{D,Pre,1} = \mu_{T,Pre,1}$$

and

$$\mu_{D,Pre,2} = \mu_{T,Pre,2}$$

where

$\mu_{D,Pre,1}$ is the mean score for *data analysis approach*, beginning of the semester, Subscale-1,

$\mu_{T,Pre,1}$ is the mean score for *traditional approach*, beginning of the semester, Subscale-1,

$\mu_{D,Pre,2}$ is the mean score for *data analysis approach*, beginning of the semester, Subscale-2,

$\mu_{T,Pre,2}$ is the mean score for *traditional approach*, beginning of the semester, Subscale-2.

Table-1 shows means, standard deviations, t-statistics and p-values which indicated there were no statistically significant differences (p-value > 0.05) in mean responses for the two groups with respect to the two subscales. The two groups appeared to be sufficiently similar in attitudes toward the statistics field and the statistics course as measured by these subscales at the beginning of the semester.

Beginning of the semester	Data Analysis Approach		Traditional Approach			
Subscale	Mean	St. Dev.	Mean	St. Dev.	t	p
1- Attitude towards field	3.63	0.36	3.56	0.48	0.5427	0.5898
2- Attitude towards course	3.22	0.86	3.32	0.72	-0.3397	0.7355

TABLE-1. MEAN BEGINNING-OF-SEMESTER RESPONSES ON ATTITUDE

In order to check if there were statistically significant differences in mean responses at the end of the semester for the two groups with respect to the two subscales we used the following null hypotheses:

$$\mu_{D,Post,1} = \mu_{T,Post,1}$$

and

$$\mu_{D,Post,2} = \mu_{T,Post,2}$$

where

$\mu_{D,Post,1}$ is the mean score for *data analysis approach*, end of semester, Subscale-1,

$\mu_{T,Post,1}$ is the mean score for *traditional approach*, end of semester, Subscale-1,

$\mu_{D,Post,2}$ is the mean score for *data analysis approach*, end of semester, Subscale-2,

$\mu_{T,Post,2}$ is the mean score for *traditional approach*, end of semester, Subscale-2.

At the end of the semester we see significant differences between the two groups. Table-2 shows means, standard deviations, t-statistics and p-values for testing whether the two groups differed on any of the two subscales of the end of the semester administration of the survey. There are statistically significant differences ($p < 0.01$) in mean responses for the two groups with respect to the subscales. These results show that the attitude toward the statistics field and toward the course is significantly more positive with the data analysis approach.

End of the semester	Data Analysis Approach		Traditional Approach			
Subscale	Mean	St. Dev.	Mean	St. Dev.	t	p
1-Attitude towards field	3.70	0.53	3.14	0.40	3.2333	0.0022
2-Attitude towards course	3.41	0.84	2.35	0.62	3.8648	0.0003

TABLE-2. MEAN END OF THE SEMESTER RESPONSES ON ATTITUDE

We also checked if improvement in attitude from *beginning of the semester* to *end of the semester* differed in the two groups. Therefore we had the following null hypotheses:

$$\mu_{D,Post,1} - \mu_{D,Pre,1} = \mu_{T,Post,1} - \mu_{T,Pre,1}$$

and

$$\mu_{D,Post,2} - \mu_{D,Pre,2} = \mu_{T,Post,2} - \mu_{T,Pre,2}$$

Post minus Pre	Data Analysis Approach		Traditional Approach			
Subscale	Mean	St. Dev.	Mean	St. Dev.	t	p
1-Attitude towards field	0.07	0.57	-0.42	0.55	2.5321	0.0146
2- Attitude towards course	0.19	1.18	-0.96	0.80	3.0363	0.0038

TABLE-3. MEAN DIFFERENCE SCORES ON ATTITUDE SUBSCALES

Table-3 shows means, standard deviations, t-statistics and p-values for testing whether the mean difference scores for the two groups are equal. There were significant differences (p-value < 0.05) in mean difference scores for the two groups with respect to the two subscales.

To further explore the data, one-sample t-tests were performed to determine if the mean difference scores (post-pre) were significantly different from zero for the two subscales when *data analysis approach* group and the *traditional approach* group were considered separately. For the analysis we had the following null hypotheses:

$$\mu_{D,Post,1} = \mu_{D,Pre,1} \quad ; \quad \mu_{D,Post,2} = \mu_{D,Pre,2}$$

and

$$\mu_{T,Post,1} = \mu_{T,Pre,1} \quad ; \quad \mu_{T,Post,2} = \mu_{T,Pre,2}$$

Post minus Pre	Data Analysis Approach		Traditional Approach	
Subscale	t	p	t	p
1-Attitude towards field	0.7615	0.4509	-2.5425	0.0292
2-Attitude towards course	1.0088	0.3193	-4.0236	0.0024

TABLE-4. SIGNIFICANCE TEST FOR MEAN DIFFERENCE SCORES ON ATTITUDE SUBSCALES

Table-4 contains the test results. The results indicate a significant decline in attitude in traditional approach (p-value < 0.05) for both subscales. There is no significant change (p-value > 0.05) in attitude in data analysis approach responses.

CONCLUSIONS AND RESEARCH DIRECTIONS

The results of the test groups indicate that the data analysis approach for teaching business statistics potentially can provide significant and positive results. We are encouraged by these findings and are in the process of conducting a larger scale study. Also, we are interested looking at a factor analysis that would provide additional insights on the relationship between attitude toward statistics and other possible factors.

REFERENCES

[1] Carnell, L. J. (2008). The Effect of a Student-Designed Data Collection Project on Attitudes Toward Statistics. *Journal of Statistics Education*, Volume 16, Number 1.

[2] Ragasa, C.Y. (2008).A Comparison of Computer-Assisted Instruction and the Traditional Method of Teaching Basic Statistics. *Journal of Statistics Education* Volume 16, Number 1

[3]Shaltayev, D., Hodges, H., & Hasbrouck, R.B. (2010). VISA: Reducing Technological Impact on Student Learning in an Introductory Statistics Course. *Technology Innovations in Statistics Education*, 4(1)

[4] Suanpang, P., Petocz, P., & Kalceff, W. (2004). Student Attitudes to Learning Business Statistics: Comparison of Online and Traditional Methods. *Educational Technology & Society*, 7 (3), 9-20.

[5] Wise, S. L. (1985). The Development and Validation of a Scale Measuring Attitudes toward Statistics. *Educational and Psychological Measurement*, 45, 401-405.

A DIFFERENT (BETTER?) WAY OF ASSIGNING AND GRADING QUANTITATIVE HOMEWORK

Larry C. Meile, Boston College
120 Commonwealth Ave, Chestnut Hill, MA, 02467
meile@bc.edu, 617-552-0158

ABSTRACT

A method for improving grader productivity while maintaining the value of homework assignments is offered. Short one-problem quizzes are substituted for turned-in homework which encourages students complete and understand assigned homework problems while minimizing the grading effort on the part of the instructor. Problem sets are assigned but not collected nor graded (solutions are provided on-line). One of these problems is given as the quiz which is graded in detail using a uniform Excel-based grading template.

Keywords

Teaching methods, evaluation, grading, homework

One of the ongoing challenges instructors face is to get students to practice the material that is presented in class. Typically an instructor will select and assign a set of homework problems for each student to work. The students work on them, turn them in, and then they are graded either by the instructor or by someone else such as a graduate assistant. One problem with this approach is assuring that the students have done the homework on their own, not just copied someone else's work. Next, if there is to be any real feedback provided, the grading of the assigned material can be tedious and time consuming. Another problem with collecting and grading homework is that feedback does not occur until the graded papers are returned. The amount of time consumed and the desire to provide prompt feedback puts pressure on the instructor to just "check off" the papers which tends to minimize the quality of the feedback provided to the students. We have come up with a novel way of encouraging students to practice all the assigned material, yet not have to grade each and every assigned problem.

What we do is assign problems, such as those found at the end of the text chapter, that are representative of the topics we want the students to understand. Typically there will be a half-dozen or so selected problems in each problem set. This problem set would be similar to the one we would assign if collecting and grading problems except that, if we are collecting and grading them, solutions to these problems cannot be available while the students do the work. For our problem sets, we supply solutions to these problems, either by assigning problems for which answers appear at the end of the text or are provided through online sources. This way the students get immediate feedback as to whether or not they worked the problem properly.

One potential issue with having solutions available is that students don't necessarily learn how to "start" a problem. They look at the solution and feel they understand the solution procedure but

cannot come up with it when given a new problem to solve. The quiz approach requires that students understand both how to approach problems as well as the mechanics of solving them. So, if some students feel they learn better by looking at the solution and backing into how the problem should be solved, let them have at it. Furthermore, they can work in groups because no longer is the answer they develop for the homework problem part of the graded material. We neither collect nor grade the assigned problems.

How do we ensure that the students work through them? We hold periodic quizzes (typically seven in a 14-week semester) that are nothing other than one of the problems that have been assigned. This encourages the students to work all the problems and assure that they know how to get correct answers. When an assigned homework problem is used as a quiz, we typically change one or more numbers so that the students can't just parrot the answer, but have to apply the correct solution technique anew. And we will break the problem down into sub-questions so that all parts of the solution procedure can be examined and partial credit granted. Since students can be expected to have already worked the problem, the quiz need not take much time at all – typically just fifteen minutes to complete.

Quizzes are given and collected in lieu of homework, but now there is only one problem to assess, not the entire problem set. Since there is less volume, greater attention can be devoted giving a detailed assessment of the problem that was submitted. To assess the problem, a spreadsheet is devised that has the solution broken down into its constituent parts with points assigned to each part. Typically the quizzes are worth ten points and there are a handful of sub-questions, each worth a point or two. At the top of the evaluation sheet is a cell for the student's name. In the body of the sheet are the correct answers to each of the sub-questions. A column on the right hand side of the sheet accumulates the points taken off (if errors appear) and the results are automatically calculated near the bottom of the sheet. Some space is reserved at the bottom of the sheet for comments to the student (if extra comments are warranted).

The completed evaluation spreadsheet is formatted so that the entire evaluation prints out exactly on one 8 ½ by 11 sheet. This formatted spreadsheet is then replicated down the worksheet as a template for as many times as there are students in the class. The papers are evaluated using a new template for each student. This assures consistency in the grading, provides detailed feedback, and minimizes the amount of writing that has to be done on the student's individual papers.

When all the papers have been graded, the spreadsheet is viewed with "print preview" to assure that all the evaluation sheets register properly (one sheet to a page) and then they are printed. Since they print out in the same order in which the papers were graded, it is easy to collate and staple the evaluations to the back of the each submitted quiz. Having the evaluation sheet second preserves the privacy of the grade as well.

Creating the quiz and the solution spreadsheet takes an hour or so. A class of 35 students can be fully processed in a little over an hour. The downside of this approach is that it does consume some class contact time – about fifteen minutes for each quiz. The advantage is that it greatly reduces the unpleasant task of grading papers – probably the greatest negative aspect of teaching

– while maintaining the benefit of having the students do assigned problems and receive detailed feedback on their solution procedures.

An extra feature we employ requires that we have assigned seating. Once the class has met and the add/drop period is over, we circulate a seating chart and the students take a permanent seat for the rest of the session. Each seat is assigned a sequential number and the student writes this number on the top of the quiz along with his or her name. The papers are sorted in ascending seat order which speeds recording the grades and returning them at the following class. It also makes grade entry easy by simply sorting the class roster by assigned seat number.

The effectiveness of this approach has not been rigorously tested using control and experimental groups. However, demonstrated proficiency and anecdotal evidence have supported the efficacy of this method. The section in which this method was applied was one of six sections of the Sophomore Management Science course taken by all business students in the school. A common final exam is given and this section (n = 35) scored an average of 83.09 (Standard Deviation 10.61) with an overall course mean of 80.02 (ten sections n = 282) which suggests at least a solid understanding of the material. Written evaluations supported the student's appreciation of the approach with comments such as:

> "…the quizzes make sure you can grasp all of the course material."
> " Thank you for not baby-sitting us by making us do a lot of busy-work homework."
> " Doing all the homework was important for getting a good grade in this course."

In the spirit of full disclosure, there was one comment relating to this procedure under "Ways to Improve the Course."

> "Can't think of any way to improve, maybe collect homework instead of quiz on it."

Attached is an example from the Quantitative Methods class. First is the assigned problem as it appeared in the end-of-chapter problems. Next is the problem as it appears in the quiz. Third is the grading spreadsheet that is duplicated for each student.

Original Problem:

The distribution system for the Herman Company consists of three plants, two warehouses, and four customers. Plant capacities and shipping costs per unit (in dollars) from each plant to each warehouse are as follows:

	Warehouse		
Plant	1	1	Capacity
1	4	7	450
2	8	5	600
3	5	6	380

Customer demand and shipping cost per unit (in dollars) from each warehouse to each customer are:

	Customer			
Warehouse	1	2	3	4
1	6	4	8	4
2	3	6	7	7
Demand	300	300	300	400

a. Develop a network of representation of this problem.

b. Formulate a linear programming model of the problem.

c. Solve the linear program to determine the optimal shipping plan.

Quiz #4   Your Name : _____   Your Seat No. _____

Problem 17 in Chapter 6 (Herman Company) gave a scenario minimizing shipping cost with three plants, two warehouses, and four customers. Plant capacities and shipping costs ($/unit) from each plant to each warehouse are:

Plant	Warehouse 1	Warehouse 2	Capacity
1	$4	$7	450
2	$8	$5	600
3	$5	$6	380

Customer demand and shipping costs ($/unit) from each warehouse to each customer are:

Warehouse	Customer 1	Customer 2	Customer 3	Customer 4
1	$6	$4	$8	$4
2	$3	$6	$7	$7
Demand	300	300	300	400

1. (3 pts) Develop a network diagram of this problem.

2. How many decision variables are there?   _____

3. How many constraints are there? (Ignore the non-negativity requirement.)   _____

4. (3 pts) Suppose Plant 2 cannot supply Warehouse 1. In what three ways can the model presented in the book be modified to accommodate this condition? (Be specific.)

    1_____

    2_____

    3_____

5. Suppose the situation changes so that shipments between the two warehouses can be made at $2 per unit and direct shipments can be made from Plant 3 to Customer 4 at $7 per unit. All the other conditions from above (including Q. 4) still hold. Develop a network representation of this situation.

6. Write the constraints for the warehouses that include the exchange between the two warehouses. (For this question you do NOT need to include decision variables with a coefficient of 0.)

Student		Seat number	0
Grading Criteria		√ =	got it
		x =	missed it
			Points
Question	**Answer**		**Deducted**
1 Network Diagram			3
	Be sure to use arrows, not just lines to show flow.		
2 Number of Variables?	14		1
3 Number of Constraints?	9		1
4 Ways to eliminate P2 to W1.	1. Remove the decision variable P_2W_1 from the problem.		1
	2. Add a constraint stating $P_2W_1 = 0$		1
	3. Set the cost for P_2W_1 very large (Big "M")		1
5 Modify problem.	Add 2 warehouse arrows.		0.5
	P_3C_4 arrow added and P_2W_1 arrow gone.		0.5
6 Write constraints for warehouse exchange.			
	$P_1W_1+P_3W_1+W_2W_1-W_1W_2-W_1C_1-W_1C_2-W_1C_3-W_1C_4 = 0$		0.5
	$P_1W_2+P_2W_2+P_3W_2+W_1W_2-W_2W_1-W_2C_1-W_2C_2-W_2C_3-W_2C_4 = 0$		0.5
	It's also OK to indicate these are ≥ 0.		
	It's also OK to have the first set of terms negative		
	and the second set positive and the whole thing = 0		
	And if you do it this way, it's OK for it to be ≤ 0		
		Quiz Grade	0
Comments:			

Europe versus Globalization of the XXI Century

Alexandre Melnik

ICN Business School, Nancy, France
Alexandre.Melnik@icn-groupe.fr

Confronting to new global challenges, Europe - birthplace of the capitalism in XVII century and of the human rights in XVIII century - is now a **slacking player of the globalization of the XXI century**, facing its identity crisis and in quest for its new hard / soft power, in the aftermath of

- fall of the Berlin Wall (11/9 1989)
- last U.E.'s enlargements (or reunification of Europe, via the come back to Europe of the former "**kidnapped democracies**" – Milan Kundera - after the "cold war"s geopolitical rupture?) (2004 and 2007)
- rejection by Netherlands and France of the European Constitutional Treaty (2005) – a **burnout for the initial concept of the European construction** after the WWII : United Nations of Europe
- adoption of the Lisbon Treaty (December 2009)
- U.E.'s bailout for Greek debt (May 2009)
- first revision of the Lisbon Treaty (October 2010)

Now, Europe's leaders are not able to answer two crucial questions concerning its own identity:

- what is Europe today ?
- what does Europe want de be tomorrow ?

- **Geographical Entity**; in other terms: a territory from Atlantic Ocean to the Ural Mountains in Russia, from North Oceans to the Mediterranean?

- **Community of people who share history, culture and humanistic values** based on common historical background (Greek gods, Roman Law, Judeo-Christian heritage, Renaissance, First and Second Industrial Revolutions, Enlightenment, clear cleavage between State and Religion, Human Rights, Civil Code, etc) ; in other terms – " West Christian Club" cultivating the nostalgia of its glorious past ?

- **Exclusive Free Market Zone** within a group of wealthy nations; in other terms – select group of wealthy nations?

- **Counterweight to US** and to others new stakeholders of the globalization (Chine, India, Brazil, etc) ; <u>in other terms</u> – a transnational organism boosted by negative motivation (do something against vs create something for)

- **Political and Military Union**; <u>in other terms</u>, "United States of Europe", as coherent global player in World Politics?

- **New, future oriented project of the Human Civilization;** <u>in others terms</u>, pan-European Renaissance in XXI century?

Rethinking of the basic European Concept is a key Geostrategic Challenge of the XXI century's Globalization.

Three key challenges of Europe in the globalization:

- ❖ institutions

- ❖ economy

- ❖ politics

1. <u>First Challenge</u> - **Institutional challenge includes two elements:**

- ▪ E.U.'s model of functioning and its decision centre of gravity (conforming to 1,000-page Lisbon Treaty, the E.U.'s multi-headed leadership :
  - ✓ President of the European Council
  - ✓ President of the European Commission
  - ✓ country - holder of rotating presidency
  - ✓ European "Foreign Minister"
  - ✓ President of the European Parliament)

"Who do I call when I want to call Europe"? – Henri Kissinger

The E.U. has important means and structures, but its institutional constellation is too much bureaucratic, lacks visibility and is not adapted to the new European geopolitical situation

- ▪ **New enlargements (where are the frontiers of Europe?)**

  Agenda of enlargement priorities – 3 groups of countries which apply for the E.U.'s integration

  - a) Balcanic countries of the former Yugoslavia
      (Croatia, Serbia)
  - b) two countries of the former URSS (Georgia and

Ukraine) – "policy of neighbouring"
c) Turkey

Facing identity crisis, the E.U. has now lack of visibility about the further enlargements: European public opinion is not ready to welcome the potential "newcomers".

2. Second Challenge – **economy challenge includes three elements**

The part of Europe in the Global GDP is declining:
25% in 1995 (15 UE members) vs 15% in 2010 (27 UE members)

- How to find in the global economy a "Modus Vivendi" between economic excellence and European social model based on "acquis sociaux" & Welfare State?
- Europe needs a shift of mentality (how to pass from "protect the average and smother the excellence" to the permanent "race to the top" and "drive to be smarter")
- How to obtain the economic convergence between West and East of the Europe reunified after the fall of the Berlin Wall ("Mittel Europa" as opportunity, not as burden)

3. Third Challenge – **political challenge**, perhaps the most difficult because now Europe either largely absent or absent as global political player – creation of "powerful Europe" ("Europe-puissance") **includes the following elements**

- Implementation of the European "hard power" (common military force) – *the demilitarization of Europe has gone from a blessing in the XX century to an impediment to achieving real security and lasting peace in the XXI century*
- Dichotomy "New – Old" Europe (different visions of European construction in the West and in the East of Europe)
- European political leadership (France and Germany are still the alone driving force of Europe?)
- Europe vs US (rivalry or alliance)
- Europe vs Russia (Russia is threat or partner? Russia is in Europe or out of Europe?)
- Europe vs Chine (what's the U.E.'s China strategy vs US well-articulated China approach : "engage economically, encourage democratically and criticize on human rights when appropriate")
- Europe toward climate change (after the failure of the Copenhagen summit in December 2009)

ALL THREE E.U. CHALLENGES ARE VERY HUGE, BUT POSSIBLE TO MEET.

Projection into future: "Old Continent – New Ideas" – reinvention of the European Dream

Europe needs **a strategic long-term vision** of how to deal with Globalization without denying its historical model.

Europe needs more than just the realization that the old ways of doing business and geopolitics no longer suffice. It requires a strategy of how to merge the challenges of a globalised and more dynamic world with the accomplishments in the social sphere ("acquis sociaux") that Europeans are rightly proud of. It requires a new geopolitical strategy, as it was the case in the XIX century with the Wien Treaty of 1815 and in the XX century – after the WWII.

In conclusion, despite all the current difficulties, I believe in Europe not only as a global economic and political player, but also as a new, very attractive, future oriented "**soft power**" model of the Human civilisation in the XXI century.

This project of the powerful Europe must be inspired by its common values

- individual human rights
- freedom
- State of Law
- separation of the State and the religion
- rejection of discrimination, particularly against women
- social protection
- solidarity between generations
- efficient public services

In this spirit, Europe needs, first of all, a real **shift of mentality** of its decision makers. Because even IT and R&D expended, it is not clear whether they would have the intended effect. The biggest psychological impulse to boost productivity and to "go global" is the survival instinct: the drive to be smarter and faster, "**race to the top, never to the down**". Nevertheless, now days plenty of barriers to competition hinder in Europe that instinct: its decisions makers seem to protect the average, but smother the excellent.

The "Old Continent" has to meet the new challenges of the XXI century and to reinvent its European Dream.

Conception to Commercialization with Immersion

Dr. Marie-Eve Faust, faustm@philau.edu

Dr. Susan Christoffersen, ChristoffersenS@philau.edu

Philadelphia University, School of Business Administration

ABSTRACT

Innovative product development is crucial in today's competitive economic environment. In order to create a learning environment where we foster the required skills, we brought an interdisciplinary product development class to Hong Kong. Literature shows that great innovations, from conception to commercialization, can result from immersion, placing people in a new cultural environment, to generate great new ideas. A "*regard neuf*" at ordinary things, observations on how people from another country live in their day-to-day life can generate a multitude of ideas for new products.

In addition to immersion in a foreign culture, multidisciplinary collaboration also contributes to innovative product development. Once an idea is identified, a good multidisciplinary team of designers, engineers and marketers is essential to the development and commercialization of the product. One needs to incorporate the voice of the customer in the design; this is known as being very powerful, therefore tools such as the Quality Function Deployment (QFD) or Kano's model enhance the process of new product development.

In this research we validate the importance of all of the above. A multidisciplinary team of sixteen undergraduate students native from various places in America were brought together in Hong Kong (a new environment for all of them) to see if they could come out with a

variety of new, innovative products that could either be commercialized in HK's market or America's market. The student products exceeded our expectations. The course included visits to vertically integrated companies (from ready-to-wear to high end products), exposing the students to best practices of managing innovation 'from concept to commercialization'. Our pilot test validates the importance of multidisciplinary team and moreover it confirms the importance of the immersion in a new environment as a way to innovate in this global world.

QUANTITATIVE PREDICTIVE CAPACITY OF HUMAN DEVELOPMENT INDEX IN WIRELESS TELEPHONY OPERATIONS.

Andrey Fendyur[a], Vas Taras[b],

[a]Haskayne School of Business, University of Calgary;

[b]Bryan School of Business and Economics, University of North Carolina at Greensboro.

[a]*Andrey Fendyur is PhD student in Operations Management at the University of Calgary, AB, Canada,* *afendyur@ucalgary.ca*; [b] *Vas Taras, PhD, is Assistant Professor in International Management at the University of North Carolina (Greensboro), NC, USA*

ABSTRACT.

The research explores effects of Human Capital on the wireless telephony operations. The novelty and contributions of the paper are: (1) HDI, as a proxy for Human Capital, can be a predictor for analyzing and forecasting wireless telephony penetration; (2) the predictive capacity of HDI on wireless telephony expansion is found to be more than that of pure monetary indicator (GDP per capita); (4) less developed fixed line network does not result in increased mobile use per capita. The findings can be used both in academia for studying wireless telephony operations and in industry for analysis and decision making.

INTRODUCTION.

Wireless telephony as business research topic and its role in business and communication.

Telephony in general and wireless one in particular, has received little attention in research as a social and economy phenomenon [7]. According to [4], there was no analysis done on aggregate impacts of that phenomenon in society and economy. Research community did not reward telephony to be a research topic as much as Internet [1], [4]. [1], for example, awarded "Mega-Innovation" status only to Internet, not to wireless telephony, nevertheless the mobile telephony was more widespread and its economy role as a means of communicating for business reasons was substantial [4]. [4] argued that the wireless telephony had been disseminated wider in more strata of society and economy, and spending for the mobile telephony was higher than that for the Internet provider services. That phenomenon was formulated by [9]: "The advent of inexpensive mass-produced mobile communications in particular, has avoided scholarly attention, perhaps because it seems pedestrian compared to the nebulous depths of cyberspace. Yet the cellular telephone, merely the first wave of an imminent invasion of portable digital communications tools to come, will undoubtedly lead to fundamental transformations..."

The penetration of wireless phones was so fast and wide, that they took their role not only in economy, but also in society: for example, the number of mobiles in use overpassed the number

of TV sets in 2001 [5]. That means, the mobile telephony industry became not only a service industry for economy use, but also a self-sufficient commercial industry – like automobiles, agriculture, computers etc. Thus, knowing predictors for mobile telephony deployment in various markets is of academic and industry's interest and practical importance – both for research, conceptual reasons and for practical business purposes.

From our empirical experience, industrial analysts typically use GDP per capita as a predictor of market potential. However, the mobile telephony expansion has happened worldwide, independent of cultural or monetary issues. [3] pointed out at a wide spread of mobiles in Italy, though the country is "technophobic" and "computers and other modern technologies have a difficult stand" there. [4] indicated that mobiles were well-accepted in Scandinavian countries, where consumers were inverted and did not communicate much. [7] highlighted that mobiles were wide-spread even in illiterate populations of poor countries – thus, involving them in both economy activity and telecom services consumption. In those countries, mobiles substituted use of computers in business in case of illiterate business operators. Also [7] and [4] indicated that a wider use of mobiles in Europe vs USA might have come from the fact of less involvement in driving, therefore, more time available for communicating while in train or other public transportation.

The study of predictors for mobile phones use is also important because some hypotheses (for example, [4]) indicated possible substitution of many uses in PCs and laptops while selection of options available through a mobile device increases (access to the Internet, use of business email through mobile phone devices etc.). That can open new horizons for the mobile telephony industry and make competition within it more severe, thus, asking for more reliable predictors for academic research and business forecasting. Geser [4] articulated that need as: "On the theoretical level, this situation calls for the development of highly elaborated analytical concepts and typologies suited for grasping the major differences in usage patterns, as well as the various symbolic meanings attributed to mobile phones, messages and users; on the methodological level, it implies the need for survey studies, as well as ethnographic approaches, for assessing such variables empirically in quantitative as well as qualitative ways". The current paper addresses both qualitative aspects of the search for mobile use predictors and quantitative indicators to be used in academic and business circles.

LITERATURE REVIEW.

Telephony as an industry and the use of telephony in business.

Telephony played a big role in shaping contemporary business operations. According to [4], fixed line telephony sustained organizational structure of businesses and consolidated operational environment. [7] noted: "Telephone is a key element in the building of corporate empires. Apart from easing the violation of laws and the realisation of exchanges without leaving traces [13], it permits the physical separation of the offices from the factories, allowing the managers to keep the control of the production. Therefore, the telephone plays a role in the urban concentration of financial and business activities. The telephone helped in the development of larger metropolitan systems with a more diversified and complex structure it is also a central element in the work

organisation and communication inside the skyscrapers, the symbols of corporate capitalism that arose at the beginning of the 20th century." [7].

Therefore, telephony plays an important role in maintaining operational unity of companies. Emergence of wireless (mobile) telephony helped to overcome the limitation of being immobilised for communicating long-distance for business reasons. [4] formulated that as "the significance of the mobile phone lies in empowering people to engage in communication, which is at the same time free from the constraints of physical proximity and spatial immobility."

Then, [4] indicated natural boom for spreading mobile telephony worldwide: the wireless telephony addressed both business and social needs. Therefore, the two main reasons for mobile telephony density in a particular society could be indicated as (1) business needs, which could be tracked through analyzing density of mobile networks vs GDP per capita, and (2) social needs, that could be analyzed through Human Development Index (HDI) across nations. An assumption behind the working hypotheses in the paper is that HDI constitutes a better predictor of mobile network density in a market due to its composite nature: both the monetary (business) component (GDP per capita) and social capital components (life expectancy, education level) are included in HDI.

The degree of economic development, which is related to the fixed line and wireless telephony penetration, is the function of the institutional development [1], [2], [5], [13], [8]. Its proxy could be "GDP/capita" [1]. From that, I could suggest to keep development level in monetary expression as a potentially moderating variable in my research: according to literature above, the variable "GDP/capita" may influence other variables related to economic development, including the industrial productivity growth rate. In line with the theory in current literature [1], [2], [5], [13], [8]., I will check for moderation and mediation effects of the "GDP/capita" on the mobile telephony penetration across nations.

Human Development Index background.

[1], [2], [5], [13], [8]. suggested that the economic development and technology use proliferation are subject to availability of skilled workforce and human development. The causation comes from presumption of better efficiency and productivity from skilled workforce versus less trained workforce. Nowadays, many industrial technologies require trained workers to operate facilities. That assumes the role of personnel training and education in increasing industrial technologies penetration of the market. A proxy for the human development could be the Human Development Index [10].

The Human Development Index [10] is a composite measurement, ranking countries by means of incorporating three elements: (1) life expectancy at birth; (2) knowledge, education, competence and intellectual development measured through aggregation of adult literacy rate of a particular country (2/3 weight), plus combined primary, secondary, and tertiary gross enrolment ratio (1/3 weight); (3) standard of living, expressed through the natural logarithm of GDP (USD) per capita adjusted to purchasing power parity.

The index is calculated and published by the United Nations Development in its Human Development Report since 1990. The current issue published in October 2009, reflects data

available as of 2007. Values below 0.5 indicate "low development", while 0.8 and more – "high development".

Though critics argue on HDI's simplicity, limitations, etc., the index is a well-known and widely accepted indicator [10].

Innovative Role of Wireless Telephony.

According to [4], cell telephony possesses innovative potential in the evolutionary prospective. That is framed by two factors: 1. Physical proximity needed for interpersonal communication, and 2. Stable dwelling places required for development of complex ways to exchange information and solve challenges. Also [4] drew several implications from those factors: the first one implied variability among different groups (geographically, industry-wise, etc.) since the different availability of physical proximity options enables various types of developments (both in society and in economy). From [4], the second factor assumed that more complex and developed forms of cooperation and organizational structure (including industrial entities) could be developed if participants interacted closely for long periods of time. In other words, good communications could trigger social and industrial development. That was the case of ancient times, when agriculture enabled people to reside in settlements, and close continuous communications caused emergence of civilizations, like ancient Egypt, Mesopotamia, India [8], [9]. In modern world, the implication of the two principles listed above implied that better communications were features of more balanced human-economy development. The dimension of balanced human-economy development is different from pure monetary measurement of development: the balanced human-economy development is incorporated in the Human Development Index (HDI), which consists of three components. Contrary to HDI, the monetary measurement of development is specified as GDP per capita.

WORKING HYPOTHESES.

From the innovative, business and communication role of the wireless telephony highlighted in the literature review section above, we developed the following working hypotheses:

Hypothesis 1: *Human Development Index might be a better predictor of mobile use density across market than a pure monetary income per capita predictor.*

An influential role of the income per capita for the human development index (which is reflected in the natural logarithm of GDP per capita PPP included in HDI formula) triggered our interest in testing existence of possible mediation/moderation influences of "GDP per capita PPP" on the use of HDI as a predictor for "Mobiles per capita" criterion variable. That caused to develop Working Hypothesis 2:

Hypothesis 2: *Monetary indicators, such as "GDP per capita PPP", can mediate/moderate the effect of Human Development Index on mobile use density across markets.*

On top of that, literature on the topic suggested that the widespread of the wireless telephony use could be attributed to the low density of fixed line network [4]. The logic behind that assumption was based on the substantial investment need for development of fixed line networks in so-called industrializing nations. "Industrializing nations" is a term used to refer to rapidly industrializing countries, which recently started their growth from beginning levels of infrastructure and economy indicators [4]. To skip that investment need and accelerate development, the industrializing nations often choose to build modern wireless telephony networks instead of traditional fixed lines [4]. Therefore, Hypothesis 3 aimed at testing that assumption:

Hypothesis 3*: there might be negative correlation between the fixed line per capita and mobiles per capita indicators: the lower the fixed line density, the higher the wireless telephony density.*

DATA COLLECTION, METHODS AND RESULTS.

Data used in the research came from reputable sources [10], [11], [12]. By default, data are latest available as of February, 2010. The number of mobile phones per country, and population parameters [14] were used to calculate the criterion "Mobiles per capita" variable. HDI values were obtained from the UNDP'2009 report. "GDP per capita PPP USD" variable was obtained from [14], too.

Top 122 data points were considered, to be representative of the general countries population: small countries with special conditions for wireless telephony operations (such as small island countries, were wireless penetration is limited due to geographical reasons etc, were not taken into the analysis, see the descriptive table below).

TABLE 1. Principle descriptive characteristics of variables.

Variables	Mean	Standard Deviation	Number of cases
Mobiles/ capita	0.8303	0.3988	122
HDI	0.7617	0.1688	119
GDP per capita	14852	14726	122
Fixed line/capita	0.1972	0.1814	162

The software used was SPSS version 17. The methodology employed for detecting predictive capacity of indicators included descriptive analysis of data, parametric correlation study, non-parametric correlation study, simultaneous and stepwise regressions. The major research methodology for moderation check was multiple regression analysis, and Sobel test – for mediation assessment. There were two main reasons for that selection: first, the multiple regressions method is the first choice method to identify influence of predictor variable on the criterion variable given they all are continuous numerical, especially if interaction effects suspected [15], [16]; second, the multiple regressions method is an effective and objective quantitative instrument in itself [15], [16], [19]. The Sobel test is also a must do method for mediation assessment [17], [18]. Thus, the conventional regression relationship analysis was performed coupled with analysis of probable mediation or moderation effects. The traditional

regression analysis is complemented with the multiple regression analysis to illustrate more precise and accurate nature of findings obtained through multiple regression analysis: the use of multiple regression analysis allows identify moderation effects. The Sobel test serves the same role, showing mediation effects presence. Otherwise, wrong inferences can be obtained and wrong conclusions drawn about which factors (if) influenced the industrial production growth rate in 2009.

Parametric correlations (see Table 2 below) produced evidence of significant strong positive correlation (Pearson's correlation coefficient) between the number of mobiles per capita and GDP per capita (0.697 significant at $p<0.01$ level); between the number of mobiles per capita and HDI (0.814 significant at $p<0.01$ level). In line with the working hypothesis, the correlation between the number of mobiles per capita and HDI was stronger than between the number of mobiles per capita and GDP per capita (0.814 versus 0.697, both significant at $p<0.01$ level).

TABLE 2. Parametric (Pearson's r) bivariate correlations, selected variables.

Variables	Mobiles/ capita	HDI	GDP/capita
Mobiles/capita	1	.814*	.697*
HDI	.814*	1	.778*
GDP/capita	.697*	.778*	1

* Correlation is significant at the 0.01 level (2-tailed).

The number of valid cases used (119) in the analysis provides a sustainable prevention of deviations in findings against possible problems with a violation of the normal distribution assumption essential for regression [16]: the central limit theorem might be valid at that number of data points. However, to check out the reliability of the findings from parametric (based on normality of distribution assumption) correlation study, we performed a non-parametric correlation study (non-parametric techniques are rank-based and do not assume normality of distribution). The non-parametric correlation study results are presented below (Table 3). Two non-parametric correlation coefficients were used to provide more verification to parametric correlation study results.

TABLE 3. Non-parametric bivariate correlations, selected variables.

Non-parametric correlation coefficient	Variables	Mob per capita	HDI	GDP per capita
Kendall's tau_b	Mob per capita	1.000	.613*	.655*
	HDI	.613*	1.000	.843*
	GDP per capita	.655*	.843*	1.000
Spearman's rho	Mob per capita	1.000	.815*	.845*
	HDI	.815*	1.000	.964*
	GDP per capita	.845*	.964*	1.000

* Correlation is significant at the 0.01 level (2-tailed).

Kendall's tau b and Spearman's rho non-parametric correlation coefficients confirm that HDI is a strong and significant predictor of the number of mobiles per capita. In the current non-parametric study, the value of the correlation between GDP/capita and the number of mobiles/capita is more (0.655 vs 0.613 for Kendall's tau b and 0.845 vs 0.815 for Spearman's rho, all significant at $p<0.01$ level).

That is different from the parametric study results described above. That could be attributed to the lower sensitivity of non-parametric tests versus parametric ones [16]. Therefore, the non-parametric correlations confirm reliability of the finding on HDI's significance as a predictor for the number of mobiles per capita. Taking into account a higher sensitivity of parametric correlation coefficients versus non-parametric ones, we believe it is more appropriate to use Pearson's correlation coefficient to rank strength of effect of HDI/GDP per capita on the number of mobiles per capita. The second reason for that decision comes from the theory's support (not only monetary (GDP/capita) but also human capital reasons lead to a certain number of mobiles per capita, therefore, a combinatory (monetary human capital) indicator (HDI) is a better predictor that only monetary (GDP/capita) one). Pearson's correlation coefficient for HDI (0.814) is higher than that for GDP/capita (0.697). That supports the theory-based working hypothesis 1.

The bivariate correlation study (Pearson's r) between the "Mobiles/capita" and "Fixed lines/capita" indicators produced results contrary to working hypothesis 3: the two indicators correlate positively, the correlation coefficient is high (0.721) and significant at $p<0.01$ level. That rejected working hypothesis 3 and enabled to conclude that the deficit of fixed lines in industrializing nations was not the cause for proliferation of the wireless telephony. In fact, the finding pointed out at the role of general development level of infrastructure and GDP/capita in increasing the density of fixed line and mobile telephony coverage: the higher the GDP/capita, the higher is the fixed line and mobile telephony coverage density. That finding is both theoretically and practically interesting. It can be used both in scholar research on telecom/development and in industry for forecasting telecom operations/investments on particular markets.

To test the predictive capacity of the HDI and GDP/capita, we performed simultaneous and stepwise regression analysis. The purpose of the simultaneous regression analysis was to identify predictive capacity of the model with both HDI and GDP/capita. The stepwise regression analysis was performed to identify statistically the most effective predictor of those two. Results are presented below (Table 4).

TABLE 4. Simultaneous and stepwise regression models.

Parameter/Model	Simultaneous	Stepwise
Predictors included in model	HDI, GDP/capita	HDI
R^2	0.672	0.663
Adjusted R^2	0.667	0.660
F change	118.963	229.955
Significance of F change	<0.01	<0.01
Standardized coefficients	HDI: 0.693	HDI: 0.814
	GDP/capita: 0.155	

t-values	HDI: 8.191	15.164
	GDP/capita: 1.830	
p	HDI: <0.01	<0.01
	GDP/capita: 0.07	
df	2, 116	1, 117

Although GDP/capita is highly correlated with the criterion variable of Mobiles/capita (Pearson's r = 0.697, significant at p<0.01 level, Table 2), GDP/capita becomes insignificant as a predictor if HDI is present in the simultaneous model. Along with a higher correlation coefficient of HDI with Mobiles/capita (Pearson's r = 0.814, significant at p<0.01 level, Table 2), that finding supports hypothesis 1: HDI, as a proxy for Human Capital, is a better predictor of mobiles used per capita than a pure monetary predictor "GDP/capita".

The stepwise regression analysis was used to further test that conclusion. The stepwise regression is considered more parsimonious and keeps only important predictor(s) in the model [16].

Results of the stepwise regression analysis (Table 4) support that conclusion, too: GDP/capita is not included as a predictor in the model, while HDI is; both models explain similar amount of variance in the "Mobiles/capita" criterion variable (67% and 66%, see above in Table 4) but the stepwise model consists of only 1 predictor (HDI, more parsimonious); HDI has a higher beta value if GDP is out of model (beta$_{HDI}$=0.693 in the simultaneous model and 0.814 in the stepwise one, all significant at p<0.01 level). F change is also more definite if only HDI is included in the model (230 vs 119, p<0.01). That conclusion also fits in line with hypothesis 1 that HDI, as a proxy for Human Capital, is a better predictor of mobiles used per capita than a pure monetary predictor "GDP/capita".

Then, we did moderation research with the multiple regression method. The suggested moderator was "GDP per capita". Results and brief interpretation will be provided below (Table 5 & 6).

TABLE 5. Models summary.

Model	Predictors (constant included)	R^2	F change	df
1	HDI	0.666	300.838*	1, 151
2	HDI, GDP/capita	0.706	20.691*	1, 150
2	HDI, GDP/capita, Interaction Term "HDI" * "GDP/capita"	0.737	20.961*	1, 149

*p<0.01

TABLE 6. Selected parameters for predictors.

Model	Predictor	Beta	t-value
1	HDI	0.816	17.345*
2	HDI	0.619	9.933*
	GDP/capita	0.282	4.549*
3	HDI	0.599	10.261*
	GDP/capita	4.085	4.905*
	Interaction Term "HDI" * "GDP/capita"	-3.794	-4.578*

*p<0.01

"Mobiles/capita" criterion variable study demonstrated support of the moderation effect of "GDP/capita" variable on HDI. The model with the interaction term (model 3 in the Table 5) demonstrates a significant R^2 increase (to 73.7% from 70.6% of the variance explained) and F change (df1, 151) = 20.961, significant at p<0.01 level. The interaction term is significant at p<0.01 level (t=-4.578, Table 6). Thus, the moderation effect of "GDP/capita" on HDI toward the wireless telephony use density gets support from the multiple regression study.

Then, we checked for possible mediation effects of "GDP/capita" on HDI in influencing the wireless telephony use density. The most common method used in literature for that purpose is the Sobel test [17], [18]. There are several variations of the test due to inclusion/exclusion of the error term. According to literature mentioned, they produce same results. We show outcomes from all variations of the test to compare findings. We used the Sobel test and the online software for calculation provided by Preacher & Leonardelli (http://www.people.ku.edu/~preacher/sobel/ sobel.htm). The summary of results is shown below (Table 7).

TABLE 7. The Sobel test findings.

Input Parameter	Value		Test Version	Test Statistic	SE	p
a	66145.11		Sobel test	12.03419054	0.104707	<0.01
b	0.00001905		Aroian test	12.03419054	0.104707	<0.01
s_a	5496.432					
s_b	0.000		Goodman test	12.03419054	0.104707	<0.01

All variations of the Sobel test showed the same inferences: there is a mediation effect of "GDP/capita" on "HDI" in determining the wireless telephony use density (p<0.01).

The results presented above prove the initial finding of higher importance of HDI in predicting degree of mobile telephony penetration in a given market (through the number of mobiles per capita as the dependent variable). The complex influence of GDP/capita (moderation and mediation) on the predictive capacity of HDI is shown. That novel predictor could be used for both academic research, business planning and forecasting in academia, public policy and the wireless telephony industry.

DISCUSSION AND CONCLUSIONS.

The research addressed the question of which factor among selected most influenced the wireless telephony penetration across markets. Also, moderating and mediating interactions of those factors were investigated.

The factors were selected on the grounds of economic and operations theory. Inclusion of factors into the analysis was justified by theory and supported through references to respective directions of research and scholars working in those areas. Predictors selected for analysis were

"GDP/capita" to incorporate possible influence of the institutional development stage and monetary issues; "HDI" to incorporate the role of human capital in mobile industry penetration across markets. On the basis of relevant research into related literature, several working hypotheses were drawn.

Hypothesis 1 propositioned that the human development level (through its proxy HDI) plays role in the wireless telephony penetration across markets. That hypothesis received support from findings: HDI was a significant and the most substantive predictor for the "Mobiles/capita" criterion variable. Hypothesis 2 assumed that HDI be moderated and mediated by GDP/capita. The assumption was supported by the multiple regression moderation and the Sobel test mediation analysis. That is in line with existing theory: human development is subject to investments in human development, also GDP is a component of the HDI.

Hypothesis 3 implied that the lack of the fixed line penetration stimulates the wireless telephony penetration across markets. The findings from our research showed that hypothesis 3 had no support from fixed line and wireless telephony density data obtained from 150 markets worldwide. The explanation of the phenomenon may be in the area of institutional (infrastructure) development, whose proxy could be GDP/capita. Since both the fixed line and the wireless telephony require substantial investments in network set-up, the higher degree of development implies more opportunities and more resources available for any telephony network set-up. Thus, markets with higher GDP/capita command more resources for any telephony development, and that results in high positive correlation between the fixed line and wireless telephony penetration.

LIMITATIONS AND FUTURE RESEARCH.

The complexity of the wireless telephony penetration phenomenon gives grounds for applying SEM method to researching into it. SEM use was out of scope the current study. In future research, SEM will be a valuable method for studying the wireless telephony penetration and its relations with other phenomena. That method enables to produce complex constructs for testing sophisticated hypotheses. SEM is in focus for future research. For doing SEM, the further theory development is required to acquire more variables which will enable to develop a structural model. This research did not cover all potentially eligible variables since that was not the research question attempted.

Fundamental theoretical research into the wireless telephony penetration rate will be a promising pursuit as well. The fact of both moderation and mediation influence from GDP/capita toward HDI raises the question: what interaction was more powerful in practice and why? How to separate the unique incremental variation attributed to each of those interactions? Are there transformations which can clear data from one of those interactions?

REFERENCES.

[1] Castells, Manuel (1998): The Information Age: Economy, Society, and Culture. Oxford: Blackwell.

[2] Fortunati, Leopoldina (2000): The Mobile Phone: New Social Categories and Relations. University of Trieste

[3] Fortunati, Leopoldina (2002): Italy: stereotypes, true and false. (In: Katz, James E. / Aakhus, Mark A. (eds.) Perpetual Contact. Mobile Communication, Private Talk, Public Performance. Cambridge University Press, Cambridge, pp. 42-62).

[4] Geser, Hans (2004) Towards a Sociological Theory of the Mobile Phone. University of Zurich, Release 3.0, May 2004. http://socio.ch/mobile/t_geser1.pdf as of 26.12.2009

[5] Katz, James E. / Aakhus, Mark A. (2002): Introduction: Framing the issues. (In: Katz, James E. /Aakhus Mark A. (eds.): Perpetual Contact. Mobile Communication, Private Talk, Public Performance. Cambride University Press Cambridge, pp. 1-14).

[6] Kaasa, A. (2009). Effects of different dimensions of social capital on innovative activity: Evidence from Europe at the regional level. Technovation, 29, 218-233.

[7] Lasen, Amparo (2002): The Social Shaping of Fixed and Mobile Networks: A Historical Comparison, DWRC, University of Surrey.

[8] Lenski, Gerhard / Nolan, Patrick / Lenski Jean (1995): Human Societies. An Introduction to Macro Sociology. McGraw-Hill, New York, (7th ed.).

[9] Townsend, A M. (2000): "Life in the Real-Time City: Mobile Telephones and Urban Metabolism". Journal of Urban Technology, (7)2: pp. 85-104.

[10] UN Development Program Report, 2009: http://hdr.undp.org/en/media/HDR_2009_EN_Complete.pdf Human Development Report 2009

[11] UN documents: http://hdr.undp.org/en/statistics/indices/hdi/question,68,en.html as of November, 11th, 2009

[12] World Telecommunication Development Report (2002): "Reinventing Telecoms" & Trends in Telecommunication Reform 2002 "Effective Regulation". (ITU 15. 3. http://www.itu.int/newsroom/wtdc2002/backgrounder.html

[13] Aronson, SH, 1977. The Social Impact of the Telephone. MIT Press Cambridge, MA.

[14] CIA Factbook. https://www.cia.gov/library/publications/the-world-factbook/ Accessed on June 17th, 2010.

[15] Aiken, L. S., & West, S. G. (1991). Multiple regression: Testing and interpreting interactions. Newbury Park, CA: Sage Publications.

[16] Hair J., 2009. Multivariate Data Analysis, 7th Ed. Prentice Hall, 2009.

[17] Preacher & Hayes, 2004. SPSS and SAS procedures for estimating indirect effects in simple mediation models. Behavior Research Methods, Instruments, and Computers 36 (4), 717-731.

[18] Preacher & Leonardelli, 2006. Calculation for Sobel Test: Interactive calculation tool for mediation tests. http://www.people.ku.edu/~preacher/sobel/sobel.htm (accessed November 5 2009).

[19] Cohen & Cohen, 1983. Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences. Erlbaum, Hillsdale, NJ.

COLLABORATIVE TECHNOLOGY AND TEAM-BASED PROBLEM SOLVING

Hayward Andres, Department of Management, North Carolina A&T State University,
1601 East Market Street, Greensboro, NC 27411, hpandres@ncat.edu

ABSTRACT

The purpose of this study was to use a qualitative behavioral observation approach to examine the effects of collaboration mode (collocated face-to-face versus non-collocated technology-mediated) on team-based problem solving behaviors – team learning, team reflexivity and team mental model development. Media synchronicity, communication grounding, and the team cognition theoretical frameworks are used to explain the effects of technology-mediated collaboration on intra-team communication and interactions while engaged in a problem solving task. Multivariate analysis of variance results revealed that relative to face-to-face collaboration, technology-mediated collaboration, through diminished synchronicity effects on grounding, experienced greater difficulty in effectively engaging in behaviors related to team learning, team reflexivity, and shared mental model development. Further, direct observation reveals that team-based problem solving can be characterized as a multidimensional process comprised of team learning, reflection, and shared understanding. The study also demonstrates the viability and utility of direct observation for studying team cognition process behaviors.

INTRODUCTION

Teams are an important component in the completion of organizational goals and objectives and the formation and use of teams (e.g., non-collocated geographically dispersed) has changed through the use of information and communication technologies (ICTs) (e.g., videoconferencing systems with shared whiteboards; enterprise-wide collaborative team work systems). Consequently more research on technology-mediated collaboration should consider taking a process analysis (e.g., direct observation) approach that examines dynamic and emergent behaviors that occur during team-based problem solving.

LITERATURE REVIEW

Dennis et al. [6] identified five capabilities (symbol sets, transmission velocity, parallelism, rehearsability, and reprocessability) of communication media (e.g., face-to-face, videoconferencing, phone, or email) that influence communication synchronicity and thus the successful performance of information exchange and shared interpretation. Communication requires creation and transmission of gestures, verbal and/or text-based content (i.e. symbols). The greater the media's ability to facilitate parallel transmission within a group and rehearsal of intended content, the more effective and efficient information is exchanged. Good communication is essential for *grounding* which is the process of making sure one's utterances are understood in communication with others and is the basis on which common ground or shared understanding is facilitated [2]. According to Clark and Brennan [2], real spoken conversation is often characterized by incomplete sentences, overlapping turns, pauses, and unintelligible utterances. The 'principle of least collaborative effort' asserts that dialogue is a

coordinated activity between the sender and recipient and participants in communication will try to minimize their collaborative effort needed to achieve mutual understanding [3]. Consequently, poor communication results in the need for relatively more expended collaborative effort aimed at acquiring and ensuring an understanding of communicated content.

Slavin [11] and later Edmondson [7] conceptualized learning as an iterative cycle of information exchange and reflection. In other words, team learning involves the exchange of facts and concepts, experimenting with ideas, joint reflection on them, and the collective restructuring and fine tuning of them. These team learning behaviors have also been referred to as team cognition [4]. Cognition in both individuals and teams has also been defined as the acquisition and sense-making of information to construct a mental model [8]. During team cognition, shared information is organized into mutually held mental models comprised of coherent chunks of causally-related facts or knowledge structures about the task and team and is used to guide behavior and decision-making. Task-related knowledge refers to knowledge of task procedures, strategies, contingencies, and constraints. Team-related knowledge refers to awareness of the knowledge, skills, abilities, attitudes and behavioral tendencies of team members. Finally, in order to *monitor* and *assess* conformance to a team-based problem solving task, a team engages in meta-cognition (i.e. reflective behaviors) to evaluate the correctness and viability of the task execution process and to revise task strategy if warranted [7], [10]. There are three central elements to the concept of reflexivity—reflection, planning, and adaptation.

RESEARCH MODEL AND HYPOTHESES

The research model is depicted in Figure 1 below. The model draws on the media synchronicity, communication grounding, and team cognition theoretical frameworks to explain the effects of technology-mediated collaboration on team cognition process behaviors. The model suggests that during problem solving collaboration mode will impact mental functioning and information processing behaviors (i.e. cognition) at the team level.
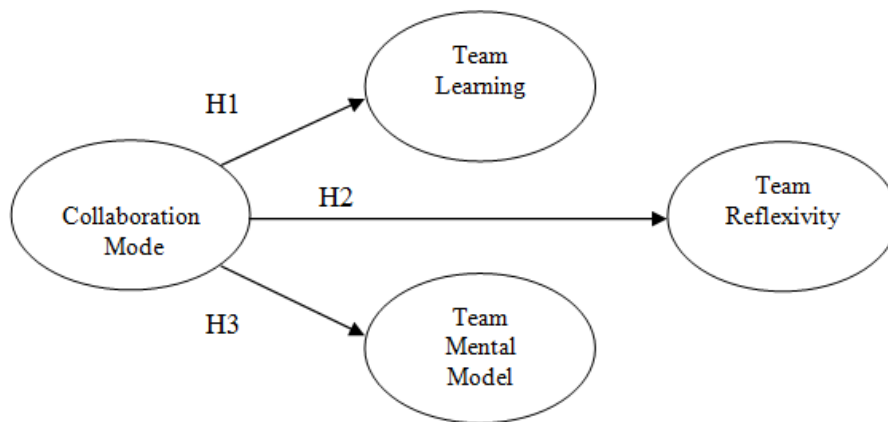
Figure 1. *Research Model*

Collaboration Mode and Team Cognition

According to MST, a media's ability to facilitate optimal communication performance (i.e. its level of synchronicity) is essential for team-wide information exchanges and interpretation necessary for learning at the team level and the development of shared meaning. In other words, collaboration mode (i.e. extent of physical proximity and communication media) must afford minimal collaborative effort cost to support information exchange and shared interpretation [2]. Consequently, there must be a fit between media capabilities and the breadth and depth of information processing needed to address the task so that collaborative effort is minimized. The specific mechanisms that enable the development of shared understanding include the ability to represent information using a rich symbol set, provide immediate feedback, minimize team member disengagement, and maintain a shared focus and interpretation of the task context. It is also argued that the level of synchronicity associated with a collaboration mode will impact a team's ability to effectively engage in team reflexivity behaviors. Team reflexivity requires the ability to create and maintain a shared focus on task solution content, task status, and an appropriate future task state. For example, Miranda and Saunders [9] noted that communication media not only affects the 'objective' quality of information exchanged (e.g., clarity, completeness) but also dictated the ability to maintain common ground and minimize discussion confusion both of which are essential to the team reflexivity process.

In summary, MST and grounding in communication theory suggest that the cost of increased interaction and increased cost of grounding (e.g., rate of clarifications, rate of acknowledgements, redundant exchanges) is likely to be more prevalent in non-collocated technology-mediated collaboration as compared to collocated face-to-face interaction. Moreover, collaboration mode will differ in the extent to which teams are able to 1) construct content the recipient is likely to clearly understand, 2) recognize that the recipient does or does not understand, 3) repair misunderstandings via feedback, 4) infer team member intention or needs, and 5) keep track of what has been discussed so far. Thus the following hypotheses regarding team cognition are proposed.

HYPOTHESIS 1. *Groups collaborating in the face-to-face settings will exhibit more effective team learning behaviors than in the technology-mediated settings.*

HYPOTHESIS 2. *Groups collaborating in the face-to-face settings will exhibit more effective team reflexivity behaviors than in the technology-mediated settings.*

HYPOTHESIS 3. *Groups collaborating in the face-to-face settings will develop a more accurate shared mental model of the task requirements and task status than the technology-mediated groups.*

RESEARCH METHODOLOGY

To test the research model and hypotheses, a laboratory experiment was conducted to examine the effects of two different modes of team-based problem solving – face-to-face and non-collocated technology-mediated. The participants were forty-eight Management Information Systems undergraduate students familiar with the Systems Development Life Cycle approach to software design and knowledge of structured programming. The teams, comprised of four

members, were required to enhance the functionality of a hypothetical university information system. The experimental task required each team to construct software design documentation that included (1) a hierarchy chart, (2) a list of function prototypes, and (3) pseudocode for each function identified as part of a solution to the problem.

Measures

Given that the final ratings were averaged among the raters, the $a_{wg(j)}$ interrater agreement index was used to assess inter-rater reliability [1]. In providing their observation ratings of team learning, shared mental model and task reflexivity, three trained observers used a rating scale that ranged from 1 (very low) to 7 (very high). The inter-rater agreement index for all scale ratings ranged from $a_{wg(j)} = 0.96$ to $a_{wg(j)} = 0.98$ indicating very good inter-rater agreement [1]. The team learning rating scale developed for this study was comprised of five items that reflected the degree of 1) team-wide information exchange, 2) experimentation with and evaluation of alternatives, 3) usefulness of ideas proposed, and 4) confirmed consensus on proposed ideas [11], [7]. The team mental model rating scale was comprised of four items that assessed behaviors that reflected the degree of 1) consensus on solution correctness, 2) confusion of task requirements and status, and 3) extent of needed explanations and clarifications [4]. Team reflexivity was assessed using items that provided observer ratings of the extent that team members 1) reflected on task procedure correctness, 2) reflected on solution content correctness, and 3) made an effort to create and maintain a positive climate. These behaviors are consistent with aspects of team reflexivity [5].

Data Analysis and Results

Multiple analysis of variance (MANOVA) was chosen to analyze the data in order to test whether there are statistically significant mean differences in collaboration mode on the linear combination of team learning, team reflexivity and team mental model. In addition, MANOVA offers the opportunity to find significant results where separate analysis of variance (ANOVA) test might not display significant results. The control variable, 'ability' did not exhibit significant effects and was therefore excluded from the analysis. Assumptions of data reliability required by MANOVA were assessed for compliance prior to data analysis [12]. Table 1 below shows the means, standard deviations, and correlations among the dependent variables used in this study.

Table 1. *Means, standard deviations, and correlations*

	Face-to-face		Technology-Mediated		ρ		
Dependent variable	Mean	Std. dev.	Mean	Std. dev.	1.	2.	3.
1. Team Learning	123.00	7.90	82.33	7.90	1		
2. Team Reflexivity	119.50	5.57	101.66	5.57	.79**	1	
3. Team Mental Model	79.00	5.15	54.00	5.15	.65*	.65*	1

** $p < .01$ * $p < .05$ (2-tailed).

Examination of Table 2 below reveals that a multivariate collaboration mode main affect was statistically significant (Wilk's Lambda = .309, $F(3, 8) = 5.970$, $p = .019$, partial $\eta^2 = .691$)

indicating that that collaboration mode exhibited a consistent simultaneous differential impact on team learning, team reflexivity, and mental model.

Table 2. *Mutivariate tests.*

Treatment Effect		Value	F	Hypothesis df	Error df	Sig.	Partial η2
Collaboration Mode	Pillai's Trace	.691	5.970	3	8	.019	.691
	Wilks' Lambda	.309	5.970	3	8	.019	.691
	Hotelling's Trace	2.239	5.970	3	8	.019	.691
	Roy's Largest Root	2.239	5.970	3	8	.019	.691

Univariate test results from the series of ANOVA tests appear in Table 3 below. The ANOVA of team learning was significant, $F(1,10) = 13.24$, $p = .005$, $\eta^2 = .57$, indicating that collaboration mode did contribute to mean differences in team learning indicating support for hypothesis 1. Face-to-face settings had a team learning mean score of 123.00 while the technology-mediated teams had a team learning mean score of 82.33 (see Table 1 above). The impact of collaboration mode on team reflexivity was also significant - $F(1,10) = 5.12$, $p = .047$, $\eta^2 = .34$ thus providing support for hypothesis 2. Team reflexivity was higher for face-to-face settings (mean score of 119.50) than in technology-mediated settings (mean score of 101.66). In support of hypothesis 3, collaboration mode has a significant impact on team mental model development ($F(1,10) = 11.76$, $p = .006$, $\eta^2 = .54$). The partial eta^2 (η^2) values of .57 for team learning, .34 for team reflexivity and .54 for team mental model indicate relatively strong effects sizes thereby indicating a strong relationship between the outcome variables and collaboration mode [12].

Table 3. *Univariate tests.*

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial η2
Collaboration Mode	Team Learning	4961.33	1, 10	4961.33	13.24	.005	.57
	Team Reflexivity	954.08	1, 10	954.08	5.12	.047	.34
	Team Mental Model	1875.00	1, 10	1875.00	11.76	.006	.54

DISCUSSION/CONCLUSION

The findings of this study suggest that collaboration mode impact on team functioning goes beyond facilitating communication and structuring social interactions to also affect cognitive processes (i.e. idea interpretation and evaluation, extent of shared understanding) and meta-cognitive processes (i.e. reflection of task strategy, solution content, and anticipation of appropriate future task state). The significant multivariate main effect supports the conception of a multidimensional aspect of team cognition – team learning, team reflexivity and shared mental model development. The three dimension are distinct in nature but yet conceptually related.

REFERENCES

[1] Brown, R.D. and Hauenstein, N.M.A. (2005), "Interrater agreement reconsidered: An alternative to the r_{wg} indices", *Organizational Research Methods,* 8(2), pp. 165-184.

[2] Clark, H. and Brennan, S. (1991), "Grounding in communication" in Resnick, L.B. and Levine, M. (Eds.), *Perspectives on socially shared cognition,* American Psychological Association, Washington, DC, pp. 127-149.

[3] Clark, H. and Krych, M.A. (2004), "Speaking while monitoring addressees for understanding", *Journal of Memory & Language,* 50(1), pp. 62-81.

[4] Cooke, N.J., Kiekel, P.A., Salas, E., Stout, R., Bowers, C. and Cannon-Bowers, J. (2003), "Measuring team knowledge: A window to the cognitive underpinnings of team performance", *Group Dynamics,* 7(3), pp. 179–199.

[5] De Dreu, C.K.W. (2007), "Cooperative outcome interdependence, task reflexivity and team effectiveness: A motivated information processing approach", *Journal of Applied Psychology,* 92(3), pp. 628-638.

[6] Dennis, A., Fuller, R.M. and Valacich, J.S. (2008), "Media, tasks and communication processes: A theory of media synchronicity", *MIS Quarterly.* 32(3), pp. 575-600.

[7] Edmondson, A. (1999), "Psychological safety and learning behavior in work teams", *Administrative Science Quarterly*, 44(2), pp. 350–383.

[8] Klimoski, R.J. and Mohamed, S. (1994), "Team mental model: Construct or methaphor", *Journal of Management*, 20(2), pp. 403-437.

[9] Miranda, S.M. and Saunders, C.S. (2003), "The social construction of meaning: An alternative perspective on information sharing", *Information Systems Research,* 14(1), pp. 87-107.

[10] Schippers, M.C., den Hartog, D.N. and Koopman, P.L. (2007), "Reflexivity in teams: A measure and correlates", *Applied Psychology*, 56(2), pp. 189-211.

[11] Slavin, R.E. (1996), "Research on cooperative learning and achievement: What we know, what we need to know", *Contemporary Educational Psychology*, 21(1), pp. 43-69.

[12] Tabachnick, B.G. and Fidell, L.S. (2000), *Using Multivariate Statistics,* (4th ed.), Allyn and Bacon, Needham Heights, MA.

KNOWLEDGE PERSPECTIVES IN PROJECTS:
UNDERSTANDING THE ROLE OF TIME

Dezhi Wu
Southern Utah University
351 University Boulevard, Cedar City, UT 84720
Email: wu@suu.edu Phone: 435-865 8444

Katia Passerini
New Jersey Institute of Technology
University Heights, Newark, NJ 07102
Email: pkatia@njit.edu Phone: 973-642 7328

ABSTRACT

The purpose of this research is to investigate individual perceptions of time and time management strategies that knowledge workers utilize to achieve their productivity in the execution of their daily tasks, projects and routines. Projects have specific time durations from the beginning to the end, which often need to be broken down into smaller temporal elements (e.g., milestones) and require learning, knowledge creation and capture throughout different project phases. We aim to observe how personal time management strategies promote knowledge management in projects. Adopting the knowledge-based perspective on time management enables understanding how often individuals capture, store, transfer and apply knowledge throughout the workday, and thus may help in creating operational efficiencies in organizations.

Keywords: knowledge management, project management, temporal structures, time management.

INTRODUCTION

When dealing with competing time demands across multiple projects and activities, people often make judgments about priorities based on their past experiences, values and interpretations that are difficult to articulate. Yet, each individual knows exactly how he/she will react to a new deadline even without such articulation. In this study, we speculate that organizations that pay more attention to understanding, codifying, and transfering the tacit time management strategies of their knowledge workers may better achieve their productivity goals. This research was designed and carried out to investigate individual time management experiences, which we plan to further categorize into different types of strategies that promote knowledge creation, storage, transfer and application. The paper aims to address some broad research questions:

RQ1: Which time management strategies do individuals employ when dealing with projects and completing their daily tasks in organizations?

RQ2: How are individual time management strategies influenced by organizational practices (that is organizational time demands)?

RQ3: How do individual time management strategies contribute to knowledge creation, storage, transfer and application in organizations? More specifically, how do individual time management approaches impact project knowledge construction processes?

The nature of time is multifaceted, and it impacts the management of projects, such as project deadlines, milestones and project goals, etc. This paper looks at the connections between project management and knowledge management, and at how time affects this connection. Such a link enables us to extend the use of knowledge management taxonomies to code and interpret how knowledge workers structure and manage their time in organizational processes and projects. Moreover, through capturing and analyzing individual time management profiles, we will gain useful insights on knowledge practices associated with projects and organizational routines, which may in turn affect organizational learning and performance.

THEORETICAL FOUNDATIONS:
TIME, PROJECT AND KNOWLEDGE MANAGEMENT CONNECTIONS

A review of the extant literature leads us to explore the distinct and unique features of time management strategies as well as their integration, driven by project deadlines and knowledge preservation and creation needs in organizations. Time, project and knowledge management appear connected in multiple ways, as depicted in Figure 1, which uses keywords to illustrate both the unique features and the connections among these three areas.
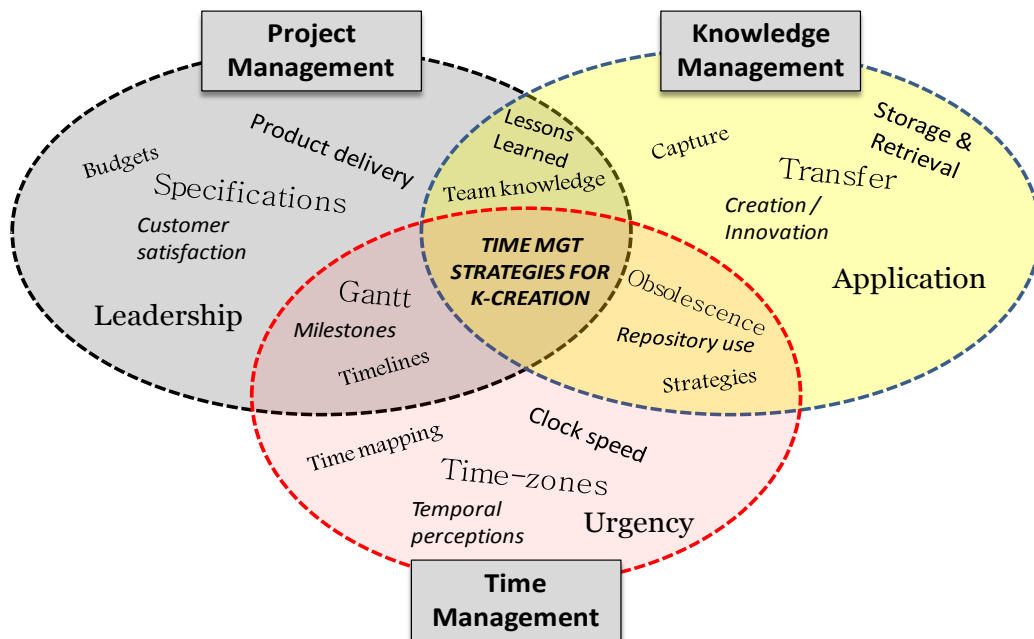
Figure 1: Relationships among Time Mgt., Knowledge Mgt. and Project Mgt.

Time Management and Project Management

As the most precious resource of an organization, time has long been a key indicator to plan, measure and evaluate organizational performance. Time management is necessary to execute various project activities in organizations, and actively engages actors who carry out their tasks at the individual, group, organization or inter-organizational levels. These actors exhibit various temporal perceptions (e.g., different experiences on the speed of time passing) and temporal personalities (e.g., different time urgency orientation) that impact the planning, execution and coordination of various organizational activities. A prior study [10] indicates that different deadline perceptions and behaviours among team members affect the ability of project teams to meet deadlines.   Understanding how these perceptions shape individual and organizational actions is essential to increase the efficiency and the effectiveness of institutional responses.

The connections between time management and project management are relatively well-grounded in the core aspects of the project management discipline. The definition of a "project" itself embeds the notion of time, with a project being a temporary endeavour that presupposes a beginning and an end. One of the key metrics for measuring project success is temporally-bound as projects must be "on-time," "on-budget" and "according to specifications." Prior time management research suggests that time is not perceived only through its "clock speed," although it is typically interpreted as a series of linear deadlines that are used for project control. Time-speed is also based on individual perceptions and time-orientations. Such perceptions pave the way to more or less successful interventions, such as the ability to create connections across diverse actors with timelines acting as boundary-spanning objects.

Thoms and Pinto [9] closely link the notion of project leadership to individual temporal perceptions and the development of related temporal skills. They map individual perceptions of past, present, and future arguing that the alignment between project needs and individual temporal perceptions may lead to better project outcomes. Actors in projects regulate their interactions and enhance their understanding of the project tasks by charting visual time representations (e.g., GANTT, PERT, milestones and other timeline-driven charts). Yakura [13] defined such visual timelines as temporal boundary objects, which enable interactions across various project agents who interpret and weigh these timelines based on their roles. Time-based visualizations foster a better understanding of how project practices are organized and sequenced; and drive visually-displayed project functions such as scheduling, allocation, task synchronization, and negotiation [13].

Project Management and Knowledge Management

Reich et al. [8] assert that much PM research has focused on the "action perspective" of projects. The action perspective centers on tasks, budget, people, and schedule management with a goal of controlling risks (embedded in size, complexity, organizational support and change management) that may hinder project success. They recognize and advocate the recent growth of a new

perspective in which knowledge and learning are paramount and take a key role in meeting fundamental project objectives ("knowledge-based perspective). A project success is measured by the amount of new knowledge and learning that it generates (or preserves) rather than simply through its product/service outcomes. The researchers assert the urgency of adopting a knowledge-based perspective because lack of attention to new learning leads to knowledge-loss at multiple phases of a project. The failure points (or knowledge loss points) could be, for example, the inability to use lessons learned as part of the project inputs; and/or incomplete knowledge integration and transfer in the planning stages of the project, including knowledge dissipation among project phases due to volatility and turnover among the governance team and project team members.

In summary, Reich et al. [8] establish that the lack of organization, socialization, integration, and transfer of knowledge negatively impact project performance. Their research also recognizes that such losses are affected by a temporal dimension. They use a mapping technique similar to Thoms and Pinto's [9] past, present and future project dimensions with a focus on temporal states such as project initiation (past resources focus), execution (present), and outputs (future products and learning achieved). Their model links project performance to project learning, based on knowledge creation, integration, transfer and coordination processes.

If we consider that the purpose of knowledge management is to create, organize, store, transfer, and coordinate the adoption of good business practices to improve organizational performance [5] [6], it follows that using a knowledge-based perspective in any project undertaken by the organization should enable focusing not only on project success but on overall organizational growth through avoiding knowledge dissipation. Since the realization of knowledge gains (or the avoidance of knowledge losses) and project performance have been identified as being bound to individual time-orientations in Thoms and Pinto's work [9], our research uses knowledge-based taxonomies to understand the temporal-orientation of various organizational players in specific settings and roles.

RESEARCH METHODOLOGY AND CURRENT STATUS

This study is exploratory and utilizes an interpretive approach based on qualitative data collection approaches. Semi-structured interviews, field observations and content analyses were used to identify knowledge workers' time management patterns and decision making processes when they dealt with deadlines. In-depth interview studies and field observations were conducted to capture individual time management practices. One of the two study sites was a US public research university in the East Coast, where twenty knowledge workers were interviewed in terms of their long-term and short-term time management strategies. All interviews were audio-recorded and resulted in over 350 pages of transcripts. The research was later replicated at another US University in the West with sixteen information technology professionals. We completed the collection and data analysis from the first site, and we are currently working on data analysis from the second site (and its integration).

While organizational knowledge management focuses on institutional practices as key analytical variables, individual knowledge management reflects on how individuals gather, classify, store, search, retrieve, and share knowledge in their personal practices [4]. Personal knowledge is interrelated with the organizational knowledge with an emphasis on personal information sharing, social interactions, and learning [11]. Because our objective is to categorize the types of knowledge utilized in individual time management practices, such as managing their daily tasks (within projects) or routines (within processes), Alavi and Leidner's [1] [2] knowledge taxonomies serve as the primary theoretical foundation for this research. Alavi and Leidner [2] classify knowledge into four different categories: (1) explicit and tacit, which refers to the nature of knowledge; (2) individual and social, which relates to its key actors; (3) creation, retrieval, transfer and application, which connects to the key knowledge processes; and (4) procedural, conditional, relational and pragmatic, which is relevant to defining the objectives of specific actions (enhance a process or establish connections).

In terms of the nature of time, we refer to "explicit knowledge" as articulated, generalized and codified understanding of time structures and deadlines, and "tacit knowledge" as rooted in actions, experiences, and dependent on a specific temporal context [7]. When we categorize time actors/users, we expand from "individual and social" to four types including "individual, group, organization and inter-organization" to more accurately represent the individuals' internal and external environments. It is important to make this distinction because much research on time management has recognized the important role that institutional time practices and deadlines play in organizations [3].

Regarding time management processes and activities, we directly use Alavi and Leidner's [2] classification which refers to "knowledge creation, storage/retrieval, transfer and application" process. However, we extend the classifications to a time-scheduling perspective. More specifically, we focus on understanding on which activities knowledge workers spend most of their time by evaluating the drivers of organizational knowledge capture. For example, "*knowledge creation*" involves scheduling to develop new content or enhance the existing content; "*knowledge storage/retrieval*" involves using scheduling tools to codify and retrieve content on a timely basis and is articulated through written documentation, and structured information in databases; "*knowledge transfer*" is interpreted as using time to transfer (information and knowledge) between individuals, from individuals to groups, between groups, across groups, and from the group to the organization; "*knowledge application*" involves time on action and execution.

Finally, to understand the objectives of individual time use, we adopted Alavi and Leidner's categories such as "procedural, conditional, relational and pragmatic" again within a temporal context. For example, *pragmatic* knowledge is regarded as use of time for practical outcomes; *procedural* knowledge refers to the use of time to better understand processes and procedures; *conditional* knowledge means to understand the temporal context (know-when of a task, routine or work environment); and *relational* knowledge is time-focused on socialization and increase of knowledge exchanges.

PRELIMINARY FINDINGS AND UPCOMING TASKS

To analyze the qualitative data collected, we broke down the long transcripts into a subset of short statements (i.e., coding units) organized according to the knowledge taxonomies above described. After two independent researchers coded all units, the inter-coder reliability was calculated using Cohen's Kappa values. The results show an overall acceptable level of reliability (Cohen's Kappa > 0.61). In the second study site, data was encoded by two additional researchers (for a total of four coders).

In terms of nature of time, 92.8% of coding units were coded as "explicit" time knowledge and only 7.2% belong to "tacit" time knowledge, since most professionals were able to decode and explain the strategies they use for time management. They explicitly clarified time boundaries and approaches for meeting exact deadlines. The determinants of the temporal norms and boundaries mainly involved individual (39.4%) and organization (48.9%) actors. Knowledge transfer (30.2%) and application (42.2%) processes played a key role in personal time management practices. The primary purpose/use of time was conditional (31.9%: know-when to complete a task) and pragmatic (38.8%). Table 1 presents an example of the typical statements coded under selected knowledge types identified in Alavi and Leidner [2] taxonomy.

RELATED TO
KNOWLEDGE PROCESSES AND ACTIVITIES

Creation	**Involves scheduling to develop new content or enhance existing content.** Example: *"Well for the year of my sabbatical my short term goal is to write some papers and get the first draft done of the book."*
Storage/ Retrieval	**Involves using scheduling tools to codify and retrieve deadlines and is articulated through written documentation, structured information in databases etc.** Example: *"Well. I use my Palm. But, I don't really…I mean it's kind of like I put down meetings there I don't really…and I put down tasks that I have and make sure that I don't forget them."*
Transfer	**Refers to using time to transfer (information and knowledge) between individuals, from individuals to groups, between groups, across groups, and from the group to the organization.** Example: *"But I really enjoyed time I spend with my students, one on one. That is one of my favorite times during the week. So I enjoy it and I enjoy what comes out of it because students really respond to that individual attention and it's very rewarding for me to do that."*

Application	**Involves focusing on actions and application.** _Example:_ _"And it was managing your email traffic. They told us to create folder, where you have folders A, B, C, and D, like your priorities. But the folder D is where you put the e-mails that maybe are your friends' emails. They don't need to be responded immediately, and can be responded later."_

RELATED TO
THE PURPOSE / USE OF TIME

Procedural	**Know-how** Example: _"Inside of my calendar, there are clips. I have like file folder labels; then I have little post-it notes. And I use colors to show the different classes."_
Conditional	**Know-when** Example: _"I don't have to come to school on Thursday, Friday, Saturday or Sunday."_
Relational	**Know-with** Example: _"If there are parties held for the department, I usually hold the parties, or I attend the parties, and I do that simple because really nice if I go…other people notice, and people appreciate that."_
Pragmatic	**Useful and practical time management** Example: _"I try to put due dates that are earlier than the real due dates." "Well not newspapers, usually during the meeting, I am doing grading and skimming articles, reading articles."_

Table 1: Taxonomy of Time Management Coded Instances (adapted from [12]).

Based on the analysis of the data from the first study site, we identified temporal behaviours of various knowledge workers. We also carefully analyzed personal profiles and calculated the frequencies of the recurrence of the same type of statements (same type of temporal perception profiles) for each type of temporal knowledge. Our upcoming focus is to identify specific individual time management profiles that leverage the knowledge-based perspective, such as a knowledge task/process. This will expand earlier project management research that clustered users based primarily on their time orientation. We expect to be able to identify specific individual profiles (like those in Waller et al. [10] described next) that will inform us on how individuals in organizations spend most of their time in relation to building and extracting knowledge from projects and processes. This will provide an opportunity to trace a preliminary organizational picture and will suggest ways to reduce knowledge-loss across project activities and practices.

A prior study [10] indicates that different deadline perceptions and behaviors among team members affect the ability of project teams to meet deadlines. Combining individuals' time urgency and time perspective on present and future orientations, Waller et al. [10] proposed four individual temporal personality prototypes including *relator*, *visioner*, *crammer*, and *organizer*, which are useful to understand individual time behaviors. *Relators* and *visioners* have low time urgency and are not likely to pay attention to deadlines. They have little desire to schedule or control deadlines and often underestimate the passage of time. They are likely to focus on present-moment relations with others. *Relators* can enjoy the present and also lose track of time. *Visioners* have low levels of impatience and irritability. They underestimate the passage of time and generally do not pay attention to deadlines. However, when they do pay attention to future orientation, they attend to deadlines. In comparison, *crammers* and *organizers* have high time urgency. *Crammers* are characterized by the need to exert control over deadlines, competitiveness, achievement focus, and show high impatience and irritability. They have day-to-day orientations without focusing on future goals. *Organizers* are characterized by achievement focus, high time awareness, and the need to define clear schedules.

With a team full of *relators* and *visioners,* it may be difficult to meet the project deadlines. Reversely, it will be easier to meet deadlines if teams are composed of *organizers* and *crammers*. Therefore, understanding the above four individual temporal personality prototypes can be useful for better time coordination, activity execution and decision making in organizations. Moreover, in a highly interdependent and time constrained work context, a shared understanding of time resources and individual temporal behaviors will even provide more support on project coordination and execution.

STUDY IMPLICATIONS AND LIMITATIONS

It will be beneficial to achieve an in-depth understanding of individuals' temporal behaviours, which are useful for project managers to better manage their project teams, and eventually to improve organizational productivity. As mentioned, Waller et al. [10] identified four individual temporal prototypes (*relator*, *visioner*, *crammer,* and *organizer*) according to time urgency and time perspectives. In reality, individual temporal prototypes may be much more complicated, since many dynamic factors shape and re-shape individuals' temporal behaviours. However, to some extent, the identified temporal prototypes can provide helpful guidelines to efficiently manage project teams. When configuring a project team, project managers should not ignore temporal personality, since it is predicted that if all team members have a low time urgency character, it is difficult to meet project deadlines. In contrast, a team full of high time urgency individuals (i.e., *crammers* and *organizers*) is much more likely to meet project deadlines [10]. Especially individuals with high time urgency and high future time perspective (i.e., *organizers*) can become key personnel for the project, since they usually proactively plan and manage their deadlines with high creativity.

In practice, many projects do not follow regular routine practices, so project activities can have ever-changing needs, which require new learning and knowledge creation efforts. Project leaders can then match temporal personalities to team configuration, such as to assign *visioners* and

organizers to do problem-solving especially in the project initiation phase, since a prior project management study [9] shows that individuals with different temporal personalities can best match different project tasks in different phases. In addition, effective learning, knowledge creation and capture should also take place to avoid knowledge losses which may result in the project failure. Many other contextual factors may also impact project activities, which often cause unusual events with time constraints. For such unusual emergent projects, low time urgency individuals (such as *relators* and *visioners*) should not be assigned to project teams, as they have low time awareness and often ignore the passage of time, which can cause serious project delays.

The biggest limitation of this research is its limited generalizability which is due to our focus on specific study sites, such as academic institutions. While these sites are instrumental when taking the knowledge-based perspective, as the main focus of an academic institution is knowledge transfer and creation, we aim to expand our study to more diverse business organizations, in particular, project-oriented organizations with different levels of actor involvement. We speculate that the insights obtained from this study will improve project management practices. Indeed, a successful project manager should be aware of the temporal perceptions of both project stakeholders and project teams. Better knowledge creation in organizational projects and practices will support more effective and efficient project coordination as well as the delivery of high-quality and more creative products.

REFERENCES

[1]   Alavi, M. and Leidner, D. E. (1999). Knowledge management systems: Issues, challenges and benefits, *Communications of the Association for Information Systems (CAIS),* 1, 1-37.

[2]   Alavi, M. and Leidner, D. E. (2001). Knowledge management and knowledge management systems: Conceptual foundations and research issues, *MIS Quarterly*, 25, 1, 107-136.

[3]   Blount, S. and Janicik, G.A. (2001). When plans change: Examining how people evaluate timing changes in work organizations, *The Academy of Management Review,* 26, 4, 566-585.

[4]   Grundspenkis, J. (2007). Agent based approach for organization and personal knowledge modelling: knowledge mgt perspective, *Journal of Intelligent Manufacturing*, 18, 4, 451-457.

[5]   Nonaka, I. (1994). A dynamic theory of organizational knowledge creation, *Organization Science*, 5, 1, 14-37.

[6]   Nonaka, I. and Takeuchi, H. (1995). The knowledge-creating company, Oxford Univ. Press, NY.

[7]   Polanyi, M. (1967). The tacit dimension, Anchor Books, New York.

[8]   Reich, B.H., Gemino, A. and Sauer, C. (2008). Modeling the knowledge perspective in IT projects, *Project Management Journal*, 39, Supplement, 4S-14S.

[9]   Thoms, P. and Pinto, J. K. (1999). Project leadership: A Question of timing. *Project Management Journal*, 30, 1, 19-26.

[10]   Waller, M., Conte, J., Gibson, C. and Carpenter, M. (2001). The effect of individual perceptions of deadlines on team performance. *Academy of Management Review*, 26, 4, 586-600.

[11]   Wright, K. (2007). Rethinking knowledge work: Supporting work and learning through personal knowledge management, KRW Knowledge Resources. [On-line http://www.knowledgeresources.ca/Knowledge_Resources/PKM_Planning.html]

[12]   Wu. D., Passerini, K. and Bartolacci, M. R. (2008). By when do you need this done? Discovering knowledge workers' time management practices, Proceedings of the 14th Americas Conference of Information Systems (AMCIS 2008), Toronto, Canada.

[13]   Yakura, E. K. (2002). Charting time: Timelines as temporal boundary objects. *Academy of Management Journal*, 45, 5, 956-970.

Knowledge Management: A Gap Analysis Framework

Richard Briotta
Bay Path College
rbriotta@baypath.edu

Today organizations view knowledge as an important strategic asset. As such, an organization must focus on gathering and developing knowledge and expertise that can be transformed into competitive advantages. This study expands upon strategic and knowledge management literature by proposing a gap analysis framework. This framework classifies the organization's current knowledge into three categories, core, advanced and innovative. The results of this analysis indicate knowledge gaps in an organization's capability to address the changes in its competitive environment. As a result, the framework can be utilized as an effective tool to determine the needed knowledge for formulation of competitive strategies.

The CONCEPTUAL MODEL of KNOWLEDGE ACQUISITION SYSTEM in ACCOUNTING and LEGAL FIRMS

Eric Y. Cheng, SUNY-Canton, 315-379-3904, cheng@canton.edu

James Fok, Hong Kong Polytechnic University, (852) 2766-7370

ABSTRACT

Knowledge acquisition is recognized as a critical knowledge management process for developing knowledge management systems. For accounting and legal industries, to transform tacit knowledge from the professional experts to explicit knowledge could be problematic as human judgment could vary substantially among the experts. In addition, maintenance of the systems could require regular interaction between the system engineers and the experts which could be time consuming and costly. This paper addresses the motivators and barriers of knowledge acquisition during the development of knowledge based systems. A conceptual model for integrated knowledge based system is introduced to resolve certain barriers during knowledge acquisition. This analysis could also be used to help accounting and legal industries to understand how to better manage existing invaluable knowledge assets.

 Keywords: knowledge management, tacit knowledge, knowledge acquisition, accounting

INTRODUCTION

Though the focus on computer applications in accounting remains in how to use advanced decision support systems to support auditing [8], there is still needs to develop a knowledge management system that can acquire tacit knowledge effectively and efficiently. Extensive researches and studies have been conducted by many scholars to resolve the deficiencies during knowledge acquisition process. J.W. Moody et. al. [11] investigated the applicability of cognitive interview in the knowledge elicitation phase of expert system development. P.H. Hendriks [14] studied the organizational impact of knowledge based systems. N.F. Matsatsinis et. al. [13] proposed a methodology which combines interviews with experts and known literatures for eliciting expertise. W. P. Wagner et. al. [21] presented a mapping between the body of knowledge acquisition empirical studies and the different problem domains within accounting and finance as a guide for choosing knowledge acquisition techniques.

The architecture of an integrated knowledge system is also explored. H.C. Tu and J. Hsiang [4] introduced an architecture of intelligent information retrieval agents

to support group knowledge retrieval. E. Davenport [2] addressed the issues of knowledge management under communities of practice infrastructure. H.R. Nemati et. al. [5] proposed the adoption of knowledge warehouse, which integrates knowledge management, decision support, artificial intelligence and data warehousing. H.C.W. Lau et. al. [3] proposed an integrated knowledge system model which contains data warehouse, OLAP system and case base reasoning component to resolve the concern about shortage of knowledge workers.

LITERATURE REVIEW

Knowledge management in accounting industry

Tom C. Davis [19] addressed the importance of knowledge management to public accounting firm and that better use of practice's knowledge to improve firm's efficiencies is vital. The global big four accounting firms are well aware of the importance and the firms invested considerable resources in developing knowledge based infrastructure. Knowledge based systems are widely adopted in large accounting firms. V. Karen et. al. [20] assessed the suitability of judgmental auditing tasks for expert systems development. They noticed that the Big 6 accounting firms (some of which have merged and consolidated to be the Big 4) have developed auditing expert systems to support their auditing business.

On the other hand, the small accounting firms could also adopt commercially available expert system development tools in building their knowledge based systems. The commercial packages contain basic components of expert system, which include, knowledge base, inference engine, user interface and explanation facility. In principle, the small accounting firms can directly develop their knowledge base systems by using the packages without any support from knowledge engineers.

The development of knowledge based system for accounting profession also received attention from scholars. J.K. Lee and M.W, Jeong [10] developed a prototype IAPS (Intelligent Audit Planning System) to support the auditor assignment process for audit engagements. S. Lee and I Han [16] introduced EDIRDB, a prototype audit support system to support EDI auditors. I. Comyn-Wattiau and J. Akoka [6] presented the application of INFAUDITOR, an audit expert system, to logistics information systems auditing. In view of the difficulties during knowledge acquisition process in developing expert system, M. Anandarajan and M. Anandarajan [12] proposed to adopt machine learning techniques for auditors's going concern reporting. System to be built under neural network model was examined.

Knowledge management in legal firms

T. Anandanpillai and T.A. Barta [17] described a case-based reasoning system for housing discrimination laws in the United States. Y.H. Tan and W. Thoen [22]

presented INCAS, a legal expert system for contract terms in electronic commerce. T.A. O'Callaghan et. al. [18] introduced SHYSTER-MYCIN, a legal expert system which combines case-based legal expert system (SHYSTER) with a rule-based expert system (MYCIN).

Dan Hunter [1] advocated the idea of commercializing Legal Neural Networks and argued that the time is ripe for practitioners to adopt neural network technology for legal inference. He illustrated the usage by referring to an example of a court in granting bail to an offender. The input layer captures factors such as habitual criminal, sex crime etc. and signals (yes =1 or no = 0) will be assigned. The signals will be multiplied by allocated weightings in the hidden layer to determine whether the value will exceed the threshold value. If the value exceeds the threshold value, the neurode will pass a signal to the output layer to suggest the nature of the bail granted. Dan Hunter [1] admitted that the neural network, as their intelligence rests with the weightings on links between neurodes, could be "quintessential black box" and unable to provide reasons for generating the outputs (conclusions). It's considered that legal neural network can be applied as a quick predictive tool.

Motivators and Barriers for Knowledge Management

This section examines the difficulties of knowledge acquisition of knowledge based system. There are typical and well researched barriers including: difficulties of knowledge acquisition, communication between domain expert and system engineer, Technical support and social background, psychological concern, and financial constraint. Based on the review of related research studies and reports on knowledge management, 18 motivators and 18 barriers are identified as relevant to the knowledge acquisition of knowledge base systems for accounting and legal firms.

The motivators and barriers are summarized in the following table 5.

Table 5. Motivators and Barriers of Knowledge Management

Motivators	Barriers
1. Committed management support	1. High cost of development
2. Perceived ease of use	2. Poor relationship between knowledge engineer and experts
3. Establish knowledge sharing culture	3. Lack of commitment from senior management
4. Incentive schemes for knowledge sharing	4. System too complicated
5. User innovativeness reward	5. Lack of proper information infrastructure
6. Infrastructure of knowledge network	6. Absence of knowledge policy
7. Relative advantage against competitors	7. Inadequate time to support knowledge acquisition
8. Link with business objectives	8. Inadequate technical support
9. Supportive working environment	9. Afraid of management change
10. Improve skills of domain experts	10. Personal benefits not perceived
11. Better decision making	11. Chinese culture not support knowledge transfer
12. Better customer handling	12. Difficult to integrate with daily work
13. Sharing best practice	13. Lack of training
14. New ways of working	14. Unable to measure the financial benefits derived
15. Staff attraction	15. Not an industry norm
16. Compatible system	16. Inadequate channels for knowledge flow
17. Create new business opportunities	17. Poor communication among employees
18. Increase market value and profit	18. Organization structure does not support knowledge sharing

Conceptual model of integrated system for accounting and legal firms

This conceptual model takes into account of recent developments in addressing the specific requirements in accounting and legal industries. Intelligent agent and organization impact will also be incorporated into the system. As IT strategies to align with business strategies is essential for all organization, strategic knowledge will be stored in the integrated system to facilitate alignment. The Nonaka's knowledge spiral model [7] was adopted for the illustration of the system framework.

The conceptual framework is depicted in figure 3 below.
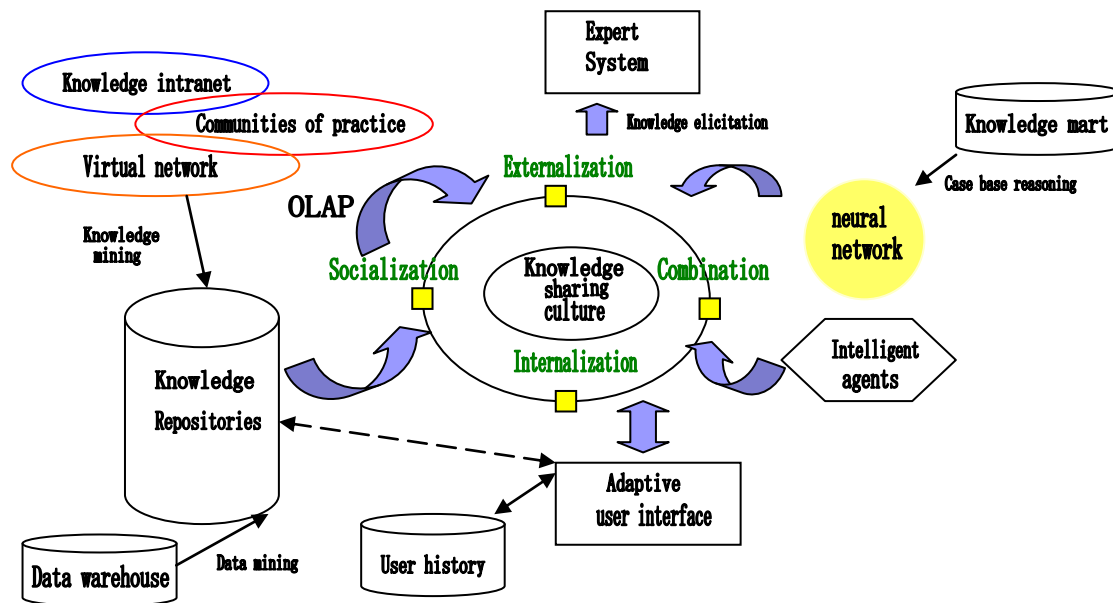
Figure 3: Conceptual integrated knowledge based system

Both machine learning and human expert knowledge transfer are incorporated in the integrated system. On line knowledge support from professional bodies such as Hong Kong Law Society and Hong Kong Society of Accountants will be linked to the integrated system. Knowledge mining software will be installed for knowledge discovery. Tacit knowledge stored in knowledge repositories will be converted to explicit knowledge. The domain experts, inspired by new knowledge discovered, will assist the domain experts in developing new knowledge. The new knowledge can be elicited into the system. As most of the experts in accounting and legal firms are accustomed to computerized environment, it is appropriate to adopt expert system shells for the domain experts to communicate directly with the system.

Case-based reasoning is useful for both tax specialists and lawyers. The knowledge- based neural network will provide quick solution for decision support. Neural network integrated with case-based reasoning mechanism will provide proposed solution which is useful for referencing. Intelligent information retrieval agents as proposed H.C. Tu and J. Hsiang [4], will assist the organization to obtain information and knowledge through the internet. The agent could assist in document classification and will be beneficial for categorizing the new knowledge acquired.

Adaptive user interface will allow customization to meet the different user background. As the domain experts are both knowledge provider and knowledge receiver, adaptive system environment will ease the acceptance to the system. The organizational impact should not be overlooked. Knowledge sharing culture has to be developed through training, incentive scheme, profit sharing and other reward systems

in order to ensure the system could be operated effectively. The knowledge intranet will serve as a channel for communicating strategic knowledge.

The conceptual model overcomes certain difficulties in knowledge acquisition, such as communication problem, interpersonal relationship as well as certain psychological concern. For further enhancement of the proposed model, it is useful to identify the main motivators and barriers of knowledge acquisition.

DISCUSSION

The studies are relevant to knowledge acquisition as domain experts are also the users of the knowledge base system. Supportive social environment is crucial for knowledge transfer, in particular, the consultants in accounting firms. Perceived ease of use as already covered in TAM, which is an important factor in developing knowledge base system. Perceived consequence could be a motivator for knowledge acquisition if the experts can perceived benefit derived from the system. It is easier for the domain experts to accept the system if the role of system is inclined for decision support instead of decision making. The experts could be afraid of being replaced by the system.

P. Poon and C. Wagner [15] studied the critical success factors for the development of information systems. They defined five evaluation criteria for system success.

The five criteria are:-
1. The system is made available for access by users.
2. The system is used by intended users.
3. Users are satisfied with the system.
4. The system has positive impact on senior management and the organization
5. The system tends to spread across the organization.

The drivers for system success are named as critical success factors and they adopted J.F. Rockart and D.W. Delong [9] together with other researchers' studies for examining the impact of ten critical success factors to implementation of information system.

The ten critical success factors are:-
a. Committed and informed executive sponsor
b. Operating sponsor
c. Appropriate IS staff
d. Appropriate technology
e. Management of data
f. Clear link to business objectives
g. Management of organizational resistance

h.   Management of system evolution and spread

i.   Evolutionary development methodology

j.   Carefully defined information and system requirement

The result of their research confirmed J.F. Rockart and D.W. Delong's [9] study as well as the two additional critical success factors.

The identified critical success factors could be relevant to knowledge acquisition in knowledge base system. Committed management could drive the experts in devoting their time and effort for the development of the knowledge base system. Quality of the IS support is important. In the absence of knowledge engineers, the organization may require to appoint external IT consultancy firm to provide IT support to the domain experts for using the expert system shells or resolving any technical problems which may arise. The critical success factors include organizational factors. Link to business objective could also be crucial. Alignment with business strategies will support the stability of IT infrastructure which is essential for sustained knowledge acquisition. Operating sponsor is important as the position will manage the details of implementation from user's side. Representative of the domain experts could be useful for clarifying the intended use of the knowledge based system which could reduce friction between the domain experts and the management during knowledge acquisition process. Management of organization resistance could relate to organization culture. Knowledge sharing culture may encounter resistance and education and incentive schemes could help the management to reduce conflicts. Management of data refers to the access of data, externally and internally. Knowledge network and knowledge repositories will provide adequate access to required knowledge by users. Evolution of the system is inevitable and the domain experts should be motivated for modification and enhancement of the system.

REFERENCES

[1] Dan Hunter, 1996.   Commercialising Legal Neural Networks.   The Journal of Information, Law and Technology. elj.warwick.ac.uk.

[2] E. Davenport, 2001.   Knowledge Management issues for online organizations: "Communities of Practice' as an exploratory framework. Journal of Documentation, Vol. 57, No.1, January 2001.

[3] H.C.W. Lau, C.W.Y. Wong, I.K. Hui, K.F. Pun, 2003. Design and implementation of an integrated knowledge system. Knowledge-Based Systems 16 (2003) 69-76.

[4] H.C. Tu, J. Hsiang, 2000.   An architecture and category knowledge for intelligent information retrieval agents. Decision Support Systems 28 (2000) 255-268.

[5] H.R. Nemati, D.M. Steiger, L.S. Iyer, R.T. Herschel, 2002. Knowledge warehouse: an

architectural integration of knowledge management, decision support, artificial intelligence and data warehousing. Decision Support Systems 33 (2002) 143-161.

[6] I.Comyn-Wattiau, J. Akoka, 1996. Logistics Information System Auditing Using Expert System Technology. Expert Systems with Applications, Vol. 11, No. 4, pp. 463-473.

[7] I. Nonaka, 1994. A dynamic theory of organizational knowledge creation, Organization Science 5 (1) (1994) 14-37.

[8] J.E. Hunton, J.M. Rose, 2010. 21st Century Auditing: Advancing Decision Support Systems to Achieve Continuous Auditing, Vol. 24, No. 2, Accounting Horizons, American Accounting Association, pp.297-312

[9] J.F. Rockart, D.W. Delong, 1988. Executive Support Systems The Emergence of Top Management Computer Use. Dow Jones Irwin, Homewood, IL.

[10] J.K. Lee, M.W. Jeong, 1995. Intelligent Audit Planning System for Multiple Auditors: IAPS. Expert Systems with Applications, Vol. 9, No. 4, pp. 579-589.

[11] J.W. Moody, R.P. Will, J.E. Blanton, 1996. Enhancing Knowledge Elicitation using the Cognitive Interview. Expert Systems With Applications, Vol, 10, No. 1, pp. 127-133.

[12] M. Anandarajan, A. Anandarajan, 1999. A comparison of machine learning techniques with a qualitative response model for auditor's going concern reporting. Expert Systems with Applications 16 (1999) 385-392.

[13] N.F. Matsatsinis, M. Doumpos, C. Zopounidis, 1997. Knowledge Acquisition and Representation for Expert Systems in the Field of Financial Analysis. Expert Systems with Applications, Vol. 12, No. 2 pp. 247-262.

[14] P.H. Hendriks, 1999. The organizational impact of knowledge-based systems: a knowledge perspective. Knowledge-Based Systems 12 (1999) 159-169.

[15] P. Poon, C. Wagner, 2000. Critical success factors revisited: success and failure cases of information systems for senior executives. Decision Support Systems 30 (2001) 393-418.

[16] S. Lee, I. Han, 2000. EDI controls design support system using relational database system. Decision Support Systems 29 (2000) 169-193.

[17] T. Anandanpillai, T.A. Barta, 1999. A case-based reasoning system for housing discrimination law. Expert Systems with Applications 16 (1999) 315-324.

[18] T.A. O'Callaghan, J. Popple, E. McCreath, 2003. SHYSTER-MYCIN: A Hybrid Legal Expert System. cs.aun.edu.au.

[19] T.C. Davis, 2000. Knowledge Management in a CPA Firm – Your Firm's Future Depends On It. www.itpna.com.

[20] V. Karan, U.S. Murthy, A.S. Vinze, 1995. Assessing the Suitability of Judgmental Auditing Tasks for Expert Systems Development: An Empirical Approach. Expert Systems with Applications, Vol. 9, No. 4, pp. 441-455.

[21] W.P. Wagner, J. Otto, Q.B. Chung, 2002. Knowledge acquisition for expert systems in accounting and financial problem domains. Knowledge-Based Systems 15 (2002) 439-447.

[22] Y.H. Tan, W. Thoen, 2000. INCAS: a legal expert system for contract terms in

electronic commerce. Decision Support Systems 29 (2000) 389-411.

Incentives for Multi-Dimensional Knowledge Sharing and Learning in Teams

Shankar Sundaresan

sundares@camden.rutgers.edu

School of Business

Rutgers, The State University of

New Jersey, Camden

Zuopeng Zhang

zzhan001@plattsburgh.edu

School of Business and Economics

State University of New York

at Plattsburgh

Abstract

We present a model for designing incentive structures for knowledge sharing taking into account the interactions between knowledge dimensions as well as team settings. Solving the organizational decision problem, we examine whether a linear reward system can be designed and implemented for multi-dimensional knowledge sharing and learning in a team setting. In addition, we study the effect of different types of interactions among knowledge dimensions on the design of organizational reward system and the output allocation rule. Our research presents a formal model of multi-dimensional knowledge sharing and learning and attempts to develop valuable insights for practitioners to manage knowledge assets with effective reward systems.

Introduction

With the advent of new information technologies, knowledge management (KM) has entered into a new era of development. For instance, social networking software enables knowledge workers to socialize, collaborate, and share knowledge. XML-based data structures facilitate the codification and exchange of information and knowledge. With knowledge management systems integrated with such technologies, companies can effectively implement various strategies for knowledge sharing and learning in organizations, such as replication, adaptation, codification and personalization.

However, technology is only one of the many enablers that facilitate knowledge sharing and streamline the knowledge management processes. Social exchange theory has identified many other factors that impact knowledge sharing and learning, including perceivable organizational support, learning capacity, innovative work behavior, information culture of organizations, social status, value of knowledge, participation inequality and conversational interactivity. Most notably,

knowledge sharing faces a major challenge because most knowledge workers tend to resist sharing their knowledge with others in the organization ([4] [3]). Offering incentives is one way to offset workers' resistance to share knowledge.

Incentives are identified as an important element to facilitate knowledge sharing and learning in knowledge management systems ([1] [2]). Some researchers have investigated the reward systems for knowledge sharing in organizations. For instance, [5] study the design of a reward system for knowledge sharing and compare the performance of an individual-based reward system with that of a group-based system. While their research only focuses on knowledge sharing (not learning aspects), [6] explore the joint role of incentives and information systems in knowledge sharing and learning in organizations, and briefly discuss a setting of knowledge in two independent dimensions with individual rewards. However, they neither consider a team-based multi-dimensional knowledge setting nor the interactions between knowledge dimensions.

Our study attempts to further the research in designing incentives for knowledge sharing and learning by incorporating the following three elements: (1) multi-dimensional knowledge setting; (2) knowledge sharing and learning in a team setting; (3) interactions between knowledge dimensions. In particular, we address the following research questions. First, how should a firm design the reward system for knowledge sharing in a multi-dimensional knowledge setting? Second, how do the different types of interactions among knowledge dimensions affect the design of the reward system? Third, how does the reward system for a team differ from that for independent workers?

Next, we present a brief outline of our model of knowledge sharing and learning with multi-dimensional knowledge context, and discuss our analysis approach.

Outline of Model and Analysis

We consider a firm that employs n knowledge workers in a team to complete a project. Upon completion of the project, the team generates an observable output x. We next describe the time sequence of five stages in our model.

1. The firm announces an allocation rule s for distributing the output x and a reward system R_s for knowledge sharing. The allocation rule $s(\Gamma_i, x)$ depends on each worker's reported knowledge level Γ_i and the actual output x. The reward system specifies how workers will be rewarded based on the amount of knowledge they share with other workers through the KMS.

2. Each worker i reports her knowledge level Γ_i to the firm. We assume that each worker i possesses knowledge in two dimensions and she will report her knowledge level in these two

dimensions, i.e.,

$$\Gamma_i = \begin{pmatrix} \gamma_{i1} \\ \gamma_{i2} \end{pmatrix}.$$

The reported knowledge level Γ_i is observable to all the workers in the team.

3. Workers share knowledge and learn from knowledge providers. To maximize her total expected payoff, worker i determines the amount of knowledge K_i^s to share (to obtain sharing reward) and the amount of knowledge K_i^l to learn (to improve productivity). Worker i incurs a sharing cost $C_s(K_i, K_i^s)$ and a learning cost $C_l(K_i, K_i^l)$ when she shares knowledge and learns, where

$$K_i = \begin{pmatrix} k_{i1} \\ k_{i2} \end{pmatrix}, \quad K_i^s = \begin{pmatrix} k_{i1}^s \\ k_{i2}^s \end{pmatrix}, \quad \text{and} \quad K_i^l = \begin{pmatrix} k_{i1}^l \\ k_{i2}^l \end{pmatrix}.$$

Both the sharing and learning cost concavely decrease in k_{ij} and convexly increase in k_{ij}^s or k_{ij}^l, $\forall j = 1, 2$, implying that a worker with a higher knowledge level incurs lower sharing and learning costs.

4. Workers jointly complete the project with their upgraded knowledge $\bar{K}_i$ after learning, generating an output x, where

$$\bar{K}_i = \begin{pmatrix} \bar{k}_{i1} \\ \bar{k}_{i2} \end{pmatrix}.$$

When working on the project, each worker i exerts her effort e_i, while incurring a cost $C(e_i, \bar{K}_i)$. The effort cost is a convexly increasing function of each worker i's effort level e_i and concavely decreasing function of i's upgraded knowledge level $\bar{k}_{ij}(\forall j = 1, 2)$ in each dimension. The output x depends on each worker i's updated knowledge level $\bar{K}_i$ after learning, the effort e_i she exerts, and a factor θ that reflects the stochasticity of the outcome. Specifically, x concavely increase in e_i and $\bar{k}_{ij}(\forall j = 1, 2)$.

5. The firm allocates the realized output x and rewards workers for their knowledge sharing and learning, based on the pre-announced allocation rule s and reward system R_s.

Given this setting, a knowledge worker i determines her effort level e_i, the amount of knowledge K_i^s to share and K_i^l to learn, and her knowledge level Γ_i to report, so as to maximize her total

expected payoff $E[\pi_i]$, where

$$\pi_i = s(\Gamma_i, x) + R_s(\Gamma_i, K_i^s) - C(e_i, \bar{K}_i) - C_s(K_i, K_i^s) - C_l(K_i, K_i^l). \tag{1}$$

For simplicity, we assume there are two knowledge workers in the team. Hence, the firm makes decision on the allocation rule s and the reward system R_s for two workers to maximize its total expected payoff $E[\pi]$, where

$$\pi = x(\bar{K}_1, \bar{K}_2, e_1, e_2, \theta) - \sum_{i=1}^{2} [s(\Gamma_i, x) + R_s(\Gamma_i, K_i^s)], \tag{2}$$

Formally, the firm's decision problem [P] can be formulated as

$$\max_{s, R_s} \quad E[\pi] \tag{3}$$

subject to each worker i's incentive compatibility (IC) constraint

$$(\Gamma_i^*, e_i^*, K_i^{s*}, K_i^{l*}) \in \underset{\Gamma_i, e_i, K_i^s, K_i^l}{argmax} E[\pi_i], \tag{4}$$

and individual rationality (IR) constraint

$$E[\pi_i] \geq 0. \tag{5}$$

Conclusion

In analyzing incentive design for knowledge sharing, prior research has considered neither the interactions between knowledge dimensions nor the complexity of team setting. We address these issues in our model and present the design of incentive structures. Specifically, we solve the organizational decision problem and examine whether linear reward system can be designed and implemented for multi-dimensional knowledge sharing and learning in a team setting. In addition, we study the effect of different types interactions of among knowledge dimensions on the design of organizational reward system and the output allocation rule. In summary, our research presents a formal model of multi-dimensional knowledge sharing and learning and attempts to develop valuable insights for practitioners to manage knowledge assets with effective reward systems.

References

[1] Argote, L., B. McEvily, & R. Reagans. *Managing Knowledge in Organizations: an Integrative Framework and Review of Emerging Themes*. Management Science, 2003, 49(4), 571–582.

[2] Ba, S., J. Stallaert, & A. B. Whinston. *Research Commentary: Introducing a Third Dimension in Information Systems Design –the Case for Incentive Alignment*. Information Systems Research, 2001, 12(3), 225–239.

[3] Bock, G. W., & Y.-G. Kim. *Breaking the Myths of Rewards: An Exploratory Study of Attitudes About Knowledge Sharing*. Information Management Resources Journal, 2002, 15(2), 14–21.

[4] Ciborra, C. U., & G. Patriotta. *Groupware and Teamwork in R&D: Limits to Learning and Innovation*. R&D Management, 1998, 28(1), 43–52.

[5] Lee, D. J., J. H. Ahn. *Reward Systems for Intra-organizational Knowledge Sharing*. European Journal of Operational Research, 2007, 180(2), 938–956.

[6] Sundaresan, S., & Z. Zhang. *Knowledge Sharing and Learning in Organizations: Role of Incentives and Information Systems*. Working Paper, Rutgers—The State University of New Jersey, Camden, 2011.

KNOWLEDGE RESOURCE, PRODUCT INNOVATION PERFORMANCE, AND THE MEDIATING ROLE OF KNOWLEDGE UTILIZATION CAPABILITY

Yi-Pei Li

National Dong Hwa University, Dahan Institute of Technology, Taiwan, pn1880@ms23.hinet.net

Yuh-Yuan Tsai

National Dong Hwa University, Taiwan, yytsai@mail.ndhu.edu.tw

ABSTRACT

The purpose of this study is to understand how a firm employs knowledge to enhance radical and incremental product innovation performance. Therefore we are interested in the relationship between knowledge resource, knowledge utilization capability, and production innovation performance. In this study, we first discuss the impact of knowledge utilization capability on product innovation performance; then we examine the relationship of knowledge resources and knowledge utilization capability, and the mediating role of knowledge utilization capability.

Keywords: knowledge resource, exploration, exploitation

INTRODUCTION

Throughout a board set of industries, innovation underlies the ability of firms to sustain competitive advantages. Innovation is among the essential processes for success, survival, and renewal of organization, particularly for firms in either fasted-paced or competitive markets [5]. In the field of product innovation research, radical and incremental innovations are particularly important for firms to develop.

A central argument of the knowledge-based view is the differences in the knowledge resources and innovative capabilities of different firms are the main determinants of differences in their performances. Scholars stress the importance of intangible assets for attaining superior performance and a sustainable competitive advantage [8]. Among intangible assets, knowledge resources are the most important resources a firm controls. Of particular relevance for this research, scholars suggest that knowledge resources are principal inputs into the innovation process [17].

Although knowledge is a pivotal organizational resource contributed to innovation. However, possessing knowledge resources is not enough. A firm's ability to acquire, create and use knowledge effectively is recognized as a key factor determining organization success or failure [1]. But existing literature is unclear about how the interplay between knowledge resources and knowledge utilization capabilities contribute to product innovation performance. Our study is an attempt to address this issue and refine comprehension of the knowledge-innovation link. Therefore, the purpose of this study is to understand how the different types of knowledge resources influence radical and incremental product innovation performance through knowledge utilization capabilities.

KNOWLEDGE UTILIZATION CAPABILITY AND PRODUCT INNOVATION

Successful product innovation demands that a firm must exploit its existing techniques and skills while trying to avoid their rigidity effects by renewing and replacing them with entirely new ones. Innovation requires the mastery of two divergent tasks. On the one hand, a firm must center its attention on existing techniques to cultivate valuable and commercially viable products. On the other hand, a firm must continually acquire a diverse and novel body of knowledge to develop future technology [14, 16]. In order to master these two tasks, a firm needs to develop different knowledge utilization capabilities. To accomplish the former task demands that a firm must search local knowledge to refine and extend its existing product innovation techniques and processes, the knowledge utilization capability here is knowledge exploitation. Its aims are greater efficiency and reliability of existing innovation activities. To achieve the later task demands that a firm must search distant knowledge to acquire entirely new technologies and skills, the knowledge utilization capability here is knowledge exploration. Its objective is to attain flexibility and novelty in product innovation.

Radical innovation creates new platforms of innovation, and incremental innovation improves and refines these platforms until new platforms emerge [3]. Radical innovation is more likely to create a new sustainable competitive advantage, but is relatively rare [19]. Therefore, a firm requires incremental innovation to keep its competitive advantage until an opportunity for radical innovation presents itself. In general, radical innovations have greater value to firms than incremental innovations, particularly when radical innovators have deep pockets and strong market power [6]. In order to sustain existing competitive advantages and creating new one, a firm should try to develop radical innovations and never give up engaging in incremental innovations.

Incremental innovations are product improvements and line extensions that are usually aimed at satisfying the needs of existing customers. They involve small changes in technology and little deviation from the current product-market experience of the firm [2]. Previous research suggested that a firm can find out the solutions to customer problems through the search and refinement of existing technologies and processes [e. g., 4]. Through searching for solutions to customer problems in the neighborhood of the firm's current experience, exploiting local knowledge can help the firm increase efficiency and productivity [14]. Because refining and extending existing product innovation techniques and processes focus attention on productivity and performance improvement in existing products, the capability to exploit local knowledge increases incremental innovations. But the aims of knowledge exploitation are greater efficiency and reliability; it means that the firm may focus its attention on variety reduction and productivity improvements in existing products [7]. Therefore, we suppose that knowledge exploitation may prevent radical innovations.

Proposition 1: Firm's capability on knowledge exploitation is (a) positively related to incremental product innovations and (b) negatively related to radical product innovations.

Incremental innovations offer better value for existing customers. But existing customers may lose due to intensified competition or preference change. In order to achieve sustainable development, the firm must consider how to provide entirely new value for emerging customers. A firm can create new value and competitive advantage through the development of radical innovations. Scholars explained that radical innovations involve fundamental changes in technology for the firm, typically address the needs of emerging customers, are new to the firm and/or industry, and offer substantial new benefits to customers [6]. The development of radical innovations depends on the use of distant knowledge [13]. New combinations of distant knowledge may produce path-breaking innovations or enable the transition from an entrenched set of techniques and design to a new technological paradigm [16]. The capability to explore distant knowledge may help the firm to acquire new technologies and skills that a firm needs to produce entirely new innovative products. Therefore we posit that the capability to explore distant knowledge may produce radical innovations. Because the objective of knowledge exploration is to attain flexibility and novelty in product innovation, it involves experimentation for novel ideas and centers the firm's attention on developing radical rather than incremental innovations.

Proposition 2: Firm's capability on knowledge exploration is (a) positively related to radical product innovations and (b) negatively related to incremental product innovations.

KNOWLEDGE RESOURCES, KNOWLEDGE UTILIZATION CAPABILITY AND PRODUCT INNOVATION

Knowledge Resources

According to a knowledge-based view of organizations, the principal function of a firm is the creation, integration, and application of knowledge [8]. From this perspective, the development of sustainable competitive advantages is the identification, development and application of the firm's unique knowledge resources. It is widely accepted that a firm's capability to innovate is closely tied to its ability to utilize its knowledge resources. In this study, we use human capital resource and organizational capital resource as two types of knowledge resource. Human capital reflects the overall skill, expertise, and knowledge levels of an organization's employees [20]. Organizational capital is the institutionalized knowledge and codified experience residing within and utilized through databases, patents, manuals, structures, systems, and processes [22].

A firm's human capital resource influences its propensity to acquire new technologies and skills. Scholars suggested that the essence of human capital resource is the intelligence and creativity of the individuals [11]. The bright and skilled employees may encourage the questioning of prevailing norms and originate new ways of thinking [21]. It is these creative, bright, and skilled employees with expertise in their role and functions who constitute the predominant source for new ideas and knowledge in an organization [18]. These employees are willing to explore a variety of new and alternate knowledge domains. Access and exposure to diverse knowledge helps the firm to acquire new technologies and skills. Therefore, individuals and their associated human capital will help the firm to build the capability of knowledge exploration.

An organization's preserved knowledge influence its propensity to reinforce its existing techniques and processes. The mechanisms to use preserved knowledge is most evident in a firm's organizational capital resource. The hallmarks of organizational capital resource are the institutionalized knowledge, and codified experience residing within manuals, databases, patents, and licenses, along with the establishment of structures, processes, and routines that encourage repeated use of this knowledge [9]. The firm tends to use preserved knowledge in solving the problems about existing products or services, because preserved knowledge is generally perceived to be more reliable and robust than other knowledge. When a firm uses this knowledge repeatedly, it can reinforce its existing techniques and processes. Therefore, organizational capital resource will help the firm to build the capability of knowledge exploitation.

Proposition 3: The level of a firm's human capital is positively associated with the level of its capability on knowledge exploration.

Proposition 4: The level of a firm's organizational capital is positively associated with the level of its capability on knowledge exploitation.

The mediating role of knowledge utilization capability

Knowledge has become a critical ingredient for gaining a competitive advantage, particularly in the new economic landscape [8]. Knowledge resources are pivotal resources for the firm to develop its innovations. However, possessing knowledge is not enough. A firm's ability to acquire, create, transmit, and use of knowledge effectively is recognized as a key factor determining organization success or failure [1]. Firms with highly skilled and knowledgeable employees have higher levels of human capital and are more likely to create knowledge, make correct decisions and hence result in better organizational innovativeness (Hitt et al., 2001). Individuals and their associated human capital are crucial for an organization to search for distant knowledge and increase its capability to absorb new knowledge. Therefore, human capital will influence radical innovations through the influence on knowledge exploration.

Organizational capital is codified, and its creation, preservation, and enhancement occur through structured repetitive activities (Nelson & Winter, 1982). When a firm uses its preserved and codified knowledge repeatedly, it will refine the existing techniques [10], and help the firm to offer better value for existing customers. In general, the firm can use its organizational capital to strengthen the techniques and processes relevant to existing products, and provide quality or function upgrading product to maintain a close relationship with existing customers. Therefore, we posit that organizational capital will influence incremental innovations through the influence on knowledge exploitation.

Proposition 5: Firm's capability on knowledge exploration mediates the relationship between human capital and radical product innovations.

Proposition 6: Firm's capability on knowledge exploitation mediates the relationship between organizational capital and incremental product innovation.

CONCLUSION

Figure 1 shows the conceptual framework of this study. Based on this framework, knowledge exploration and knowledge exploitation are two types of knowledge utilization capability a firm can use to develop its desired product innovations. The bright and skilled employees can help the firm to search distant knowledge and develop the capability on knowledge
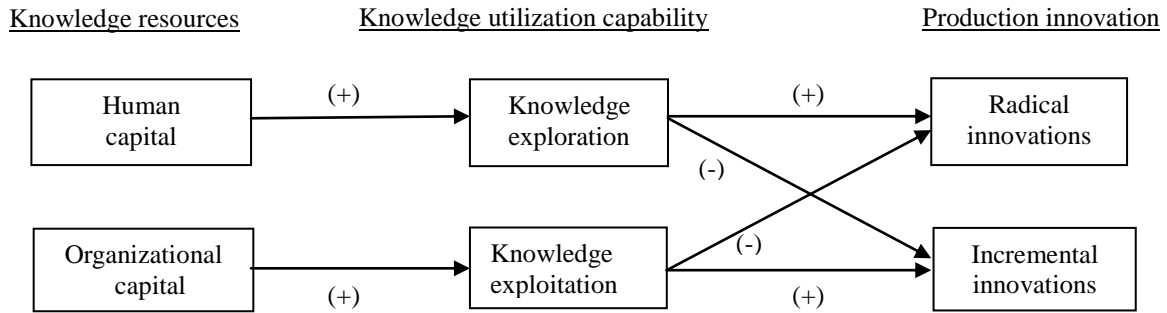
Figure 1 conceptual framework

exploration. Thus, a firm's efforts at hiring, training, and other human resources activities are important if it focus on knowledge exploration. To developing the capability on knowledge exploitation, the firm can focus on the institutionalized knowledge accumulated in and utilized through its patents, databases, structures, and systems. Although knowledge resources such as human capital and organizational capital are pivotal inputs in innovation process, their influence on product innovations are mediated by a firm's capability on knowledge exploration and exploitation. However both knowledge exploration and knowledge exploitation have opposing influence on incremental and radical innovations. Therefore, managers may need to combine high knowledge exploration with low knowledge exploitation to develop radical innovations, and vice versa to develop incremental innovations.

REFERENCES

[1] Alavi, M. & Leidner, D. E. (2001). Knowledge management and knowledge management systems: Conceptual foundations and research issues. *MIS Quarterly*, 25, 107-136.

[2] Atuahene- Gima, K. (2005). Resolving the capability-rigidity paradox in new product innovation. *Journal of marketing*, 69(4), 61-83.

[3] Baker, W. E. & Sinkula, J. M. (2007). Does market orientation facilitate balanced innovation programs? An organizational learning perspective. *Journal of Product Innovation Management*, 24, 316-334.

[4] Benner, M. J. & Tushman, M. L. (2003). Exploitation, exploration, and process management: The productivity dilemma revisited. *Academy of Management Review*, 28, 238-256.

[5] Brown, S.L. & Eisenhardt, K.M. (1995). Product development: Past research, present findings, and Future directions. *Academy of Management Review*, 20(2), 343-378.

[6] Chandy, R. K. & Tellis, G. J. (2000). The incumbent's curse: Incumbency, size and radical product innovation. *Journal of Marketing*, 64(3), 1-18.

[7] Daneels, E. (2002). The dynamics of product innovation and firm competences. *Strategic Management Journal*, 23, 1095-1121.

[8] Grant, R. M. (1996). Toward a knowledge-base theory of the firm. *Strategic Management Journal*, 17(10), 109-122.

[9] Hansen, M. T., Nohria, N. & Tierney, T. (1999). What's your strategy for managing knowledge. *Harvard Business Review* , 1-10.

[10] Hitt, M. A., Bierman, L., Uhlenbruck , K., Shimizu, K. & Kochhar, R. (2001). Direct and moderating effects of human capital on strategy and performance in professional service firms? A resourced- based perspective. Academy of Management Journal, 44(1), 13-28.

[11] Jones, G. B. & Schneider, W. J. (2006). Intelligence, human capital, and economic growth: An extreme-bounds analysis. *Journal of Economic Growth*, 11, 71-93.

[12] Katila, R. & Ahuja, G. (2002). Something old, something new: A longitudinal study of search behavior and new production introduction. *Academy of Management Journal*, 45, 1183-1194.

[13] Madanmohan, T. R. (2005). Incremental technical innovations and their determinants. *International Journal of Innovation Management*, 9(4), 481-510.

[14] March, J. G. (1991). Exploration and exploitation in organizational learning. *Organization Science*, 2,71-87.

[15] Nelson, R. & Winter, S. (1982). *An Evolutionary Theory of Economic Change*. Cambridge, M. A.: Belknap Press.

[16] Nerkar, A. & Roberts, P. W. (2004). Technological and product-market experience and the success of new product introductions in pharmaceutical industry. *Strategic Management Journal*, 25, 779-799.

[17] Rosenkopf, L. & Almeida, p. (2003). Overcoming local search through alliances and mobility. *Management Science*, 49, 751-766.

[18] Snell, S. & Dean, J.W., Jr. (1992). Integrated manufacturing and human resource management: A human capital perspective. *Academy of Management Journal,* 35, 467-504.

[19] Sorescu,A. B., Chandy, R.K. & Prabhu,J. C. (2003). Source and financial consequences of radical innovation: Insights from pharmaceuticals. *Journal of Marketing*, 67, 82-99.

[20] Subramaniam, M. & Youndt, M. A. (2005). The influence of intellectual capital on the types of innovative capabilities. *Academy of Management Journal*, 48(3), 450-463.

[21]Tushman, M. L. and Anderson, P. (1986). Technological discontinuties and organizational environments. *Administrative Science Quarterly*, 31, 439-465.

[22] Youndt, M. A., Subramaniam, M. & Snell, S.A. (2004). Intellectual capital profiles: An examination of investments and returns. *Journal of Management Studies*, 41, 335-362.

AN INTER-ORGANIZATIONAL KNOWLEDGE MANAGEMENT SYSTEM IN HEALTHCARE: BUILDING INNOVATIVE INFORMATION INFRASTRUCTURE FOR SOCIAL GOOD

Elliot B. Sloane, Drexel University, 3141 Chestnut Street, Philadelphia, PA 19104, USA
James P. Keller, ECRI Institute, 5200 Butler Pike, Plymouth Meeting, PA 19462, USA
Eric Sacks, ECRI Institute, 5200 Butler Pike, Plymouth Meeting, PA 19462, USA
Q B. Chung, Villanova University, 800 Lancaster Avenue, Villanova, PA 19085, USA

ABSTRACT

Patient safety and medical errors is a well-documented global problem, and in most healthcare settings, patient safety is a reactive, not proactive, activity because there are few comprehensive systems to resolve the problem. Knowledge management (KM) in the traditional sense has been confined to narrow and individual organizational settings since knowledge has been viewed as a proprietary resource that could enhance an organization's competitive advantage. However, the endemic failure to facilitate the flow of knowledge through organizational boundaries is causing risks of patient injuries and deaths. Since healthcare is in the realm of public good and as much as possible the risk of future incidents like the one at Johns Hopkins should be minimized, the traditional proprietary approach to KM, which primarily sought competitive advantage of individual organizations, should be rethought. This paper describes ECRI Institute's Alerts Tracker, a novel inter-organizational knowledge management system (KMS) designed by a nonprofit healthcare agency called ECRI Institute.

Keywords: Knowledge management, Knowledge management systems, Patient safety, Inter-organizational Knowledge management systems (IOKMS), Globalization

INTRODUCTION

Patient safety and medical errors is a well-documented global problem. It is being addressed by US agencies such as FDA, VA, NASA, AHRQ, and NIH, as well as WHO, PAHO, and multiple international government bodies. The problems are related to many causes, including training, complex error sequences, and device and drug failures and interactions. Problems caused by this latter group, which spans a wide technology base, could often be avoided or eliminated if users were alerted to, and acted on, known recalls, product updates, hazards, safety alerts, and other information resources. In most healthcare settings, patient safety is a reactive, not proactive, activity because there are few comprehensive systems to resolve the problem.

There are many facets of medical technologies that make this more difficult to solve than might be apparent on the surface. First, there are so many technologies—devices, disposables, drugs, biologic agents, computer systems, and hybrid implementations—that categorization of the available information about those technologies and matching that information to existing inventories is daunting. Second, any single hospital or healthcare institution has only a limited portfolio of the potentially-affected products (e.g., only some of the affected lot or serial

numbers for a specific model). Even if the hospital knows that it has, or had, an affected product model, it may be unable to pinpoint the configuration, serial, lot, or batch numbers or location of such items in the facility because this level of detail may not have been or is not able to be recorded in their inventories. Lastly, there is no standardized system that allows hospital staff to organize the information and manage the follow-up process to conclusion.

Knowledge management (KM) in the traditional sense has been confined to narrow and individual organizational settings since knowledge has been viewed as a proprietary resource that could enhance an organization's competitive advantage. However, in this situation, the endemic failure to facilitate the flow of knowledge through organizational boundaries is causing risks of patient injuries and deaths. As an example, in one high-profile incident at Johns Hopkins University Hospital in Baltimore, Maryland, two patients may have died from exposure to contaminated bronchoscopes that were part of an ongoing product recall. A recall notice had been sent to customers by the product's manufacturer. Unfortunately the recall was delivered to a receiving dock at Johns Hopkins and was not passed on the relevant users of the products in the bronchoscopy department. Since healthcare is in the realm of public good and as much as possible the risk of future incidents like the one at Johns Hopkins should be minimized, the traditional proprietary approach to KM, which primarily sought competitive advantage of individual organizations, should be rethought.

This paper describes ECRI Institute's Alerts Tracker, a novel inter-organizational knowledge management system (KMS) designed by a nonprofit healthcare agency called ECRI Institute. ECRI Institute provides a common core knowledge base of medical device and supply recalls, safety alerts, hazard reports, and related critical information that are organized around a universal nomenclature system. From each participant's point of view, ECRI Institute's KM is only a subset of their own proprietary KMS, since they can extend ECRI Institute's information with locally relevant information when and as needed. The main contribution of this paper is two-fold. First, we identify the viability of inter-organizational KMS (IOKMS) and how it enhances the value positions of multiple parties, both collaborators and competitors, which deviates from the traditional notion of KMS intended for competitive advantages of single organizations. Second, through a comparative analysis of major alternatives that can be viewed viable for improving the social good in the domain of healthcare technology risks, we show the KMS has substantial unique advantages.

The organization of the paper is as follows. The next section describes the context of the problem, namely patient safety and medical technology hazards, from the perspective of KM. It will be followed by a general overview of KM and KMS and its applications to the medical field. We provide a detailed description of ECRI Institute's Alerts Tracker, an IOKMS, followed by a comparative analysis of alternatives. We conclude the paper with discussions about the worldwide implications of this novel approach as well as a potential extension to other domains of the public good. Also discussed are the benefits to the individual organizations in this emerging form of 'cooptition' that has heretofore only been applied in commercial settings.

CONTEXT OF THE PROBLEM

In a 1999 Institute of Medicine (IOM) report, patient safety was identified as a very serious national issue. According to the publication, up to 90,000 patients in the US are injured or killed each year by medical errors. There have been two published updates on this subject by the IOM since then, and a federal agency, the Agency for Healthcare Research and Quality (AHRQ), was formed in 1999, in part, to address this issue. AHRQ has held numerous conferences and commissioned extensive research since then, and has found that the problem of patient safety and medical errors is, in fact, a very serious international issue. AHRQ has provided millions of dollars in research funding to begin to organize the body of knowledge about patient safety and medical errors in order to eliminate the problem. The U.S. Food and Drug Administration (FDA) has stated that the most prominent causes of medical errors is "user error"—i.e., incorrect application of a medical technology by a doctor, nurse, or patient. It also requires that healthcare organizations track the status of certain implantable devices and high risk equipment so that newly reported risks or problems with these devices can be quickly resolved for patients that have these devices.

Many other organizations have created policies or departments to address medical errors. For example, the US Veterans Hospital Administration (VHA) has a center for patient safety (www.patientsafety.org/gov) that develops policies and trains its national staff. The US Military has a similar program. The Joint Commission for the Accreditation of Healthcare Organizations (JCAHO), the major accrediting body for hospitals, nursing homes, and home care agencies, has incorporated patient safety and medical error reduction requirements in all of its inspection and training programs. It has established specific National Patient Safety goals that have addressed use of technologies like infusion pumps and patient monitors. JCAHO also requires that hospitals identify and implement processes for monitoring and acting on equipment hazard notices and recalls. Table 1 lists many of the related international agencies and web sites that focus on this problem.

As will be discussed in detail shortly, the incomplete, incorrect, and inefficient distribution of information about medical device hazards, defects, limitations, and recalls is one very well documented cause for user errors. For example, unintended free-flow of intravenous medication has been responsible for numerous medication overdoses and even patient deaths. These incidents have led to infusion pump redesign and widespread publicity about infusion pump-related overdoses. However, incidents have continued to occur because many institutions do not have systems in place to effectively monitor for medical product hazards and recalls or to disseminate such information to appropriate staff. This article describes how the comprehensive reporting of medical technology hazards and recalls can be accomplished to overcome these problems.

The current article describes a novel KMS for medical device recall and hazard knowledge, which allows individual hospitals to identify risks and both manage and document the remediation process. Traditionally, knowledge management (KM) is understood as an attempt to discover, record, disseminate, and re-use organizational memory to enhance the efficiency and effectiveness of the performance of the organization. In order to support this activity, a KMS may be designed or purchased by an organization. This paper describes a Web-based medical

device hazard and recall KMS for hospital use that not only supports an organization's unique experience and needs, but also enables international sharing of discovered knowledge that can prevent critically important patient injuries and deaths in other hospitals.

Table 1: Patient Safety Issues in Reports from Various Healthcare Organizations

Source	Year	Report Title	Conclusions
Joint Commission for the Accreditation of Healthcare Organizations (JCAHO)	2005	Setting the Standard: The Joint Commission & Health Care Safety and Quality	Almost 50 percent of Joint Commission standards are directly related to safety, addressing such issues as medication use, infection control, surgery and anesthesia, transfusions, restraint and seclusion, staffing and staff competence, fire safety, medical equipment, emergency management, management of hazards and recalls, and security.
World Health Organization (WHO)	2004	World Alliance for Patient Safety	The fifty-fifth World Health Assembly in 2002 adopted a resolution urging countries to pay the closest possible attention to the problem of patient safety, and strengthen safety and monitoring systems.
Institute of Medicine (IOM)	1999	To Err is Human; Building a Safer Health System	Estimates of between 44,000 and 98,000 Americans are killed by medical errors each year. Goal of 50% error reduction in the next 5 years..
Institute of Medicine (IOM)	2001	Crossing the Quality Chasm	…Key steps to promote evidence-based practice and strengthen clinical information systems.
Institute of Medicine (IOM)	2004	Patient Safety: Achieving a New Standard for Care	To achieve this, a new health care delivery system is needed a system that both prevents errors from occurring, and learns from them when they do occur.
Institute of Safe Medical Practices (ISMP)	2000	Discussion Paper on Adverse Event and Error Reporting in Healthcare	… on Adverse Event and Error Reporting in Healthcare
Veteran's Healthcare Administration (VA)	2005	Creating a Culture of Safety	It has been reported in the medical literature that as many as 180,000 deaths occur in the United States each year due to errors in medical care, many of which are preventable.
Emergency Care Research Institute (ECRI)	2005	ECRI Patient Safety Center	As part of its mission to promote the highest standards of safety and quality in healthcare, ECRI provides public access to many of its patient safety resources.
Agency for Healthcare Research and Quality (AHRQ)	2005	Patient Safety Network	PSNet is the first comprehensive effort to help health care providers, administrators, and consumers learn about all aspects of patient safety.

KM AND KMS IN MEDICAL FIELD

The term "knowledge management" (KM) portrays an explicit, organizational-level effort that stems from the recognition that an organization's knowledge possessed by its members is a determining factor of its competitiveness and success. This notion goes farther than simply acknowledging that knowledge is a crucial organizational resource in addition to the traditional

ones—e.g., capital, labor, and land. Given the strategic significance, it is not surprising that firms allocate sizeable amount of budget to KM initiatives, and create KM-oriented top executive positions such as CKO (chief knowledge officer), while software and consulting practices specializing in KM are flourishing.

However, KM means many different things to different people. Perhaps recognizing knowledge as a key source of competitive advantages is where the general agreement stops among the various schools of thought claiming expertise in KM. In the field of KM, the diversity in philosophy and orientation is truly remarkable. A comprehensive review of the vast KM literature helps to identify a few competing themes of KM research. The first theme is concerned with what can be termed as the 'Life Cycle of Organizational Knowledge' (roughly put, creation–codification–dissemination–internalization–reuse). Depending on which phase of the lifecycle one is interested in, this thematic group is divided into various smaller camps although there is nothing small about the size of their claims or produced research output. There is a visible contrast between the Eastern KM approach and that of the West. The latter puts much emphasis on the reuse of *explicit* knowledge (hence the short-term gains are the main driver), while the former is interested in creation of *tacit* knowledge through 'nurturing' culture of knowledge community that rewards long-term, but less tangible/measurable knowledge output.

The second theme can be called the school of 'Organizational Learning' whose legacy dates much further back than the coinage of KM. Although knowledge is commonly placed on the trajectory of the 'data-information-knowledge' progression, knowledge poses a much more rigorous requirement for "duplicated possession" than data and information do. That is, the only and true way to disseminate knowledge within an organization is for the members to "learn," whereas data and information can be disseminated for immediate use. Therefore, the organizational learning school enjoys a solid legitimacy in KM research.

The third theme can be called the 'Return of AI.' This group appeals to the successful track record of the 1980s' artificial intelligence (AI) applications and knowledge-based systems (KBS) encompassing knowledge acquisition, knowledge representation, expert systems, and machine learning. While it is true that the subject matter of both AI/KBS and KM is 'knowledge,' there are two main differences: (1) the AI/KBS approach delivered tangible results with significant impact virtually in all areas of business, engineering, science, and medicine, while KM is still trying to shake off the bad names like 'fad' and 'hype'; on the other hand, (2) a KBS has a tendency to be effective in a narrowly defined domain, while KM has no such limitation in the scope of applicability.

The fourth school of thought represents a practical mindset that regards the internally developed knowledge and know-how as organizational resources deserving protection and leverage from the 'Intellectual Property Rights' perspective. It is notable that the main drivers of this theme are predominantly the practitioners in knowledge-intensive industries such as consulting, R&D, and legal counsel.

Finally, the fifth major theme is the one that puts a slightly negative spin on KM, and can be referred to as the 'KM Misnomer Movement.' Since one can accept—even so reluctantly—the notion that KM means different things to different people, this quiet voice of disapproval has not

taken the center stage of KM literature. However, those trained in the discipline of management can easily sympathize with such serious questions as "What is knowledge and can it be managed?", or at least become curious why some practitioners show downright disapproval of KM with such claims as "Forget KM; Back to information!". In fact, Nonaka *et al.* agree: "It is our strong conviction that knowledge cannot be managed; only enabled".

This brief survey of the field of KM reveals two fundamental facts about KM as we know of it today. First, there is hardly a consensus about what KM actually means, and consequently, what organizations do to implement KM tends to depend on which view or views of KM they subscribe to. Second, regardless of the definition of KM adopted from the above popular choices, the underlying principle is to advance the organizational competitiveness by taking advantage of its knowledge assets.

When an organization develops a KMS to implement its beliefs about KM, due to the fluidity and variety, the resulting product may take a wide range of forms. More fundamentally, Alavi and Leidner posit that the shape of a KMS will be dictated by how knowledge is viewed. For instance, if knowledge is considered an extension of data and information, a KMS will not look radically different from the traditional IS. On the other hand, if knowledge is understood strictly as the "state of mind," implementation of KMS will lean toward providing access to the knowledge sources rather than knowledge itself. Discussions of KM and KMS in the context of gathering, storing, and transferring knowledge stem from the view that knowledge can be understood as "object," whereas the "process" view of knowledge leads KMS implementations to facilitation of the flow of knowledge within the organization by strengthening of links among the knowledge sources with the help of information technology (IT).

A comprehensive of view of various KMS architectures is presented as Figure 1 (adopted from Alavi and Leidner). The all-encompassing technological enabler that serves as the foundational infrastructure is the corporate intranets. Upon that base, the next layer of building blocks is laid composed of various groupware technologies for communication and collaboration among the knowledge workers. Employing the groupware is intended primarily to facilitate externalization and socialization of knowledge. The top layer of the KMS architecture supports various knowledge activities—creation, storage and retrieval, transfer, and application—using a range of available IT.

Healthcare organizations have adopted the concepts of knowledge management in a variety of ways. Most have focused on identifying processes in need of improvement that have specific patient safety concerns. The most prominent application has to do with medication management (i.e., making sure the right medications are delivered to the right patients, at the right dose, at the right site, at the right time). For example, many hospitals have established automated order entry systems to help minimize errors associated with misrecording orders for high risk medications. These systems also establish rules to prevent delivery medications to the same patient that are incompatible with one another. These rules are routinely checked and updated, and then communicated to staff through the automated system.

Applicability of the current view to the problem: Figure 2 illustrates one example of how a contemporary healthcare enterprise views and manages its KMS assets (Glaser 2004). In this

idealized model, a review committee "vets" or sanctions the rules in accordance with institutional policy to ensure optimal outcomes and alignment with strategic goals. Those rules are then implemented within systems by knowledge editors. The clinicians (physicians and nurses) use the KMS to support their efforts, and the medical records of each event is stored in an analytic discovery data warehouse for ongoing quality assessment and improvement activities.
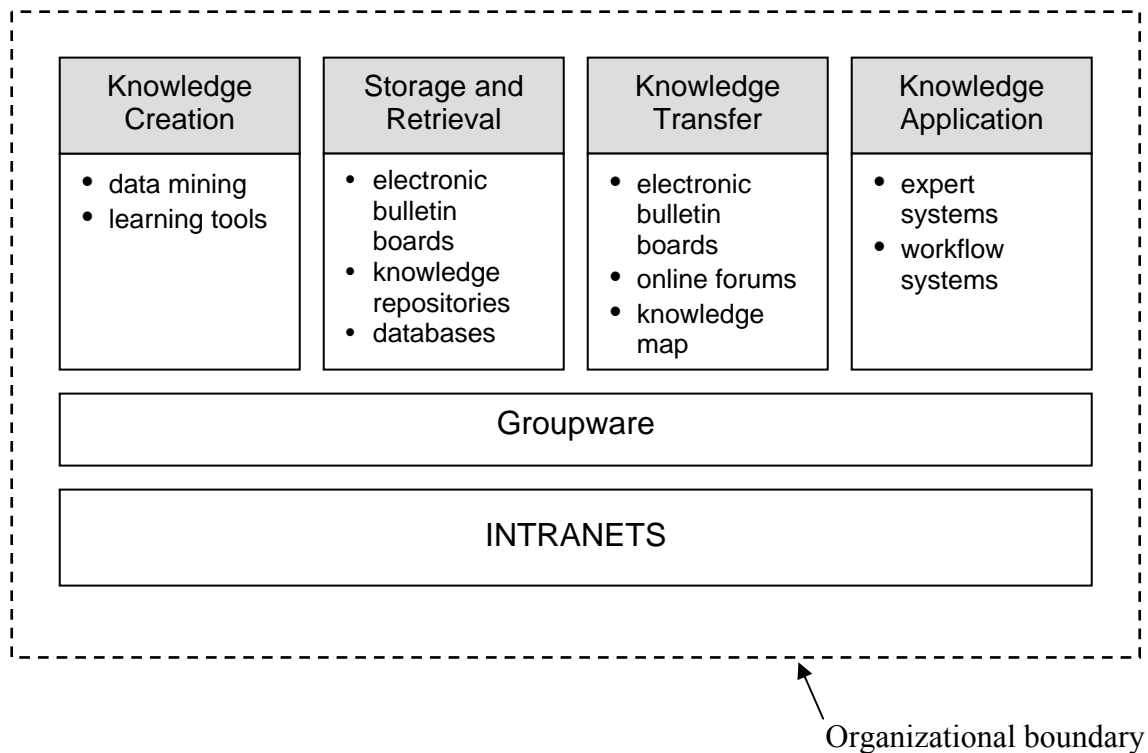
Figure 1: A comprehensive view of KMS

While most KM applications are implemented within particular individual organizations, and KM vendors' offerings tend to be confined within individual organizational boundaries due to the self-fulfilling nature of KM initiatives, it is worth noting that there have been reports of inter-organizational KM practices. Particularly, when firms in knowledge-intensive industries come together (referred to as strategic alliances, virtual organizations, or knowledge networks) with an explicit purpose of developing new ideas, products, or processes and of jointly learning from one another, they form a cooperative KM environment that fosters synergistic discoveries. Holmqvist (1999) documents a case of inter-organizational knowledge creation in the context of "imaginary" organizations in a variant form of Nonaka's framework (1994) based on socialization, articulation, combination, and internalization. According to a recent empirical study of inter-organizational KM initiatives among law firms in Europe and their impact on the participating firms' use of IT (Khandelwal & Gottschalk 2003), the technological elements have been found to be a more influential factor for inter-organizational knowledge flows than the degree of trust between the firms involved. On the other hand, from the perspective of the knowledge sharing and re-use, Majchrzak et al. (2000) obtained somewhat limited evidence of computer-mediated collaboration among design teams that were inter-organizational, distributed, and virtual in nature. These cases may be viewed as anecdotal evidence of inter-organizational

KM programs. However, when one unveils the motivational factors behind such initiatives, and also when the outcome of such inter-organizational KM initiatives are evaluated, it becomes clear that thus far, the primary driving force of KM, whether intra-organizational or inter-organizational, has been the competitive advantage of the organization, which ultimately seeks to benefit the self.
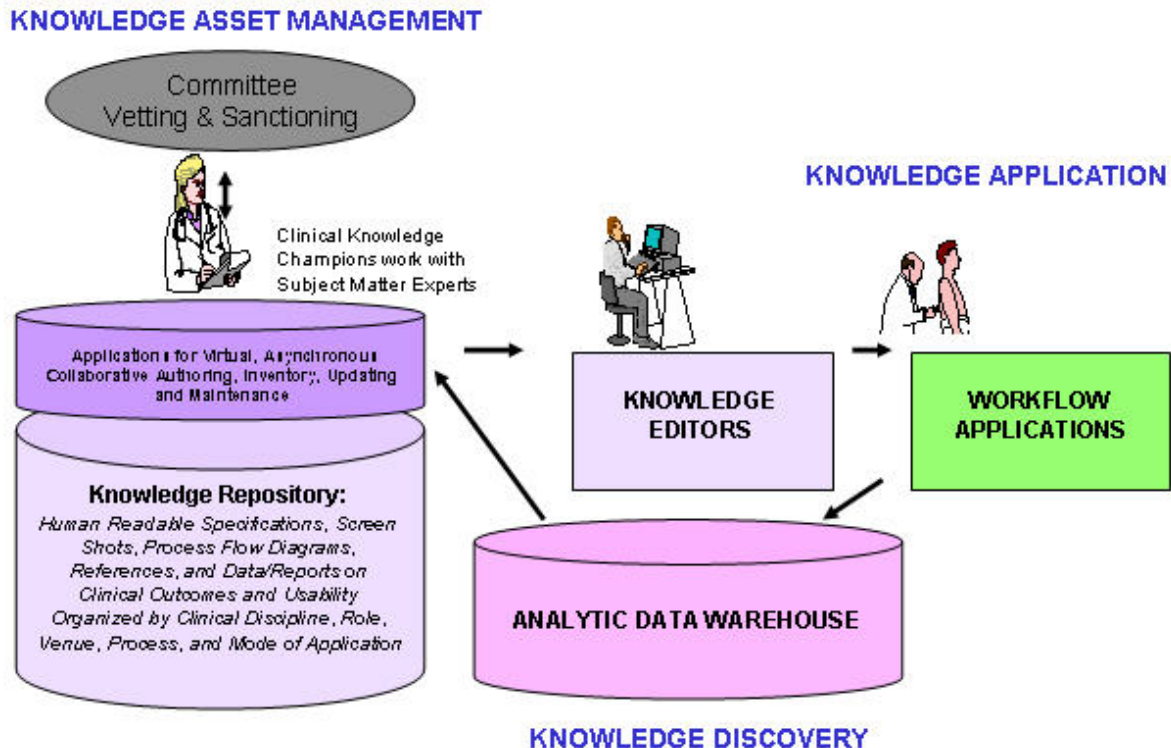
Figure 2. Partner's Healthcare Knowledge Management Framework (Glaser, 2004)

Contemporary application of knowledge management in the medical field takes many forms, including Expert Systems, Artificial Intelligence, and Decision Support Systems (DSS). Few clinical software systems, such as computerized prescriber order entry systems, include KMS rules directly, but instead rely on each hospital or clinical user to select, install, and maintain the system in accordance with internal policies and work in conjunction with general rule sets purchased by hospitals for the systems from third-party providers like First Databank. In a great many instances, such systems are based on professional bodies of knowledge, commonly referred to with terms like clinical practice guidelines and pharmacological formularies. The guidelines practice guidelines have been organized at a Website supported by and a selected number of centers of excellence.

The pharmaceutical formularies are variously created, maintained, and referenced by insurance companies and government agencies around the world as a basis for reimbursement decisions (e.g., Blue Cross, Medicare, and Veteran's Administration.) Safety aspects of the formularies, such as adverse drug interactions, are embodied in several hardware and software systems (e.g., Alaris, Partner's Healthcare, and the Arden nomenclature.) Ultimately, in most countries, the

final practice determination lies with individual physicians, based on her/his knowledge, although national health system reimbursement policies exert a strong external influence by constraining reimbursement.

Because clinical decisions are managed locally, patients with similar conditions or histories may well be treated differently in different hospitals or countries. This is giving rise to large numbers of computerized KMS systems, with an enormous number of rules. For example, Partner's Healthcare in Boston, MA, identified over 20,000 decision rules in a recent survey of its relatively small number of facilities. Because the majority of such KMS rules are locally created and maintained, such rules can be seen as an expression–or codification of–the clinical wisdom within their institutions. This use fits the common model of a KMS applied for strategic advantage in their market, because it attempts to ensure more effective, reliable, and/or safe medical care than other competing institutions can provide. Furthermore, a KMS like Partners's can constitute a form liability protection, serving to reinforce and document desired practices.

There is an inherent inefficient—and potentially unethical or immoral—aspect to such proprietary applications of KMS in healthcare, however. Consider the ramifications if, for example, a physician observes a new unsafe practice or drug interaction that is not evident to outsiders. Until or unless that physician can find a way to rapidly validate and publish such a finding, this critical knowledge will necessarily be limited to her/his business enterprise. As in the case of the recent criticism and recall of Vioxx due to cardiac risks, the inability to effectively share such knowledge has been blamed for perhaps many tens of thousands of deaths. In a similar incident, Guidant initiated a recall of thousands of implantable cardioverter defibrillators for a potentially fatal problem that it had known about for several years. The delay in establishing the recall had to do with, in part, what was believed to be a low failure rate for the defibrillator.

In order to eliminate or mitigate such serious risks to patients, a novel potential KMS architecture can be considered, in which human-risk-related knowledge might be shared via the Internet with other healthcare providers instead of being hoarded for internal operational and strategic benefit. In such an architecture, doctors, nurses, clinical engineers, or healthcare administrators with critically important safety knowledge may elect to share it in exchange for access to similar knowledge-sharing from the community. The benefit of this architecture is that safety-related knowledge is leveraged on behalf of the public good, leaving the individual healthcare institutions to their own means to compete on other dimensions of customer service or efficiency— e.g., by application of Intranet KMS tools as depicted in Figure 1.

ECRI INSTITUTE'S KMS

ECRI Institute is a nonprofit health services research agency with a mission of improving the safety, quality, and cost-effectiveness of healthcare. ECRI Institute accomplishes its mission through development and distribution of high quality information addressing healthcare technology safety, effectiveness, and cost issues. ECRI Institute maintains a presence in hospitals and health systems through regular consultation with client hospitals, on-site, by telephone, and increasingly by electronic correspondence and interactive services. ECRI Institute is a

collaborating center of the World Health Organization and has been designated as an evidence-based practice center by the federal Agency for Healthcare Research and Quality (AHRQ).

ECRI Institute launched its knowledge management system, *Alerts Tracker*, in 2003. *Alerts Tracker* is a Web-based system that supports client hospitals with management of the medical device hazard reports and recall notices issued in ECRI Institute's *Health Devices Alerts (HDA)* content product. The *Alerts Tracker* system provides automated e-mail notification of each new *Alert* to relevant personnel throughout the hospital based on a profile for each user that reflects clinical specialties and administrative roles. The e-mail notices provide basic information about new alerts and navigation to a secure Web site where recipients can read the details of each report and log responses regarding whether the hospital has affected product and, if so, what has been done to resolve the problem.

In order to show the Inter-organizational nature of the *Alerts Tracker* system, ECRI Institute's KMS, the following features are noteworthy.

The foundation of the *Alerts Tracker* system is the high quality *HDA* content that it delivers. With hundreds of hospitals using this information resource, ECRI Institute regularly receives leads, feedback, and other input from customers that results in refinement of the information available to all users. Such input includes:

- Submissions of new alerts that have not yet been reported in the system;
- Reports of problems and accidents that have occurred in the healthcare setting;
- Corrections to information presented in the system; and
- Questions regarding ambiguities in the presented information.

In the case of recalls, *HDA* editorial and research personnel verify the details of each new recall with the manufacturers and distributors involved and, with the assistance of ECRI clinical and technical staff, clarify issues that may be confusing in the original manufacturer notifications or FDA's coverage in its weekly *Enforcement Report*, providing additional commentary and recommendations as needed. To support hospitals in determining whether each recall affects their inventory and which departments should be notified, ECRI also indexes each alert to the *Universal Medical Device Nomenclature System (UMDNS)*, ECRI's *Health Devices Sourcebase,* and a list of clinical specialties and other healthcare professional classifications like risk management and clinical engineering.

In addition, ECRI's *Health Devices* and *Accident and Forensic Investigations* staff routinely issues Hazard Reports through the Alerts Tracker as result of their investigation of incidents at hospitals and other healthcare facilities. In some cases, Hazard Reports cover problems or design defects with specific products. Many Hazard Reports also cover typical and/or high severity user errors that would never result in any product recall.

Leads and input from client hospitals improves the *Health Devices Alerts* content in terms of accuracy, completeness, and timeliness. The capacity of this information refinement process is far beyond that which could be accomplished by any individual hospital or health system working independently.

The Inter-organizational KMS is also Web-accessible. Alerts Tracker e-mail notices provide users with basic information about each new *Alert* (product type or supplier & name, basic description of the problem, priority) and navigation to the full-text report on the *Alerts Tracker* Web site. Once logged into the Web site, users record information about how the alerts impact the hospital, indicating which are and are not resolved. The system is configured in the application service provider (ASP) model and as such, information recorded by users is transmitted over the Internet and stored on servers at ECRI Institute's data center. The Alerts Tracker Web site operates in a secured socket layer (SSL) so each transaction is encrypted during transmission over the Internet.

In order to facilitate shared use of proprietary knowledge sources, information logged by each user is time/date stamped and associated with the recording individual by unique user names and passwords (using Active Directories authentication). User accounts are configured into organizational hierarchies which facilitates centralized reporting by hospitals and health systems administrative staff while keeping the information confidential and inaccessible to other client institutions.

COMPARATIVE ADVANTAGES OF INTER-ORGANIZATIONAL KMS

Adopting an interactive IOKMS like Alerts Tracker that includes highly refined content that can be filtered and delivered to the right types of personnel throughout the healthcare institution saves time, improves participation by clinical staff, and maximizes the risk reduction benefit of their effort.

The combination of quality content and customizable interactive tools is necessary because medical devices and supplies present a particularly heavy hazard and recall management burden because of the sheer numbers of affected products and the challenges of their distribution throughout the healthcare facility.

Inventory databases used to manage maintenance of durable medical equipment typically contain significant levels of obsolete or "dirty" data. Bad data arises both from informal product naming practices (e.g., product name abbreviations) and from changes in product-line representation over time. For example, many defibrillators produced by Hewlett-Packard Company Medical Products Group remain in use. Hewlett Packard sold that business to Agilent Technologies Inc. several years ago. Agilent was subsequently acquired by Philips Medical Systems for its Cardiac & Monitoring Systems Division. Such company histories and the inconsistent use of product name abbreviations limit the ability of hospitals to use their equipment databases for rigorous identification of affected product.

Furthermore, most consumable supplies like catheters are not logged in inventories with specific identifiers like lot numbers or expiration dates. As such, hospitals frequently must rely on the collective knowledge of clinical staff in various departments to help isolate recalled products. This burden is substantially reduced by use of alert content that specifies which clinical and professional specialties are likely affected.

For example, a hazard report on linear accelerators would be routed to designated clinicians involved in nuclear medicine as well as supporting engineering staff. Similarly, a recall of contaminated cardiac catheters would be sent to clinicians in the cardiac catheterization laboratory, the infection control officer, and central supply staff.

DISCUSSIONS AND SUMMARY

KMS are inherently beneficial to the Medical Device Hazard and Recall Management process in healthcare facilities for a few basic reasons.

First, information capture is centralized. Automated aggregation of input from subject matter experts throughout an organization's healthcare facility or facilities supports centralized, real-time monitoring and reporting. Centralized monitoring and reporting supports risk management, quality improvement, and accreditation activities. The automated nature of information aggregation provides labor savings by comparison to the administrative and often paper-based efforts that would otherwise be necessary to compile institution-wide reports.

Also, the burden on individual users is minimized. Automated distribution further reduces the labor involved in hazard and recall management. Basing automated distribution on individual user profiles allows filtering of content to avoid overwhelming individual users with hazards and recalls that do not apply to their areas of responsibility. By tailoring content for individual users, duplication of effort is minimized and users are more likely to contribute.

An Inter-organizational Knowledge Management System, such as ECRI Institute's KMS, adds further benefits in the form of highly refined content by aggregating the feedback of a community of member healthcare organizations. First set of benefits are the timeliness and completeness. New hazards and recall notices are added as soon as one member notices that something is missing and submits it to the system. The other key benefit is quality control. Corrections based on errors identified by one member and verified by the system host are shared by all users. However, the most significant of all is that fact that hundreds healthcare facilities are acting on known dangers with medical devices every day, before the devices are exposed to their patients, staff, and visitors.

REFERENCES

References are abbreviated in the current proceedings. While the author-year format may provide some indication of the sources cited, complete citation information will be provided upon request to the corresponding author via q.chung@villanova.edu.

SOCIAL NETWORKING TECHNOLOGIES IN KNOWLEDGE MANAGEMENT

Edward T. Chen, University of Massachusetts Lowell
Operations and Information Systems Department, Lowell, MA 01854
edward_chen@uml.edu, (978) 934-2756

ABSTRACT

This paper discusses the Internet phenomenon known as Web 2.0. It explores Internet use, Internet users, and the continuous improvements being made to the Internet. The purpose of this paper is to explain the impact that social networking has on the modern enterprise. Each application is also applied to a practical business setting. The benefits and challenges of each application are discussed and examples of organizations that are implementing Web 2.0 strategies are presented. Some limitations and concerns of Web 2.0 are discussed. The paper concludes with an examination of the implications of Web 2.0 on companies and their business and marketing strategies.

Keywords: Knowledge management, Web 2.0, social networking, collaboration

INTRODUCTION

Today's knowledge management (KM) systems focus on centralized sets of repositories, organized around established business processes. The current knowledge management systems are expensive to implement and the long-term commitment of the major resources of their deployment, maintenance, and daily operation can be seen as a huge burden. Consequently, even customized solutions end up going unused, with the knowledge workers running these custom KM solutions not having the information technology (IT) tools to provide support for their responsibilities. Based on these underutilized KM systems, the continuing evolution of Web 2.0 is providing a new KM solution, a collaboration based solution [12]. Social networking technologies provide immediate solutions to the large investments for the deployment, maintenance, and daily operations for today's KM systems [3] [7]. It is time for organizations to start looking at tomorrow's knowledge management solution and realize this new solution is a more efficient and effective model for today's enterprise knowledge management systems. Each time a new system is implemented; large investments into systems that have promised automation and seamless integration to share knowledge across the organization rarely become a reality [8][16] .

To understand the most fundamental aspects of knowledge management, one must first understand the process of knowledge acquisition, the use of intellectual property and the use of non-material assets. It is the knowledge within the organization that is the basis of an organization's development and allows them to find solutions to business problems. The knowledge management system becomes an essential tool for all the organizations actions, with the goal being that decisions can be made quicker and are justified and strengthened by the

knowledge within the system itself. Using these systems to have knowledge about clients, similar to a customer resource management system also increases the level of success in providing them with the best solutions. In proving solutions to clients, the knowledge that is captured also allows the knowledge system to help drive innovation. The learning organizational culture is a requirement for the constant changes in business processes and management practices that are driven by staying focused on the KM system and continuing to improve that knowledgebase [10]. Web 2.0 social networking (SN) technologies provide organizations with a set of tools to facilitate the knowledge acquisition, transformation, and sharing. Employees using SN to interact with other coworkers and teams will create an environment where information is exchanged with ease. SN eventually promotes a learning organizational culture that engages these people, connects information, and establishes strong relationships.

The knowledge management system should help maintain continuous innovations that lead to the creation of new goods and/or services and establish new business processes. Knowledge management is a solution that requires organizational, human and technological resources to provide the assets for the system. Choosing not to focus on any of these aspects will many times lead to failure of the system. It is the values of the organization which is the general problem in the realm of knowledge. And more accurately it is the human component that will determine the level of success of knowledge management systems [10]. Building trust across the organization so everyone trusts the KM and the solutions it will provide. Trust in the people that help create the organization is important and a lack of it is one of the biggest reasons of failure from human aspect of knowledge management. Once organization implements Web 2.0 and SN tools, talent and expertise can be retained through portal, networking and relationship building.

The purpose of this paper is to explain the impact that social networking has on the modern enterprise; particularly, when it comes to collaboration and knowledge sharing. The growth trajectory of Web 2.0 software such as social networking, blogs, tags, RSS feeds, wikis, YouTube videos and widgets are presented and each component is outlined in detail. Each application is also applied to a practical business setting. The benefits and challenges of each application are discussed and examples of organizations that are implementing Web 2.0 strategies are presented. Some limitations and concerns of Web 2.0 are discussed. The paper concludes with an examination of the implications of Web 2.0 on companies and their business and marketing strategies.

WEB 2.0

Web 2.0 is defined in many ways by a variety of different sources. The term Web 2.0 describes the transformation of websites from silo information sources to interlinked computing platforms [15]. Traditionally, websites presented static information that was rarely updated. Companies and organizations published information on the web and users consumed what was offered. There was no ability to interact with others on the web. However, the emergence of Web 2.0 has transformed the way the web is used, managed and developed.

Web 2.0 allows for a richer user experience. It embodies interactive functionalities such as social networking, blogs, tags, RSS feeds, wikis, YouTube videos and widgets. These applications enable users to become active participants in the web. They are no longer forced to passively

consume the information available. They can contribute to and improve content on websites such as Wikipedia in real time. They can develop social networks with other users through platforms such as MySpace, Facebook and LinkedIn. They can share life events through YouTube and they can publish their own content on blogs.

Web 2.0 allows users to share information, opinions, and thoughts through blogs. It enables users to improve free source software and redistribute it for free. According to Tim O'Reilly, founder and CEO of Sebastopol-based O'Reilly Media, the company that coined the phrase Web 2.0, "the heart of Web2.0 is the community – building collective intelligence from the mass of people that you can reach and interact and hear from – like customers which make this community" [15].
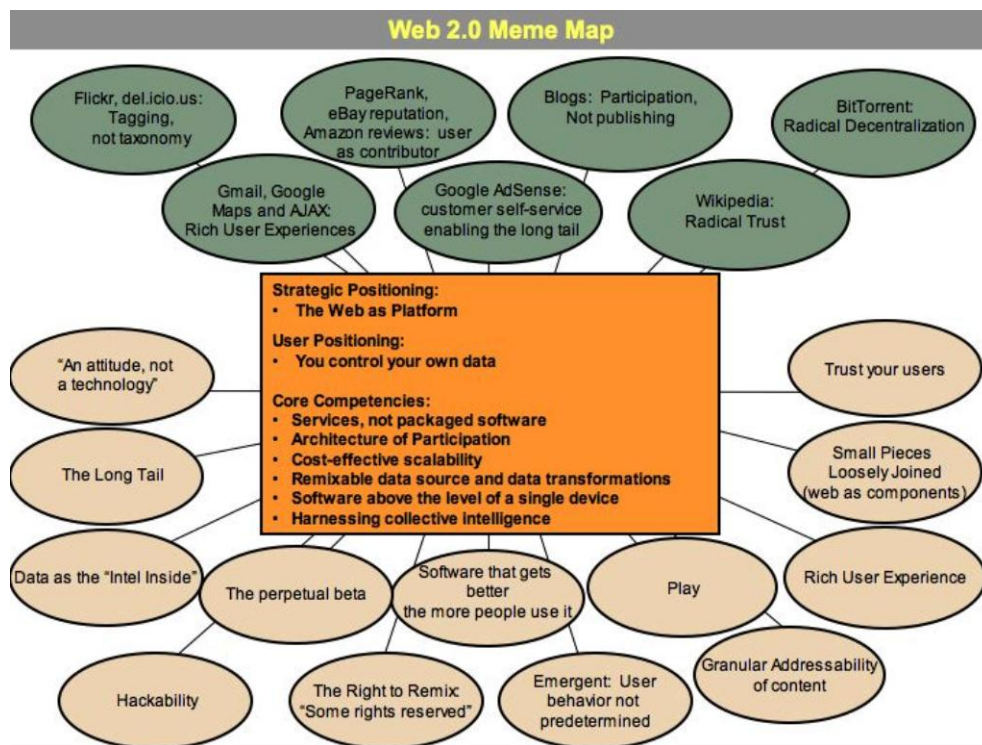
Figure 1. Web 2.0 Meme Map (Source: [15])

In addition to transforming the user experience, Web 2.0 is forcing companies to consider how they can take advantage of this groundswell of technological advancement [3]. Forrester Research reports that 29 percent of the U.S. population watches user generated videos on the web. Twenty-five percent read blogs, visit social networking websites such as Facebook and MySpace, and read reviews and ratings on line at least once a month. People are spending more of their time on-line than watching TV, reading newspapers, and listening to the radio. People ages 16-24 years old consume the least amount of television. Eighty three percent spend their time on-line playing games, downloading music, using instant messenger (IM), participating in social networks, and inhibiting virtual worlds. They are increasingly on the move. Ninety five percent of them own mobile phones [1]. Moreover, wireless hand held devices like BlackBerrys and personal digital computers (PDA) enable people to log on to the Internet from remote locations anywhere in the world at any time.

Figure 1 outlines the components of Web 2.0. It is built on the cornerstone concept of the web as a platform where the user controls the data. The core competencies transform software packages into services, a participative architecture, cost-effective scalability, interchangeable data sources and data transformations, software above the level of a single device, and the harnessing collective intelligence. On the periphery, Web 2.0 focuses providing rich user experiences where open source functionalities require trust and decentralization. Tagging is emphasized in place of taxonomy and user participation and contributions are promoted in place of traditional web publishing [15].

SOCIAL NETWORKING

Social networking (SN) builds web-based communities. SN software gives web users the ability to create profiles that foster interaction between groups of people based on interests and expertise. Typical SN applications include blogs, wikis, bookmarking/tagging, RSS feeds, and mashups. A mashup is a web page or application that combines data or functionality from two or more external sources to create a new service [6]. The result is typically a new and distinct Web service that was not originally provided by either source. Hence, SN software includes majority of the Web 2.0 technologies. Figure 2 indicates the connections of a typical SN.

SN complements traditional working practices by creating opportunities for extending sales, marketing, recruitment, research, and technical support. SNs can be leveraged as a customer relationship management tool for companies selling products and services. Using SN, these companies have been able to drive traffic to their own online sites while encouraging their consumers and clients to have discussions on how to improve or change products or services. As SN applications become integral to an organization's activities, they achieve legitimacy and value that puts at the same level as enterprise applications [11][18].

The need to create, acquire, store, organize, search, filter and visualize information for business purposes will only increase in coming years. Therefore, despite initial trepidation, enterprises; especially those with highly skilled employees working in remote locations, are coming to the realization that SN tools can help build a corporate culture in which knowledge is quickly located and shared [11].

SOCIAL NETWORKING AND KNOWLEDGE MANAGEMENT

Knowledge Management (KM) is a set of tools and processes companies use to create, track and share intellectual assets. The first wave of KM involved digitizing and tracking documents using tools like content management systems. It quickly became clear that it was too hard to share those documents, so companies adopted collaboration tools. Organizations are already actively leveraging the power of social networks to find new business opportunities, but SN tools show clear potential for improving collaboration and knowledge sharing within organizations. With social networks, companies are extending KM to make it easier to connect employees and information [8] [16] [18]. SN can be effectively used for finding expertise quickly and easily, particularly for people working remotely who feel part of the broader community with the use of SN [4] [14]. Since businesses operate globally, social networks can make it easier to keep in

touch with contacts around the world. Additionally, employees are typically familiar with SN since they use them outside of work [8] [11]. The following section of this paper addresses the business advantages of SN in many perspectives.

Existing business tools for knowledge sharing and collaboration primarily consist of email, work productivity desktop applications and portals. These tools are very structured and rigid in their set up and interaction, and do not provide a free-form medium for users to leave their impressions and opinions behind in the way that SN applications do. Bookmarking/tagging and other SN tools help bring order to the abundance of information that employees have to sift through [5][9].

SN helps people find and connect to co-workers through user profiles, expert search, and social graphs—visual maps of an employee's connections with co-workers. This makes it easier to stay in touch with a greater number of people than would be possible with one-to-one interactions. SN also helps workers find content and people relevant to their work, share information easily, and offer insights to each other on a continuous basis. For instance, experienced senior staff members can offer insights to junior staff members in small doses and in a casual style [5].

Many organizations are still divisionally segmented. SN can bridge groups so they can see what is going on outside their own area. For instance, an employee whom reads blogs outside of his/her own business group can understand the bigger picture of what is happening in the organization [5] [11].

If used properly, SN lets firms accelerate business by bringing faster response time to all facets of the business. In order to support growth, firms need ways of expediting innovation and SN tools supports the agility needed in today's economy [11]. SN tools bring people from different locations or business functions to participate in solving problems or creating innovation. Rapidly sharing ideas and complementary skills can help firms reduce development time. The theory here is that good ideas get validated and bad ideas get discarded more quickly, which leads to faster product development [5][8].

SN tools have made it possible to tap into the decision-making capabilities of the collective on a greater scale than ever before by opening discussions to a greater sample of resources, with greater disparity of areas of knowledge [13]. A growing number of applications have shown that a large enough group of diverse, independent and reasonably informed people might outperform and get to an end result that reflects a complete truth more effectively than a single expert or closed group [2][4].

For many problems that a company faces there can be a solution far outside of the traditional places that managers might search, within or outside the organization. Furthermore, decisions made at the head office may not fit local or field realities. The knowledge of those who have the necessary information from being in the field can be more effective than the use of top-down, template-based decisions. Collective intelligence accumulated via SN can help provide a diversity of viewpoints and input that can deter self-serving bias and belief perseverance, and can help combat pattern obsession and negative framing effects [2].

Most business people are familiar with SN sites like MySpace, Facebook, and LinkedIn, other online communities and SN tools. Since SN has become a part of these people's lives outside of the office, they will be eager to use these tools; potentially leading to more business collaboration [8].

Social networks are also easy and intuitive. There is great business opportunity here since there is always demand among learners for easy-to-use tools that simplify processes and SN tools can be used to replace more complicated collaboration tools. This ease of use epitomizes SN's potential for companies that want to tap the knowledge of their workers [8].

The foundation of SN is its social context. Sharing is encouraged, and the open, visible contributions and interactions reduce barriers to information flow. The personal nature and immediacy of SN can make interaction less impersonal and artificial than older bulletin board, mailing lists, and collaboration tools [9]. Adding a face and personality to the names of coworkers and business partners can go a long way toward supporting productive interaction [11]. For instance, a SN application like Facebook could act as a virtual employee water cooler [8]. Some individuals who might not otherwise interact as extensively with co-workers are actively participating. Formerly, people were forced to give up their knowledge, but with social networks, people willingly give up their knowledge.

SN exchanges are preserved, creating a record of previous conversations. Within a SN, employees search, view, bookmark/tag, rate, comment on, and edit information. In doing so, employees leave "digital fingerprints" on the content they access and these "digital fingerprints" provide insight into what is influencing the daily work of employees [7]. Capturing this valuable engagement data and making it actionable presents organizations with a clear opportunity to visualize and improve the way information is both consumed and contributed to by employees. These findings are important for businesses since they point to the nature of how people find and internalize information and more importantly how organizations can then make these traits part of their best practices [9].

LIMITATIONS AND CONCERNS

There are several limitations and concerns in bringing SN into the business environment. SN requires that firms turn over the technology experience to the end user. That is not normal, nor is it comfortable for companies, especially information technology (IT) departments [3] [18]. Also, SN applications often lack any explicit refereeing process that might provide some degree of quality assurance, which could lead to unwanted and undesirable outcomes. Communities of interest could drown out any voice of reason leaving the majority view essentially unchallenged [13]. Further control concerns include unpredictability, unassigned liability, and data leakage from staff gossiping freely in an open environment. Consequently, one of the biggest issues with respect to control is whether to include outsiders in the process [2] [18].

Companies have reservations about SN privacy and security, and rightfully so. SN opens up new avenues for the introduction of malware and phishing scams practiced by cyber-crooks. Also, businesses should be wary of potential about open access to the company servers as a result of lax and outdated attitudes toward passwords [8][18].

The assumption that an unmediated open group of resources will always come to a better conclusion than a single expert or closed group is dangerous. Companies can collect information from myriad sources and then perform some sort of averaging. In this case, the whole is equal to the sum of its parts, but the key is to maintain the right balance between diversity and expertise. Certain problems are more appropriately addressed by a diversity-based approach than others, but no amount of diversity will help if the group is completely ignorant of the issues. Therefore, firms need to decide which people to involve in group decisions and whether or not each participant should be given an equal voice. Even an application like Wikipedia, which might look simple on the surface, relies on a complex hierarchy of carefully selected editors [2][13].

Companies have generally been convinced of the value of connectivity and sharing information, but SN communities sometimes lack focus. Personal SN sites such as MySpace, Facebook, and Twitter, have been given a bad rap and are often seen as vehicles for sophomoric self-aggrandizement [5]. This has lead some to argue that SN interaction in the business environment is a nonproductive use of time. It also raises policy questions around moderating employee behavior and the use of network bandwidth [18]. Also, introducing SN into the enterprise presents a learning curve for workers whom are not familiar with SN and are used to communicating in specific ways. Firms not fully convinced of the business value of SN stand to waste significant time with employees needing training [8].

Use of SN as business tools lacks a reliable formula for measuring return on investment (ROI). When these tools are used to connect with customers and partners, there are usually ways to calculate a payback, but when companies provide them to employees, they're often going on gut instinct that SN will be good for business [5]. It is hard to come up with a reliable yardstick to measure the cost of the tool versus cost savings due to time saved or new opportunities created. There is no real way to know how solutions will fit in a firm's environment until they have been implemented and used by employees and customers [18].

DISCUSSION AND SUGGESTION

With thousands of active user groups already contributing to social networks, companies have been cautious when adopting SN technologies. SN is here to stay so it is important for businesses to find a practical way to adapt to it and work with these SN sites [18]. Following are some suggestions for firms when considering in rolling out SN applications.

Create Community

As a collaborative tool for KM, SN must be embraced by all employees and should inspire the frequent sharing of valuable knowledge. What motivates people to participate in a collective undertaking can vary widely so organizations must provide a continuous flow of new, enthusiastic participants to keep engagement high, or they need to provide incentives to sustain people's motivation over time [2] [14]. Rewards and recognition are not necessarily monetary in nature. Instead, a community of practice (CoP) should be formed around a recognized identity that all members can relate to and feel part of.

For a CoP to be successful, the community must become part of the practice itself. Community members must be able to easily see a direct benefit from being a member of the CoP and the community must take on a sufficient level of importance to its members. Otherwise, it becomes easy for them to lose interest in contributing. The "what's in it for me" factor is lost and the pressure of day-to-day business outweighs any reasons to contribute. Members of communities rely on other members for assistance with work-related issues, problem-solving and professional support. They need to know that they can trust their colleagues if they are to share openly and they must also feel that they are treated with appropriate respect [17].

Do Not Let Fear Strangle Growth

Many organizations are wary of giving a voice to employees because they do not know what they will say. Businesses also worry that employees will overdo the social aspects of these applications. This may tempt organizations to police employee-generated content, either through monitoring or pre-approving contributions. However, it is important to resist that temptation, as it will drastically affect employee participation. Employees need time to become comfortable with the idea of speaking up, sharing ideas, and participating in company-wide conversations. A SN project will likely wither before it has a chance to grow if people fear the thought police [5].

Resist Exclusivity

Business units or teams may want to build gated communities, but that approach defeats the purpose of a social network. The value of SN is in broadening the number of individuals who are generating or evaluating solutions. A company may want to tap into people or groups that it has not traditionally included when collecting and evaluating ideas. For instance, it might want to reach across business functional barriers or even groups outside of the company [2][5].

Standard Code of Conduct

When people are allowed to contribute to decisions, the likelihood that some will misbehave increases with group size. An implicit code of conduct like e-mail protocol helps govern people's behavior [2]. The development of standards would also help to establish SN more firmly in the corporate world [11][17]. This is not to be confused with policing. As mentioned earlier, policing an enterprise SN will kill it.

Select the Right SN Technology

SN technology should be as simple, effective, and transparent as possible and applied as needed in order to support key community functions and effective group development. An important element with choosing technology for a SN is to be very clear from the start on exactly how the technology will best serve the community. Even though a particular platform is either popular or easily available at a given point in time, it may not necessarily be the best solution for all communities [17].

Search underpins the value of a SN so insufficient indexing and searching capabilities will make social applications less useful. The point of SN in business is to let people provide input into the

relevancy of content and people, so make sure SN has a search engine that allows for user-generated feedback such as tags and content-rating systems [5].

CONCLUSION

SN continues to expand across businesses and enterprises. Social network software could have a more far-reaching organizational impact than technologies adopted in the 1990s. Vendors including IBM, Microsoft, Adobe, Novell and Oracle are adding SN tools to their products. Similarly, vendors such as Jive and Ektron have gotten into the act by offering SN toolboxes with their core products, and Yammer, a tool that works much like Twitter but is intended for business use, includes a SN component that gives employees personal pages.

The bottom line is that SN tools are helping businesses streamline the processes of researching projects, forming teams, and sharing knowledge. The personal nature, familiarity, and ease of use of SN attract executives and employees to improve their collaboration and relationship. Clearly, the future of collaboration and knowledge sharing can be enhanced by SN tools. We are seeing the beginnings of a new era of how information and knowledge will be discovered, created, distributed, and utilized inside organizations.

In order for companies to stay ahead of the curve and, more importantly ahead of their competitors, companies and organizations need to seriously determine how they can implement effective marketing strategies that incorporate Web 2.0 capabilities. They need to interact in a more personal way and create favorable user experiences by proactively reaching out and connecting with their customers. Web 2.0 has introduced a new level of technological sophistication. RSS feeds enable companies to stay engaged with customers by delivering updates to them directly, without requiring them to visit the company's website. YouTube marketing strategies offer companies the ability to showcase their personalities provide and offerings in a way that can be more authentic. It also provides a platform for viral marketing campaigns.

Companies that fail to think outside of the box, and that fail to think about the opportunities presented by Web 2.0 strategies, will be at risk of being surpassed by their competitors. Web 2.0 is revolutionizing the Internet and the way users interact with the Internet. It will continue to have a very powerful effect. Adoption or lack of adoption will eventually be the difference between the companies that succeed and those that do not.

References

[1] Bernoff, J., & Li, C. *Harnessing the Power of the Oh-So-Social Web*. MIT Sloan Management Review, 2008, 49(3), 36-42.

[2] Bonabeau, E. *Decisions 2.0: The power of collective intelligence*. MIT Sloan Management Review, 2009, 50(2), 45-52.

[3] Burrus, D. *Social networks in the workplace: the risk and opportunity of Business 2.0.* Strategy & Leadership, 2010, 38(4), 50-54.

[4] Cardon, P.W., Marshall, B., Norris, D.T., Cho, J., et al. *Online and offline social ties of social network website users: an exploratory study in eleven societies*. The Journal of Computer Information Systems, 2009, 50(1), 54-64.

[5] Conry-Murray, A. *Can enterprise SN pay off?* Information Week, 2009, 1224, 23-29.

[6] Dearstyne, B. *Blogs, mashups & wikis: oh, my!*. Information Management Journal, 2007, 41(4), 25-31.

[7] Diehl, A., Grabill, J.T., Hart-Davidson, W. & Iyer, V. *Grassroots: supporting the knowledge work of everyday life*. Technical Communication Quarterly, 2008, 17(4), 413-434.

[8] Fitzgerald, M. *Why social computing aids KM*. CIO. 1-5. Retrieved April 15, 2009, from tttp://www.cio.com/article/395113/Why_Social_Computing_Aids_Knowledge_Management

[9] Gupta, P. & Carpenter, H. *Enterprise wide SN*. Business Intelligence, 2009, 12(3), 26-29.

[10] Kreitzberg, A.P. *Building a Web 2.0-friendly culture: success on the Web is about people, not technology*. People and Strategy, 2009, 32(2), 40-45.

[11] Lamont, J. *Social networking gets down to business*. KM World, 2009, 18(6), 10-13.

[12] Levy, M. (2009). *WEB 2.0 implications on knowledge management*. Journal of Knowledge Management, 2009, 13(1), 120-134.

[13] Longbottom, C. & Bamforth, R. *Social networking in the public sector*. Knowledge Management Review, 2008, 11(5), 30-33.

[14] Maulana, A.E. & Eckhardt, G.M. *Just friends, good acquaintances or soul mates? An exploration of web site connectedness*. Qualitative Market Research, 2007, 10(3), 227-242.

[15] O'Reilly, T. *What is Web 2.0: Design patterns and business models for the next generation of software*. Retrieved March 11, 2009 from http://oreilly.com/web2/archive/what-is-web-20.html.

[16] Parise, S. *Social media networks: What do they mean for knowledge management?*. Journal of Information Technology Case and Application Research, 2009, 11(2), 1-11.

[17] Posey, C., Lowry, P.B., Roberts, T.L. & Ellis, S. *Proposing the online community self-disclosure model: the case of working professionals in France and the U.K. who use online communities*. European Journal of Information Systems, 19(2), 181-195.

[18] Wilson, J. *Social networking: the business case*. Engineering & Technology, 2009, 4(10), 54-56.

A Framework for Managing End-of-life Pharmaceutical Products

Nilima Shroff[1]   **Elif Kongar[2]**

University of Bridgeport, USA   *University of Bridgeport, USA*

ABSTRACT

A product is considered in its end-of-life (EOL) when it completes its service life time. Completion of service life time is either caused by deterioration or obsolescence. Deterioration implies that the product is no longer usable due to disintegration or degeneration. Obsolescence, on the other hand, refers to those products that are rendered invalid due to systemic, functional, and style mismatch or due to notification of expiration set by the manufacturer. Therefore, obsolete products are likely to preserve their initial conditions. Pharmaceutical products, specifically prescription drugs, constitute one product category that completes its service life time before it deteriorates. The mismatch between the utilization period and the obsolescence date leads to accumulation of stored (hibernating) or discarded EOL pharmaceutical products. Furthermore, following their expiration date, some of these pharmaceutical products also become toxic and hence hazardous to human health and the environment. The economically, socially, and environmentally sustainable option is to take back and reuse, recycle, and/or properly dispose of these products. Reverse logistics systems focus on such problems. With these motivations, this study proposes a reverse logistics framework that embodies environmental, economical and

[1] Nilima Shroff, MS Candidate, Department of Technology Management, University of Bridgeport, 221 University Avenue, School of Engineering, 141 Technology Building, Bridgeport, CT 06604, USA, Phone: (203) 576-4379, Fax: (203) 576-4750, E-mail : nshroff@bridgeport.edu

[2] Contact Author: Elif Kongar, Ph.D., Departments of Mechanical Engineering and Technology Management, University of Bridgeport, 221 University Avenue, School of Engineering, 141 Technology Building, Bridgeport, CT 06604, USA, Phone: (203) 576-4379, Fax: (203) 576-4750, E-mail : kongar@bridgeport.edu

physical concerns for EOL pharmaceutical products. The Information Technology (IT) infrastructure required for the proposed system along with a comprehensive overview of the existing take-back regulations in the U.S. are also provided.

Keywords: End-of-life, information technology, life cycle, pharmaceuticals industry, reverse logistics, take-back programs.

1. Introduction

A product is considered in its end-of-life (EOL) when it completes its service life time. Completion of service life time is either caused by deterioration or obsolescence. Deterioration implies that the product is no longer usable due to disintegration or degeneration [1]. Obsolescence, on the other hand, refers to those products that are rendered invalid due to systemic, functional, and style mismatch or due to notification of expiration set by the manufacturer. Therefore, obsolete products are likely to preserve their initial conditions. Pharmaceutical products, specifically prescription drugs, constitute one product category that completes its service life time before it deteriorates. Research indicates that many patients discontinue the prescribed treatment and switch over to other medications [2, 3]. The mismatch between the utilization period and the obsolescence date leads to accumulation of stored (hibernating) or discarded EOL pharmaceutical products. The Healthcare Distribution Management Association (HDMA) estimates a three to four percent return rate for pharmaceutical products for redistribution, recycle and disposal [4]. These products that are unused and/or expired constitute a significant financial market since the majority of expired drugs in their unopened original container would be expected to remain stable for an average of 57 months after their expiration [3]. Following their expiration date, some of these pharmaceutical products - such as antibiotics - also become toxic and hence hazardous to human health and the environment [5], if not disposed of properly. The economically, socially, and environmentally sustainable option is to take back and reuse, recycle, and/or properly dispose of these EOL products. Reverse logistics systems focus on this problem and involve planning, management and controlling the flow of waste with appropriate EOL processing option, viz.,

reuse, recycle and/or proper disposal [6]. A well established reverse logistics system that holds apparent and consistent information and product flow of prescription drugs would also identify and prevent counterfeit medications. With these motivations, this study proposes a reverse logistics framework that embodies environmental, economical and physical concerns for pharmaceutical products. The Information Technology (IT) infrastructure required for the proposed system along with a comprehensive overview of the existing take-back regulations in the U.S. are also provided.

2. Background Research and Literature Review

The literature offers a broad variety of studies investigating forward logistics of pharmaceutical products. Forward logistics focus on the traditional flow, from manufacturers to distributers to consumers through the retailers. Even though the issue is now of significant importance to pharmaceutical companies, health organizations and governments [7], research on green reverse logistics operations of EOL pharmaceutical products is limited.

One of the few studies that provide data on returned pharmaceuticals is proposed by Sartori [4]. Martin [8] and Teunter [9] estimated the pharmaceutical return management cost to be $2.5 billion dollars and report an approximate $5 billion dollar cost for returned products due to expiration, damage, recall or improper delivery. Hunter et al. [10] reported approximately three to six percent return rate for pharmaceutical products. Cross [11] and Sartori [4] emphasized the social and financial significance of returned prescription drugs. Sartori, investigated various alternatives to transport expired medications for recycling or complete proper disposal, and reported that counterfeit drugs are approximately ten percent of the global pharmaceutical market [4]. The issue of waste transportation is also investigated by Jennings and

Scholar [12]. Lee and Chan [13] proposed an algorithm to determine the customer's locations in order to reduce the total cost of reverse logistics transportation.

The complexity of reverse logistics systems and the importance of information technology (IT) in company performance are well documented. The proper utilization of IT leads to performance efficiency in the company by eliminating poor information flow and manual mistakes [14-20]. In terms of particular IT solutions, Lee and Chan [13] presented RFID as a solution to identify the category of the collected product in collection point. Vadde and Ilgin [21-23] also studied the economic benefits of sensor embedded products for EOL processing operations. In order to reduce the complexity of the reverse logistics systems and manage the network effectively, outsourcing is also proposed as one of the viable options [24, 25].

3. National Review

This section provides an overview of the current Food and Drug Administration (FDA) drug take-back rules and regulations in the United States. Our research indicates that all U.S. states have developed rules and regulations for drug recycling except the states of North Carolina, South Carolina and Oregon.

Even though not all medicines become toxic immediately after their expiration dates [26], there are still some drugs, such as medications to treat cancer and HIV, that become highly poisonous after the notified expiration date. In order to control the sale of expired drugs the FDA has devised various programs such as the "Shelf Life Extension Program", which tests the stability of the drugs after their expiration [3].

In 2009, the National Conference of State Legislature (NCSL) has identified the state legislations to create prescription drug recycling, repository or redistribution program for EOL medications. According to the legislation, the EOL medications should be distributed to needy

5

individuals who cannot afford to purchase the expensive medicines [27]. NCSL also reports that 36 states have laws for drug recycling. Out of these 36 states, six of them, viz., Colorado, Florida, Kentucky, Minnesota, Nebraska and Wisconsin specifically focus on accepting and redistributing cancer medications [27].

Frisman [28] reports that the state of Washington has funded a product stewardship program in which all the drug sellers in the states will be involved and responsible to pay for collecting, transporting and disposing unwanted drugs, supposed to be implemented by January, 2012. The main purpose of drug take-back programs is to allow the safe, legal, and environment friendly disposal of extra drugs which helps to reduce health and environmental impacts of consumer products [29]. The economic viability of such programs is also reported. For instance, Fryer [30] reported that the redistribution of unopened prescription medicines in Pennsylvania is estimated to save the state approximately $1 million annually. Furthermore, Pomerantz [3] emphasizes the role of the FDA in monitoring and eliminating the redistribution of recycled medicines that are impure, misbranded, expired, or counterfeit. However, despite the advantages, laws and programs for redistribution of EOL drugs, pharmacies are reluctant to participate due to the unknown liability to the participants [31]. Figure 1 depicts the state of FDA laws regulating drug recycling in the U.S.A.

4. Methodology

In this section, a reverse logistics system for pharmaceutical products is proposed. The study portrays a composite system for End-of-life processing operations of pharmaceutical products. The system is modeled as a reverse supply chain that includes institutional targets, reverse logistics operations and the information technology (IT) infrastructure required for the EOL operations. The decision input includes uncontrollable and controllable factors. The model

6

embodies various physical, environmental and financial constraints while targeting multiple objectives.
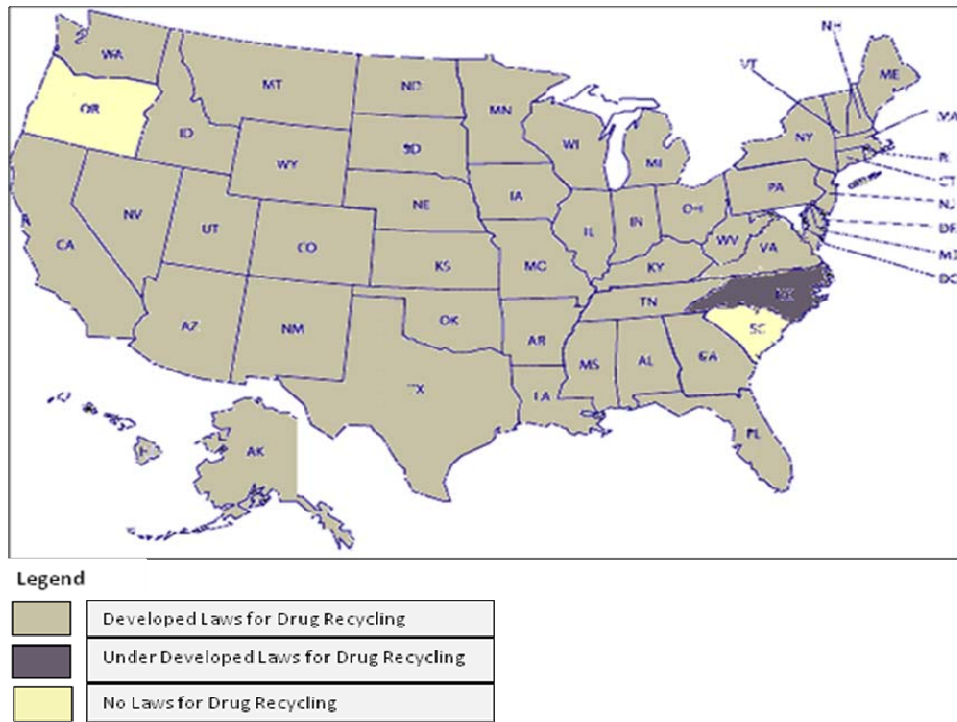
Figure 1 FDA laws in the United States of America for drug recycling processes

The proposed system consists of two main processes: 1) The collection process which includes a 'Drug Take Back' program, and 2) The inspection process.

Figure 2 depicts the forward supply chain and proposed reverse supply chain product and information flow, decision variables for the reverse logistics operations, and the EOL processing options for the pharmaceutical products.
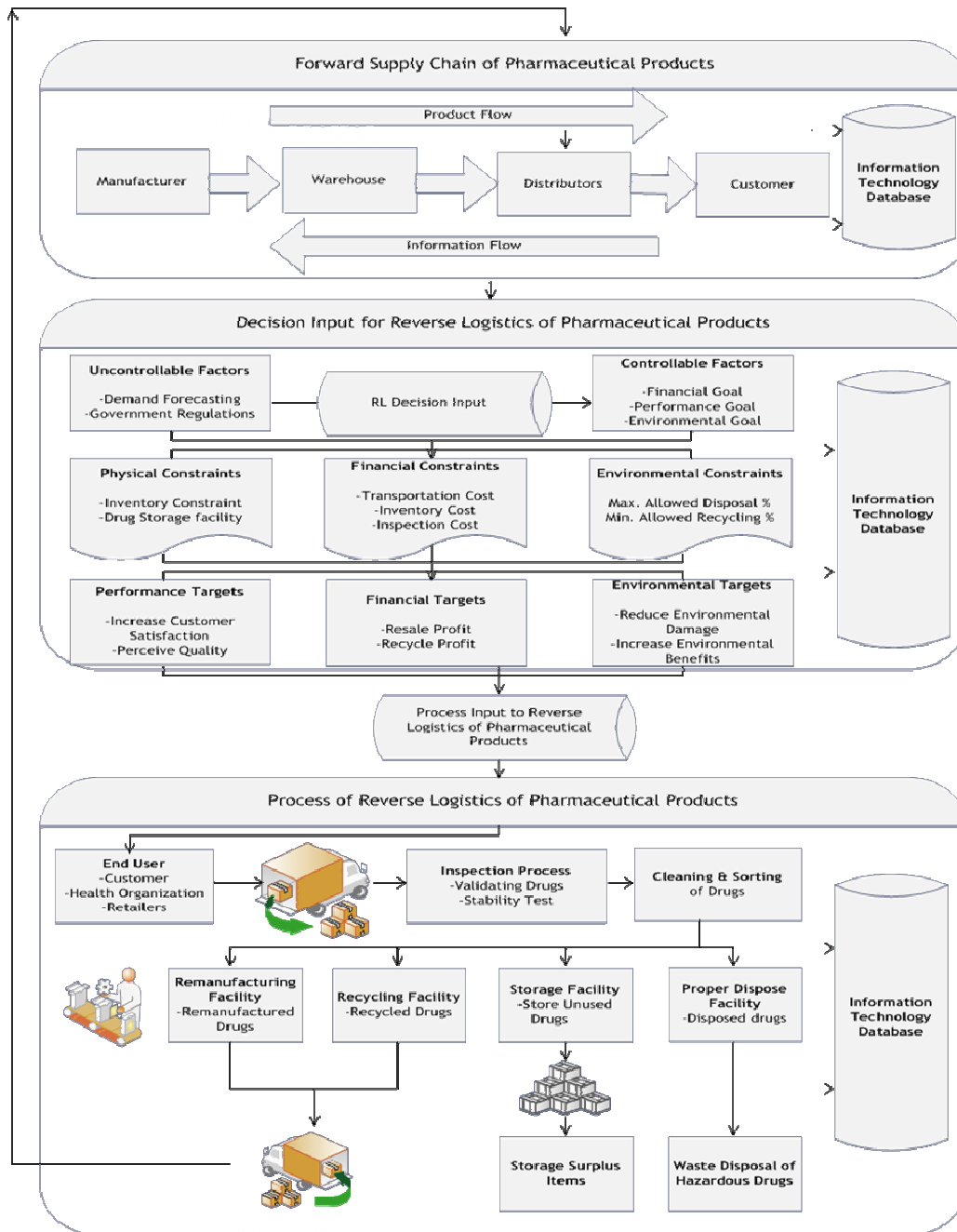
Figure 2 Product and information flow for the forward and proposed reverse logistics

system (Adopted from: [32])

4.1   Forward supply chain of pharmaceutical products

The forward supply chain starts with the end user's need and the information flow through distributors, warehouses and manufacturers in order to meet the demand for pharmaceutical products.

4.2   Decision Input for the proposed reverse logistics system

The uncertainty in take-back systems increases the risk in ensuring the sustainability of the overall system. Utilization of accurate data that is shared among the related parties in the reverse logistics supply chain is effective in reducing this complexity. The proposed decision input module aims at reducing the vagueness in the overall reverse supply chain. Long term strategic decisions ensuring economic and environmental sustainability of the take-back operations are made at this level. The module utilizes two main data categories: (1) uncontrollable, and (2) controllable factors. Physical, financial and environmental targets of the company and system restrictions are also included in this module. Uncontrollable factors include the future demand for returned products and the rules and regulations which are imposed by the FDA and other governmental institutions.

Future demand for pharmaceuticals can be forecasted using historical data reported by various organizations such as the Healthcare Distribution Management Association (HDMA) [4]. Furthermore, various statistics on the form of EOL medications, i.e., liquid, gel, pills, or powder, etc.; would reduce the system complexity while allowing more accurate estimates for the profitability of the operations. Collecting data on the end user behavior and recording the reasons of returns; e.g., medication cutoff and deterioration, would aid in determining the reusability rates of prescription drugs. Determining the appropriate EOL operations such as redistribution,

recycling or complete disposition would lead to economic and environmental sustainability in the overall system.

Controllable factors include institutional goals for financial, environmental and performance related variables.

Overall, the proposed system focuses on reducing the lead time as a result of both, poor information flow and the high numbers of actors involved in the forward and reverse logistics operations.

Here, financial targets concentrate on the resale and recycling profits obtained from the reverse logistics operations, where as environmental targets focus on reducing the environmental impact of these operations. Performance related targets aim at increasing the customer satisfaction while providing ease in take-back operations.

There are various constraints involved in the reverse logistics process of medications such as physical, financial, and environmental constraints. *Physical constraints* include capacity restrictions of the overall system including storage, manufacturing, handling, and recycling, in addition to the cooling facility allocations for perishable items. *Financial constraints* include the cost of reverse logistics of EOL medications such as transportation, storage, and inspection. *Environmental constraints* restrict the amount of disposal and impose a minimum level for recycling.

The decision phase covers all controllable and uncontrollable factors which are considered as significant input to implement the reverse logistics of EOL pharmaceutical products.

10

4.3 Product and information flow in the reverse logistics system

The proposed system is mainly partitioned into two processes: 1) Collection / End user notification and (2) Inspection.

Collection / End user notification: The FDA has initiated 'Drug Take Back' programs for EOL medications in almost all of the states in United States. Thus, the EOL medications from retailers, customers, and health organizations are accepted at various centers [29]. Once received, these medications are forwarded to next stage, namely the inspection centers. This transfer includes handling operations such as accepting, sorting, packing and the transportation of EOL drugs.

Inspection Process: Expiration dates for pharmaceutical products are set by manufacturers. The date indicates the end of the manufacturer's responsibility. To determine the expiration date, a stability test is conducted to ensure that the identity, strength, quality and purity of the drug is in compliance with the FDA regulations during its useful life [33].

In the proposed system returned EOL medications are primarily sorted according to their expiration dates. Valid medications are sent to stability tests to ensure the validity of the drugs and are then redistributed. The expired drugs are further tested for their toxicity. Toxic drugs are disposed of properly where as non-toxic drugs are sent to redistribution and/or recycling facilities. Storage is considered for the drugs that have no market value at the given time period and/or for those that will not ensure financial gain.

4.4   IT infrastructure for the EOL pharmaceutical products

The proposed system aims at increasing the data quality and reliability throughout the supply chain. The system also allows the data to be shared among each party in the supply chain through the entire product life cycle.

Figure 3 illustrates the required IT infrastructure for the reverse logistics operations. At the inspection center, the barcode of EOL medications will be scanned and related stored information will be displayed, which will help to identify the drug and its contents. The database will update the status of the drug as EOL. The main input is provided by the barcode on the packaging in prescription medications and other pharmaceutical products. The database is visible to manufacturers, distributors, retailers, and customers. The data provide ease in decisions such as inventory management, warehouse management, and transportation.

After the EOL medications are returned to the collection facility, the information on the medication is read from the barcode and entered to the database. For prescription drugs, the data include the prescription number, the name of the medication, quantity, description of the drug such as color, shape, texture, etc., filling date, patient, prescriber, manufacturer information, and expiration date.

Following the system entry the information is made available to the other parties in the system. The medications are then differentiated based on the toxicity characteristics; i.e., toxic and nontoxic. Generally, the drugs which are used for HIV Aids treatment, Chemotherapy treatment for cancer, etc., become toxic after a period of their expiration date [3].
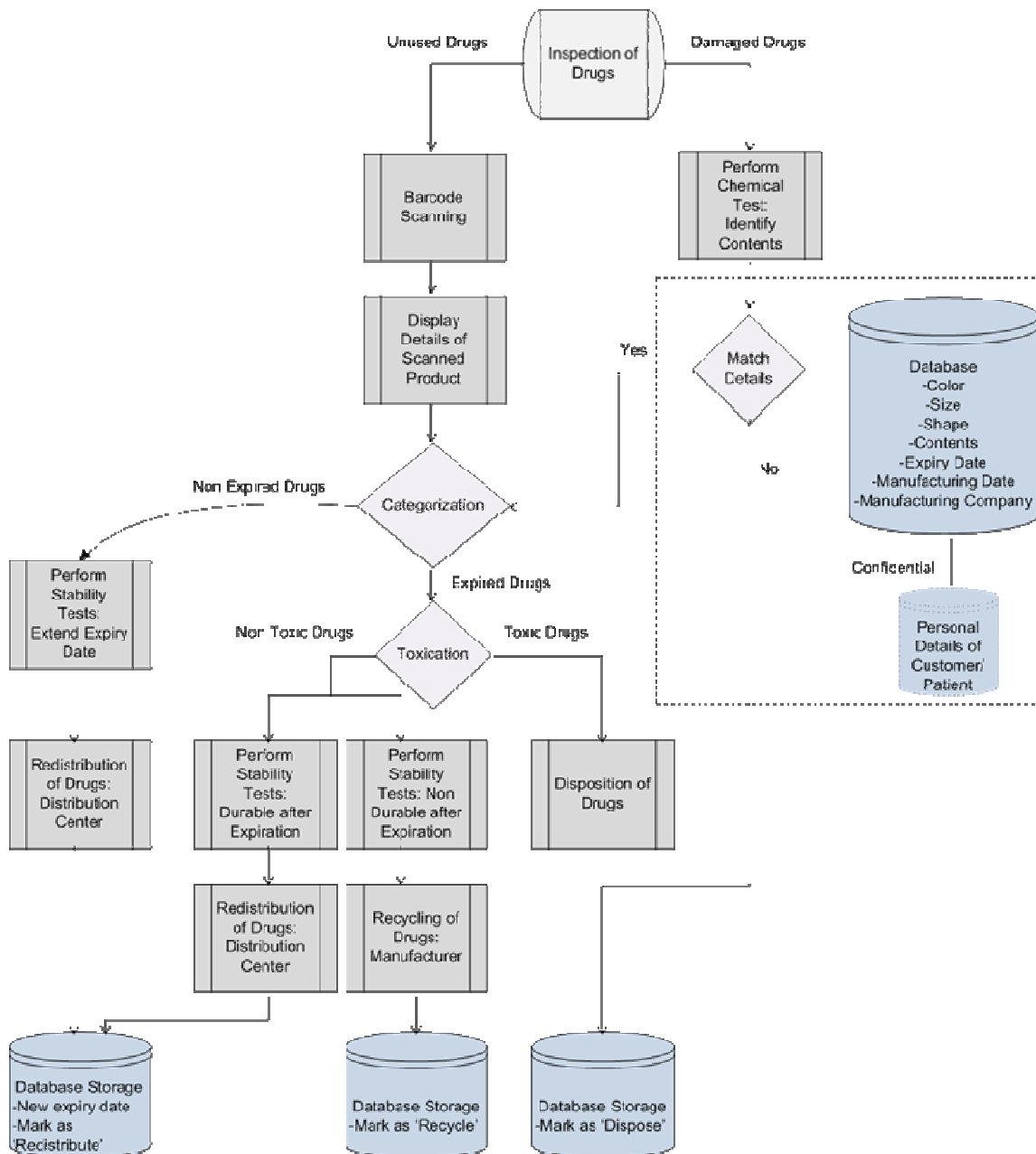
Figure 3 Data Flow Diagram of the Proposed Inspection Process

The drugs will be classified based on the expiration date and toxicity, and future actions take place as described below:

1)   Unexpired drugs*:* Perform stability tests to determine the remaining life of the drug. Accordingly, shelf life of drug is extended and sent back to redistribution centers or warehouses

for further resale. The system database will update the expiration date and store the results of stability test of the drug and mark it as 'Redistribute'.

2)   Expired and Nontoxic drugs: Perform stability tests to analyze reusability of the drug. If it is reusable; meaning, if the drug is still in compliance with the FDA standards, then it is considered as a 'still functioning' product (Kongar & Gupta, 2009), and send back to redistribution centers or warehouses for resale. The system database will update the expiration date and store the result of stability test of the drug marking it as 'Redistribute'.

If the drug is not reusable, further chemical tests will be performed to analyze its recyclability; meaning, regaining the chemical or substance value added to that particular drug (Kongar & Gupta, 2009). If it is recyclable, it will be sent to manufacturers for recycling processes.

If the tests result shows drugs are completely expired, then it will be sent for proper disposal and the system database will mark it as 'Dispose'.

Regardless of the end destination, the packaging materials of pharmaceutical products: metal, plastic, paper, and glass, are also some of the items which can be regained via recycling operations. The system database will mark these packages as 'Recycle'.

3)   Expired and toxic drugs: Even though most drugs are still usable after their expiration date, there are few exceptions that become toxic following their expiration date. For instance, liquids that contain sugar or other flavoring additives become unstable faster than the remaining drugs. Antibiotic suspensions prepared from powder also become unstable and should be disposed properly. Hence, drugs which fall under this category will be sent for proper disposal processes in order to avoid natural hazards. The system database will be updated and will mark these drugs as 'Dispose'.

4) Damaged drugs: The proposed database also considers the damaged medications. Damaged medications include broken tablets/pills or drugs with damaged wrappers/packets that would prohibit barcode scanning operations. In these cases where the system is not able to utilize barcode scanning, chemical tests will be performed to identify the contents of the drug. Then, the results of these tests including contents will be matched against the information stored in that particular drug database at the initial stage of the supply chain. Once the drug is identified, the stability tests will be performed to decide whether the drug is reusable, recyclable or subject to disposal. The system database will be updated accordingly.

It is also possible that the results of chemical test will not match the information stored in the database and the system will not be able to identify the drug. In these instances, these drugs will be sent for proper disposal to minimize the potential hazard to human health.

Once the return of EOL medication is submitted, the system will make available the information to all entities in entire supply chain. The overall goal of the information technology infrastructure is to remove inefficiencies such as delays, manual errors, and poor information and product flow from the product return process and to achieve pre-determined targets.

4.5   Challenges and Advantages of the proposed system

Cost: The proposed system includes a large variety of cost measures such as transportation, inventory, inspection, and EOL processing at every step. Furthermore, the proposed IT infrastructure introduces additional cost measures such as equipment, training, and planning to ensure system reliability and efficiency. However, effective use of information technology, government incentives and subsidizing would decrease the overall cost and even help create a financially viable system.

Pricing: When the drugs return to the retailer or manufacturer, a refund must be made based on the original price [3]. Since the price of the medication tends to fluctuate, it is quite difficult to estimate the correct pricing of the EOL drugs. Wholesale reductions, blanketing, and shelf life [34] are also factors that make pricing difficult for the credit system . The proposed IT system aims at creating an automated credit system that eliminates over-credit or under-credit of returns. Governments can also offer insurance and incentives to the customer to encourage buying recycled medications, which would help increase the demand for those medications [3].

Redistribution: The EOL medications that are not expired will be resold again if profitable. Furthermore, if the drugs are in good condition and their shelf life is also long, then those medications can be sold at higher rate than the original one [35]. The recyclers can use the difference between the acquisition cost and sale price to fund the expenses, including operation cost and profit. In addition, donation of these returned and refurbished EOL medications would add value to the public, as well as decrease the environmental damage by preventing unnecessary consumption, incineration or land filling.

Effective Process Management lower operation cost: Effective use of technology incorporates effective management processes to reduce cycle time and cost throughout the entire reverse logistics system. For example, based on history data in the system, forecast will predict the approximate quantity of returned medications at collection centers, which will lead to optimize transportation and inventory cost. The system provides a high level of visibility from one end to the other in the forward and reverse supply chain, which will result in reducing administrative cost.

5. Conclusions and future research

Three to four percents of drugs or pharmaceutical products are returned, corresponding to a significant amount in pharmaceutical manufacturing. Hence, it is important to implement a system for reverse logistics of these EOL pharmaceutical products. Redistributed, recycled or properly disposed products would help the pharmaceutical industry by regaining the chemical or substance value added to them with considerable benefits to human health and society.

The paper demonstrated an IT infrastructure that would ensure data accuracy to trace EOL pharmaceutical products and to improve the overall performance of the reverse logistics system. The uncertainty in the overall system is one of the main concerns in ensuring the financial stability of the reverse logistics operations. However, this study aimed at summarizing the current rules and regulations, highlighting the gaps in the literature, and proposed a reverse logistics system and the required IT infrastructure. The proposed system considered various controllable and uncontrollable constraints, goals and targets to put the idea of reverse logistics into practice.

Future research will include collecting data to analyze the total sale of pharmaceutical products, the number of products returned, the type of products or medication used for particular diseases, and existing ways to dispose of EOL medications in the U.S.

6. Acknowledgements and In Memoriam

References

[1] Mueller, D.B., Cao, J., Kongar, E., Altonji, M., Weiner, P.-H., and Graedel, T.E. *Service Lifetimes of Mineral End Uses, Report, USGS award Number: . November 2007.*, 2007, USGS.

[2] Leslie, D.L. and Rosenheck, R.A. *From conventional to atypical antipsychotics and back: dynamic processes in the diffusion of new medications. The American Journal of Psychiatry* 159(9): p. 1534-1540, 2002.

[3] Pomerantz, J.M. *Recycling Expensive Medication: Why Not?* 2004; Available from: http://www.ncbi.nlm.nih.gov.

[4] Sartori, G. *Reverse Logistics Role in Securing the Pharmaceutical Supply Chain.* 2010 Sep. 28; 16:[Available from: http://www.rlmagazine.com/edition16p18.php.

[5] Kümmerer, K. *Pharmaceuticals in the Environment.* Annual Review of Environment and Resources 2010; 57 -75].

[6] Hu, T.-L., Sheu, J.-B., and Huang, K.-H. *A reverse logistics cost minimization model for the treatment of hazardous wastes. Transportation Research Part E: Logistics and Transportation Review* 38(6): p. 457-473, 2002.

[7] Valero, G. *Logistics in reverse.* 2005; Available from: www.pbsmag.com/Article.cfm?ID.

[8] Martin, P. *"Reverse logistics in pharma industry", The Remanufacturing Gazette Magazine.* 2007; Available from: http://remannews.blogspot.com/2007/03/reverse-logisticsin-pharma-industry.html.

[9] Teunter, R., Inderfurth, K., Minner, S., and Kleber, R. *Reverse logistics in a pharmaceutical company: a case study.* in *Econometric Institute Report.* 2003.

[10] Hunter, T., Droege, M., Marsh, W., and Droege, W. *Effectively managing pharmaceutical returns and waste. Drug Topics* 149(2): p. 36, 2005.

[11] Cross, C. *A hard sell: Carolina Logistics Services reverses the route of controlled substances. Industrial Engineer* 39(5): p. p.24, 2007.

[12] Jennings, A.A. and Scholar, R.L. *Hazardous waste disposal network analysis. Journal of the Environmental Engineering* 110(2): p. pp. 325–342., 1984.

[13] Lee, C.K.M. and Chan, T.M. *Development of RFID-based Reverse Logistics System. Expert Systems with Applications* 36(5): p. 9299-9307, 2009.

[14] Bowersox, D.J. and Daugherty, P.J. *Logistics paradigms: the impact of information technology. Journal of Business Logistics* 16(1): p. pp. 65–80, 1995.

[15] Chandra, C. *The case for healthcare supply chain management: insights from problem-solving approaches. International Journal of Procurement Management* 1(3): p. 261 – 279, 2008.

[16] Closs, D.J., Goldsby, T.J., and Clinton, S.R. *Information technology influences on world class logistics capability. International Journal of Physical Distribution and Logistics Management* 27(1): p. pp. 4–17, 1997.

[17] Kumar, S., Dieveney, E., and Dieveney, A. *Reverse logistic process control measures for the pharmaceutical industry supply chain. International Journal of Productivity and Performance Management* 58(2): p. pp.188-204, 2009.

[18] Rogers, D.S. and Tibben-Lembke, R. *Going Backwards: Reverse Logistics Trends and Practices.* 1999, Pittsburgh, PA: RLEC Press. pp. 43.

[19]   Ross, A. *A multi-dimensional empirical exploration of technology investment, coordination, and firm performance. International Journal of Physical Distribution and Logistics Management* 32(7): p. 591–609, 2002.

[20]   Sundarraj, R.P. and Talluri, S. *A multi-period optimization model for the procurement of component-based enterprise information technologies. European Journal of Operations Research* 146(2): p. 339–351, 2003.

[21]   Ilgin, M.A. and Gupta, S.M. *Comparison of Economic Benefits of Sensor Embedded Products and Conventional Products in a Multi-Product Disassembly Line. Computers and Industrial Engineering* 59(4): p. 748-763, 2010.

[22]   Ilgin, M.A. and Gupta, S.M. *Recovery of Sensor Embedded Washing Machines Using a Multi-Kanban Controlled Disassembly Line. Robotics and Computer Integrated Manufacturing* 27(2): p. 318-334, 2011.

[23]   Vadde, S., Kamarthi, S.V., Gupta, S.M., and Zeid, I., *Product Life Cycle Monitoring via Embedded Sensors*. in *Environment Conscious Manufacturing*, Gupta, S.M. and Lambert, A.J.D., Editors, 2008, CRC Press, p. 91-104.

[24]   Cheng, Y.-H. and Lee, F. *Outsourcing reverse logistics of high-tech manufacturing firms by using a systematic decision-making approach: TFT-LCD sector in Taiwan. Industrial Marketing Management* 39(7): p. 1111-1119, 2009.

[25]   Krumwiede, D.W. and Sheu, C. *A model for reverse logistics entry by third-party providers. The International Journal of Management Science* 30(5): p. pp. 325−333, 2002.

[26]   Cohen, L.P. *Drugs Frequently Potent Past Expiration. THE WALL STREET JOURNAL*, 2000.

[27]   NCSL. *State Prescription Drug Return, Reuse and Recycling Laws.* 2010; Available from: http://www.ncsl.org/IssuesResearch/Health/StatePrescriptionDrugReturnReuseandRecycl in/tabid/14425/Default.aspx.

[28]   Frisman, P. *COLLECTING AND DISPOSING OF UNUSED MEDICATION AND RECYCLING PRESCRIPTION BOTTLES. OLR Reserach Report*, 2009.

[29]   Cotter, P. *Drug Takeback Programs – National Day Planned.* 2010; Available from: http://www.naag.org/drug-takeback-programs-national-day-planned.php.

[30]   Fryer, M. *Pennsylvania Senate OKs bill allowing jails to recycle unused medicine*. in *Leader Times*. 2010: Pittsburgh.

[31]   Agency of Human Services. *Drugs and Medical Supply Repository Study*. in *Report to the Legislature on Act 139 Sec. 6*, 2007, Department of Health.

[32]   Kongar, E. and Gupta, S.M. *Solving the Disassembly-To-Order Problem Using Linear Physical Programming. International Journal of Mathematics in Operational Research (IJMOR)* 1(4): p. 504-531, 2009.

[33]   Wan-Chih, T. *Pharmacist's Letter / Prescriber's Letter.* 2007; Available from: www.pharmacistsletter.com

[34]   IBM, C. *IBM Solution for Pharmacutical Track & Trace.* 2007; Available from: ibm.com/solutions/rfid.

[35]   Chen, M. *The enterprise value analysis based on Reverse Logistics.* 2006.

was most efficient. This study is a first step toward a more comprehensive assessment of biofuels, which will need to include additional criteria beyond those considered here. Each of these biofuels will need to be investigated for each criterion, not just from a production perspective, (i.e., quantity required and cost per unit), but also in terms of intangible criteria as well.

Keywords: Alternative Energy Sources, Biofuels, CO_2 Emission, Data Envelopment Analysis, Greenhouse Gases, Life Cycle Assessment.

INTRODUCTION

Over the last decade, the interest in use and production of renewable energy technologies has grown rapidly to include biofuels as an economically and environmentally benign option. Renewable transportation fuels can be produced from a variety of substrates, including agricultural residues, corn stover, grasses, legumes, algae, food processing wastes, and other biological materials. Feedstock preference and factory scale are geographic dependent. For example, in the U.S. corn grain is primarily used, but in Brazil sugarcane is predominant.

There have been many questions over the years regarding the sustainability of corn-based ethanol; these have been asked by the scientific community, policymakers, as well as the public itself. Many of these questions have focused on the production of the corn, manufacturing efficiencies, resource and energy inputs versus outputs during fuel

2

manufacturing (i.e., the net energy balance and the life cycle of ethanol), process economics, performance of ethanol in vehicles vis-à-vis gasoline, water consumption, land use change, greenhouse gas emissions, and the use of corn for fuel instead of food. To address these questions, many studies have been conducted to examine the overall costs and benefits of this biofuel and to assess its sustainability. Some overviews of these studies have been published as well [1-4].

Modeling and simulation studies, including Life Cycle Assessment (LCA), of the sustainability of biofuels have not yet been completely definitive. Results are dependent upon initial assumptions and system boundaries, as well as the specific steps which are included in the models. Moreover, there is no unanimity in the scientific community concerning adverse environmental impacts of energy extraction, conversion, transportation, and end use, making incorporation of environmental parameters disputable [5]. However, it is important to try to understand the implications of deploying these types of systems on a large scale, both throughout the U.S. as well as globally.

Beyond LCA, other approaches to assess sustainability of systems do exist. One of the most relevant research studies has been published by Ulutas [6]; in this research, the author aimed at applying an analytic network process (ANP) model to evaluate alternative energy sources for Turkey. Along similar lines, Junnila [7] performed a scenario analysis for a life-cycle assessment of a service sector company; the analysis involved six case companies to test the influence of 32 alternative scenarios on the

environmental impact of the median company. Furthermore, Geldermann and Rentz [8] described a multicriteria analysis (MCA) of environmentally relevant installations based on a case study of the surface coating sector. In addition, Ramanathan and Ganesh [9] developed an integrated model for the household sector of Madras, India, using goal programming (GP) and the Analytic Hierarchy Process (AHP) for energy resource allocation. Regarding alternative energy byproducts, Rosentrater and Kongar [10] performed a techno-economic simulation study to compare various fuel ethanol byproduct pelleting processes; the authors indicated that value-added processing is cost effective for a variety of scenarios for biofuel manufacturing residues.

Data Envelopment Analysis (DEA) is a well proven methodology allowing the introduction of multiple inputs and multiple outputs and obtains an "efficiency score" of each DMU with the conventional output/input ratio analysis [11-15]. The method has been extensively used in the literature to evaluate environmental sustainability. Sarkis and his colleagues utilized DEA to evaluate technical and ecological efficiencies of various industries and programs [16-20]. Chien and Lu [21] used the DEA method to estimate the technical efficiency for 45 economies in the years 2001 and 2002. Lozano and Gutiérrez [22] also proposed a DEA based approach to model the relationships among population, GDP, energy consumption and CO_2 emissions. Mukherjee [23] utilized DEA to analyze the energy efficiency for the aggregate manufacturing sector as well as for the six highest energy consuming sub-sectors for the period 1970–2001, and to propose various alternative models. Criswell and Thompson [24] used DEA

methodology to compare the technical efficiency of large-scale power systems needed to meet the growing energy needs of society.

Methods other than DEA have also been utilized to study material life cycles and eco-efficiency issues. Moyer and Gupta [25] studied minimization of life cycle scrap via a comprehensive survey of previous work on environmentally conscious manufacturing practices. Further, Isaacs and Gupta [26] analyzed the effects of substitution of high-grade plastics for steel using GP techniques. Boon et al. [27] also applied GP to examine the economic impact of aluminum-intensive vehicles on US automotive recycling. Creating a model of the automobile-recycling infrastructure, Boon et al. [28] also used the GP technique to assess material streams and process profitabilities for several different clean vehicles.

In this study, we combine LCA and DEA approaches to assess alternative energy technologies, specifically biofuels, and compare them using various sustainability criteria. This approach is illustrated with a numerical example, based on literature data for various biofuel options, providing a relative comparison for some of the technologies that were analyzed in a recent study [29].

MATERIALS & METHODS

DEA allows the introduction of multiple inputs and multiple outputs and obtains an "efficiency score" for each decision making unit (DMU) with the conventional

output/input ratio analysis [11]. DEA algorithms can be categorized using two criteria, namely, the "orientation" and the "optimality scale" criteria. The "orientation" criterion categorizes DEA algorithms into two depending on whether the definition of efficiency used in the algorithm is input- or output-oriented. *Input-oriented* DEA models are suitable for "the least input for the same amount of output" problems, whereas *output-oriented* DEA models target "the most output for the same amount of input". The "optimality scale" criterion categorizes DEA models into four based on the returns to scale. If production increases, efficiency may increase, remain constant, or decrease, thus, demonstrating "Increasing Returns to Scale (IRS)", "Constant Returns to Scale (CRS)", or "Decreasing Returns to Scale (DRS)", respectively. "Variable Returns to Scale (VRS)" refers to a case where both an increase and a decrease in returns to scale are observed at alternative levels of output. First introduced by Banker et al. [30] as an extension of the CRS DEA model, the VRS model assumes that not all DMUs operate on an optimal scale. In this study we propose a basic input-oriented Constant Returns to Scale (CRS) model, assuming constant returns to scale for all of the inputs and outputs.

DEA Methodology

DEA defines basic efficiency as the ratio of the weighted sum of outputs to the weighted sum of inputs, the relative efficiency score of a test DMU. This non-linear problem can be converted into a linear program as follows:

$$\max \quad \sum_{k=1}^{s} v_k \, y_{kp} \qquad\qquad (\,1\,)$$

$$s.\ t.\quad \sum_{j=1}^{m} u_j x_{jp} = 1$$

$$\sum_{k=1}^{s} v_k y_{ki} - \sum_{j=1}^{m} u_j x_{ji} \le 0 \qquad \forall\, \text{DMUs } i$$

$$v_k, u_j \ge 0 \qquad\qquad \forall\, k, j.$$

In Equation (1), the $\sum_{j=1}^{m} u_j x_{jp} = 1$ constraint sets an upper bound of 1 for the relative

efficiency score, and,

$$k = 1 \text{ to } s,$$

$$j = 1 \text{ to } m,$$

$$i = 1 \text{ to } n,$$

y_{ki} = amount of output k produced by DMU i,

x_{ji} = amount of input j produced by DMU i,

v_k = weight given to output k,

u_j = weight given to input j.

The CCR model given in Equation (1) must be run n times for n DMUs to obtain the technical efficiency (TE) of each DMU. The model is characterized by constant returns to scale (CRS). Please see [11-15] for further information on DEA methodology and the details of the CRS DEA model.

DEA System Definition for the Life Cycle of Biofuels

Figure 1 shows the DEA system definition for the proposed model, which describes the biofuel life cycle using two output variables, i.e., CO_2 offset (y_1), and net energy produced (y_2). The only input variable is the type of biofuel utilized (x_1).

DEA system

Inputs: x_1 = 1 unit (L)
Output: y_1 = (Net Energy (MJ/L)
y_2 = CO_2 Offset (%)
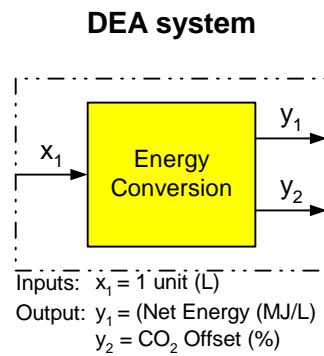
Figure 1. DEA system with input and outputs for the biofuel life cycle.

The DEA model was run by DEA-Solver-PRO v.5.0 by SAITECH, designed on the basis of the work by Cooper *et al* [15]. Table 1 depicts the data utilized in the DEA model, which was based on information found in the literature.

Table 1. Data used for the DEA model to compare various biofuels.

DMU			y_1 [Citation]	y_2 [Citation]
No	Biofuel Type	Feedstock	CO_2 Offset (%)	Net Energy (MJ/L)
1	Biodiesel	Soybean Oil	95 [31]	25.47 [32]
2	Ethanol (conventional)	Corn	30.5 [31]	5.88 [33]
3	Ethanol (cellulosic)	Switchgrass	68.6 [31]	23 [1]
4	Ethanol (cellulosic)	Hybrid Poplar	61.9 [31]	21 [3]
5	Ethanol (cellulosic)	Corn Residue	74 [31]	19.7 [3]
6	Algae biofuel	Algae	85 [34]	5.86 [35]

RESULTS & DISCUSSION

By solving the CRS DEA model, the technical efficiency (TE) score of each biofuel type (DMU) was obtained (Table 2).

Table 2. Technical efficiency and rank of each biofuel type.

DMU			TE	DMU
No.	Biofuel Type	Feedstock	Score	Rank
1	Biodiesel	Soybean Oil	1	1
2	Ethanol (conventional)	Corn	0.321	6
3	Ethanol (cellulosic)	Switchgrass	0.903	2
4	Ethanol (cellulosic)	Hybrid Poplar	0.824	4
5	Ethanol (cellulosic)	Corn Residue	0.778	5
6	Algae biofuel	Algae	0.894	3
Overall Average [TE] = 0.787				

The average technical efficiency was calculated to be 0.787 (Figure 1). Conventional ethanol (i.e., using corn as a feedstock) was the only biofuel that was below the average efficiency. This was expected, since the CO_2 offset of conventional corn ethanol was significantly lower compared to the remaining biofuels considered in the model. The low value of net energy obtained from algae was compensated by the significant reduction in CO_2 emissions it provides, making algae a "relatively efficient" option (89.4%). Among all biofuel types, soybean biodiesel was calculated to be the most efficient (100%) option. This was due to its relatively high CO_2 offset and high net energy gain.

Developing Criteria for Evaluating the Sustainability of Emerging Energy Technologies

Kurt A. Rosentrater[1]   **Elif Kongar[2]**

USDA, NC Agricultural Research Lab., USA   *University of Bridgeport, USA*

Abstract

Over the last decade, the interest in and production of biofuels has grown rapidly. Renewable transportation fuels can be produced from a variety of substrates, using various processing strategies. Feedstock preference and factory scale are geographic dependent. However, it is important to understand the implications of deploying these types of systems on a large scale, both throughout the U.S. as well as globally. To assess the sustainability of various biofuel options, it is crucial to evaluate their performance according to a number of attributes. This has been done using Life Cycle Assessment (LCA). Even so, comparisons among biofuel options are not easily accomplished. The objective of this study was to examine the efficiency of various biofuel options by using criteria which are important to LCA, namely, net energy balance and net carbon dioxide emission, and then compare them using Data Envelopment Analysis (DEA). Biofuel options examined included soy biodiesel, conventional corn ethanol, cellulosic ethanol (switchgrass, poplar, corn stover), and algae biofuel. This comparative approach is illustrated with a numerical example, which found that, using these criteria, soy biodiesel

[1] Kurt A. Rosentrater, Ph.D., USDA, ARS, North Central Agricultural Research Laboratory, 2923 Medary Ave., Brookings, SD, 57006, USA, Phone: (605) 693-3241; Fax: (605) 693-5240; E-mail: krosentr@ngirl.ars.usda.gov

[2] Elif Kongar, Ph.D., Departments of Mechanical Engineering and Technology Management, University of Bridgeport, 221 University Avenue, School of Engineering, 141 Technology Building, Bridgeport, CT 06604, USA, Phone: (203) 576-4379, Fax: (203) 576-4750, E-mail : kongar@bridgeport.edu
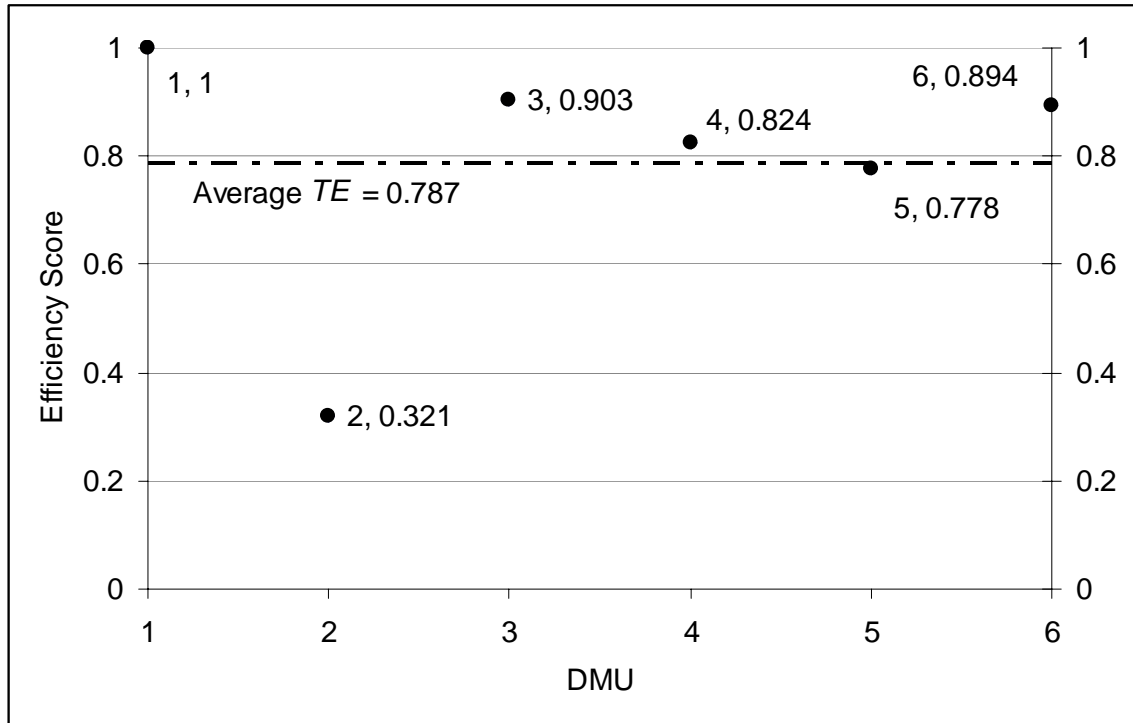
Figure 2. Efficiency score for each biofuel option (DMU).

In this study, a numerical case study has been presented as a first step towards a systematic interpretation of various biofuel options. Using this approach will eventually allow us to evaluate the overall efficiency of biofuels within a geographic area, for a variety of substrates and technologies, while emphasizing the relative importance of the different stages of the life cycle for each option. Another important extension of this study will be to include other criteria (beyond CO_2 offset and net energy). Some examples of these additional criteria are provided in Table 3. By adding such information, the role of LCA can be extended to reach out to decision makers in industry and government, as well as members of the general public, who may be able to relate to a physical system, but may have different criteria for evaluating the pros and cons of a

biofuel. Such additions will therefore enhance the practical relevance of LCA and provide guidance as the industry moves forward.

Table 3. Additional criteria for evaluating the sustainability of biofuel options.

Financial	Economic	Technological
Transportation cost	New job creation	Compatibility with current technology
Resale revenue	Governmental incentives	Peripheral equipment requirements
Processing cost	Governmental subsidy	New technology requirements
Peripheral equipment cost	Tax deduction	Technology upgrade rate
New technology implementation cost	Impact on local economy	Patent/copyright/trademark requirements
New technology equipment cost	Contribution to the national grid	Skill/training requirements
Maintenance/repair cost	Alternative energy utilization	Compatibility with national/international standards
Societal	**Processing related**	**Environmental**
Prioritization of agricultural by-products and residues	Gross energy required	Shortened crop rotation
Impact on local food supply	Gross energy produced	Reduction in green house has emission
Manufacturing safety	Cogeneration	Noise/light pollution
Impact on rural society	Quality of fuel	Gaseous/solid/liquid waste generation
Contribution to national independence	Ease in efficiency improvement	Losses in biodiversity
Self-sustainability	Net energy balance	Raw material consumption
Impact on local society	Density of maize cultivation	Water recycling/discharge rates

Our proposed approach to combining LCA and DEA has the following advantages:

- An explicit system definition in LCA terms makes the cycles and the indicators transparent. This is an important prerequisite for any comparison between energy

technologies, and, as a consequence, for determining the most relevant potentials for increasing the utilization efficiency of alternative energy sources.

- DEA can be applied in such a way that the Decision Making Units (DMUs) correspond to specific processes in the LCA. This enables us to define an efficiency indicator for each process in an LCA system.

- In contrast to conventional indicators, DEA is much more flexible, allowing for example, multi-dimensional efficiency definitions, which can simultaneously minimize or maximize multiple variables.

- Furthermore, criteria in DEA are not limited to information about physical flows alone, but can include any other variable relevant to the performance of a decision making unit, such as cost, health, or pollution.

DEA models do not require any *a priori* weights for either the input or the output variables of interest. However, the results are sensitive to the choice of dataset, and their reliability increases with sample size and, needless to say, accuracy of the data. Unfortunately, data quality and availability for new biofuels technologies are in general very low, and our quantitative results therefore need to be interpreted with caution. Indeed CO_2 offsets and net energy for each of these biofuels are somewhat disputable, as many estimates have been published over the years.

Lack of data also affected our model development, only allowing simplified system definitions. Therefore, our proposed model falls somewhat short in providing a better understanding of biofuels. The model would certainly benefit from a logistics system

infrastructure that includes additional criteria such as scale of production, production and transportation costs, geographical information, etc. But it is a first step toward developing a new approach to examining the efficiency of various biofuels.

And nevertheless, this paper has demonstrated the usefulness of combining LCA and DEA approaches for an evaluation of biofuels.

REFERENCES

[1]   Farrell, A.E. *Ethanol can contribute to energy and environmental goals (vol 311, pg 506, 2006). Science* 312(5781): p. 1748-1748, 2006.

[2]   Goldemberg, J. *Ethanol for a sustainable energy future. Science* 315(5813): p. 808-810, 2007.

[3]   Hammerschlag, R. *Ethanol's energy return on investment: A survey of the literature 1990 - Present. Environmental Science & Technology* 40(6): p. 1744-1750, 2006.

[4]   Larson, E. *A Review of LCA Studies on Liquid Biofuel Systems for the Transport Sector*. in *International Conf on GHG Control Technologies: Energy for Sustainable Development X(II)*, 2006, Trondheim, Norway.

[5]   Knight, W.A. and Sodhi, M. *Design for bulk recycling: analysis of materials separation. CIRP Annals - Manufacturing Technology* 49: p. 83-86, 2000.

[6]   Ulutas, B.H. *Determination of the appropriate energy policy for Turkey. Energy* 30(7): p. 1146-1161, 2005.

[7]   Junnila, S. *Alternative scenarios for managing the environmental performance of a service sector company. Journal of Industrial Ecology* 10(4): p. 113-131, 2006.

[8]   Geldermann, J. and Rentz, O. *Multi-criteria analysis for technique assessment - Case study from industrial coating. Journal of Industrial Ecology* 9(3): p. 127-142, 2005.

[9]   Ramanathan, R. and Ganesh, L.S. *Energy Resource-Allocation Incorporating Qualitative and Quantitative Criteria - an Integrated Model Using Goal*

Programming and Ahp. Socio-Economic Planning Sciences 29(3): p. 197-218, 1995.

[10]   Rosentrater, K. and Kongar, E. *Biofuel Coproducts: Modeling the Effects of Value-Added Processing and Logistics*. in *5th International Logistics and Supply Chain Congress*, 2007, Istanbul, Turkey.

[11]   Charnes, A., Cooper, W., and Rhodes, E. *Measuring the Efficiency of Decision-Making Units. European Journal of Operational Research* 2(6): p. 429–444, 1978.

[12]   Charnes, A. and Cooper, W.W. *Management Models and Industrial Applications of Linear Programming*. 1961, New York: John Wiley and Sons.

[13]   Charnes, A., Cooper, W.W., and Ferguson, R.O. *Optimal Estimation of Executive Compensation by Linear Programming. Management Science* 1(2): p. 138-151, 1955.

[14]   Charnes, A., Cooper, W.W., Lewin, A.Y., and Seiford, L.M. *Data Envelopment Analysis: Theory, Methodology, and Applications*. 1994, Boston: Kluwer.

[15]   Cooper, W.W., Seiford, L.M., and Tone, K. *Data Envelopment Analysis-A Comprehensive Text with Models, Applications, References and DEA-Solver Software*. 2000: Springer.

[16]   Sarkis, J. and Cordeiro, J.J. *Investigating technical and ecological efficiencies in the electricity generation industry: are there win-win opportunities? Journal of the Operational Research Society* 60(9): p. 1160-1172, 2009.

[17]   Sarkis, J. and Dijkshoorn, J. *Relationships between solid waste management performance and environmental practice adoption in Welsh small and medium-*

sized enterprises (SMEs). International Journal of Production Research 45(21): p. 4989-5015, 2007.

[18]   Sarkis, J. *The adoption of environmental and risk management practices: Relationships to environmental performance. Annals of Operations Research* 145: p. 367-381, 2006.

[19]   Sarkis, J., *Environmental benchmarking of the largest fossil-fueled electricity generating plants in the US*. in *Environmentally Conscious Manufacturing Iii*, Gupta, S.M., Editor, 2004, p. 182-192.

[20]   Revilla, E., Sarkis, J., and Modrego, A. *Evaluating performance of public-private research collaborations: A DEA analysis. Journal of the Operational Research Society* 54(2): p. 165-174, 2003.

[21]   Chien, T. and Hu, J.-L. *Renewable Energy and Macroeconomic Efficiency of OECD and Non-OECD Economies. Energy Policy* 35(7): p. 3606-3615, 2007.

[22]   Lozano, S. and Gutierrez, E. *Non-parametric frontier approach to modelling the relationships among population, GDP, energy consumption and CO2 emissions. Ecological Economics* 66(4): p. 687-699, 2008.

[23]   Mukherjee, K. *Energy use efficiency in US manufacturing: A nonparametric analysis. Energy Economics* 30(1): p. 76-96, 2008.

[24]   Criswell, D.R. and Thompson, R.G. *Data envelopment analysis of space and terrestrially-based large scale commercial power systems for Earth: A prototype analysis of their relative economic advantages. Solar Energy* 56(1): p. 119-131, 1996.

[25]   Moyer, L. and Gupta, S.M. *Environmental Concerns and Recycling/ Disassembly Efforts in the Electronics Industry. Journal of Electronics Manufacturing* 7(1): p. 1-22, 1997.

[26]   Isaacs, J.A. and Gupta, S.M. *Economic Consequences of Increasing Polymer Content on the U. S. Automobile Recycling Infrastructure. Journal of Industrial Ecology* 1(4): p. 19-33, 1997.

[27]   Boon, J., Isaacs, J.A., and Gupta, S.M. *Economic Impact of Aluminum Intensive Vehicles on the US Automotive Recycling Infrastructure. Journal of Industrial Ecology* 4(2): p. 117-134, 2000.

[28]   Boon, J.E., Isaacs, J.A., and Gupta, S.M. *End-of-Life Infrastructure Economics for "Clean Vehicles" in the United States. Journal of Industrial Ecology* 7(1): p. 25-45, 2003.

[29]   Kongar, E. and Rosentrater, K.A. *Data Envelopment Analysis Approach to Compare the Environmental Efficiency of Energy Utilization. International Journal of Green Computing* 1(2): p. 1-18, 2010.

[30]   Banker, R.D., Charnes, A., and Cooper, W.W. *Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis. Management Science* 30(9): p. 1078-1092, 1984.

[31]   McCarl, B.A. *BioEnergy Lifecycle Carbon Footprint, and Leakage*. in *Farm Foundation meeting on Lifecycle Carbon Footprint of Biofuels*, 2008, Washington, DC.

[32]   Pradhan, A., et al. *Energy Life-Cycle Assessment of Soybean Biodiesel*. in *Agricultural Economic Report*, 2009, United States Department of Agriculture.

[33]   Shapouri, H., Duffield, J.A., and Wang, M. *The energy balance of corn ethanol revisited. Transactions of the ASAE* 46(4): p. 959-968, 2003.

[34]   Stephenson, A.L., Kazamia, E., Dennis, J.S., Howe, C.J., Scott, S.A., and Smith, A.G. *Life-Cycle Assessment of Potential Algal Biodiesel Production in the United Kingdom: A Comparison of Raceways and Air-Lift Tubular Bioreactors. Energy & Fuels* 24: p. 4062-4077, 2010.

[35]   Chisti, Y. *Response to Reijnders: Do biofuels from microalgae beat biofuels from terrestrial plants? Trends in Biotechnology* 26(7): p. 351-352, 2008.

CHARACTERISTICS OF A MANUFACTURING STRATEGY IN A PRODUCTION NETWORK CONTEXT

Maricela Connie Arellano Caro

Teaching Assistant and PhD candidate at HEC Montreal
Operations and Production Management Department
3000, chemin de la Côte Sainte-Catherine
Montréal (Québec) H3T 2A7
Bureau 5.841
maricela-connie.arellano-caro@hec.ca
Tél. (514) 340-2058

ABSTRACT

A manufacturing strategy is defined by a pattern of decisions, both structural and infrastructural, which determine the capability of a manufacturing system. It also specifies how the system will operate to meet a set of manufacturing objectives which are consistent with overall business strategy. Yet, the different facets of manufacturing strategy have evolved since 70s integrating a variety of elements that are regrouped in three paradigms: competing through manufacturing, strategic choices in manufacturing, and best practice [1]. Moreover, manufacturing components – *e.g.* productions systems have been important triggers of social and economic change.

Over the last decades, the context where operations strategy acts has expanded, moving from a dominant view where a company competed through a factory to an extensive outlook where the global capabilities of manufacturing plants are emphasized [2][3].

Nowadays, companies develop production networks both as results and as drivers of globalization.

THE ADOPTION OF LEAN MANAGEMENT AND THE OUTSOURCING DECISION: AN EMPIRICAL STUDY

Ed Arnheiter, Quinnipiac University, 275 Mount Carmel Avenue, Hamden, CT, edward.arnheiter@quinnipiac.edu, (860) 523-5032.

Mary J. Meixell, Quinnipiac University, 275 Mount Carmel Avenue, Hamden, CT, mary.meixell@quinnipiac.edu, (203) 582-5206.

ABSTRACT

Outsourcing and lean management are two well-known and popular practices within the international business world. While their origins differ, both practices have come to represent broad strategic programs that seek to improve corporate competitiveness. In this research, we consider the relationship between the practice of lean and outsourcing. That is, do companies that adopt lean management practices outsource as extensively as those who don't? We use empirical data based on a survey of US manufacturers to assess the relationship between these variables, and analyze the data using contingency tables. Interestingly, we find limited evidence that implementing lean practices influences the outsourcing decision, suggesting a need for future research that investigates the contextual and moderating factors that drive this result.

Keywords: lean, outsourcing, manufacturing, empirical analysis.

INTRODUCTION

Outsourcing - the practice of moving work that had previously been performed internally to another firm, often in an international location - has had a significant effect on how manufacturing firms develop, produce and deliver products to their customers. Indeed, few practices in manufacturing have been as widely implemented as outsourcing. A 2007 survey shows that approximately 70% of manufacturers in the US outsourced at least one activity [1]. Improved computing, communications, and distribution capabilities have enabled this trend by providing manufacturers with increased availability to low cost labor and manufacturing capacity, both domestically as well as internationally.

Lean production, also known as lean management or simply "lean", is based primarily on practices developed at the Toyota Motor Corporation in Japan starting in the early 1950's. Specifically, the tools and philosophy of lean can be traced to the creation of the Toyota production system (TPS), a manufacturing approach pioneered by Japanese automotive engineers Taiichi Ohno and Shigeo Shingo [2]. The TPS is also associated with just-in-time (JIT) production methods, a key element of lean management, and for this reason the TPS represents the "gold standard" for proponents of lean management methods [3]. Another important element of lean management is variability reduction, including three primary types of variability; (1)

demand variability, (2) manufacturing variability, and (3) supplier variability. Supplier variability can include uncertainties in quality and delivery times. Partnerships and other forms of supply chain cooperation are commonly used to reduce supplier variability [3]. Lean management stresses the reduction and complete elimination of waste, and also emphasizes customer satisfaction, high quality, and comprehensive employee training and empowerment. Efforts to reduce waste are pursued through continuous improvement activities, often organized as focused, cross-functional teamwork activities known as kaizen events. Lean principles have been successfully used by many companies, including Danaher Corporation, United Technologies Corporation, Siemens Corporation, and Harley-Davidson. There is growing evidence that lean is being adopted by manufacturers to enable the continuation of production activities in-house, i.e. lean and outsourcing are increasingly viewed as substitutes. In several articles that recently appeared in the business press [4-8], firms across a variety of industries report that lean practices have improved the cost of producing internally enough to sway the balance in favor of keeping production in-house. In these cases, lean and outsourcing are treated as complementary practices. If costs can be reduced enough through lean implementation to be competitive, the activity may be retained in house; otherwise, it will be outsourced.

This phenomenon can be observed in practice in a variety of industries. As discussed in Elewaut et al. [9], banks in developing countries such as Chile strive to achieve a high level of cost efficiency by implementing lean practices to improve their processes, and then outsource what can be done more productively by another organization. For some activities, such as bill payment and check processing, it is the scale of the operation that is essential to cost efficiency, and firms that specialize have an advantage that can be acquired through outsourcing. For other activities that are commonly offered at traditional branches, a lean no-frills approach enables banks to improve cost efficiency while retaining many operations in-house. Ehret and Cooke [10] present similar arguments while exploring the extent to which the lean approach relates to outsourcing in the aerospace industry.

Jim Womack, author of Lean Thinking, argues that companies must consider these factors as "lean math" before making outsourcing decisions. He says that hidden costs, such as the costs of additional inventory of goods shipped over long distances, or the costs of engineer visits to help the new supplier, can make otherwise attractive looking deals unwise. Womack feels many companies foolishly only consider piece-price costs for production in a low-wage country, and then add in the cost of slow freight when making decisions [11].

The purpose of this research is to enrich this evidence with an empirical investigation into the relationship between the practice of lean and the degree to which companies outsource. In particular, we ask if firms that implement lean practices retain production and other activities in-house more frequently than those who do not. To help answer this question, we use empirical data from a survey of US manufacturers, and assess the relationship between these variables using contingency table analysis.

EMPIRICAL ANALYSIS

In this research, we use the Industry Week / Manufacturing Performance Institute (IW/MPI) 2009 Census of Manufacturers survey. This survey collects plant-level data on manufacturing

metrics, management practices, and financial results. The survey contains over a hundred variables that pertain to how the responding plant structures its operation and utilizes its capacity, equipment, information technology, human resources, and supply chain. The survey is sent to approximately 30,000 plant managers and financial officers, who are invited to respond either in paper format or online. The survey respondents come from a wide variety of industry segments as defined by the North American Industry Classification System (NAICS), and span twelve different industry supply chains. Other authors that have used this data include Ward and Zhou [12], Youssef et al [13], Bardhan et al [14], and Shah and Ward [15].

Outsourcing was evaluated for this study for individual functions, that is, companies were asked "To what degree have the following activities been outsourced?" The survey provided a list of eleven functions that might be outsourced (fabrication, assembly, electrical, design and/or R&D, maintenance/asset management, information technology, purchasing, transportation, customer service, human resources management and sales and marketing). For each activity, the respondent selected "all," "some," or "none" for each outsourcing category. As such, we evaluated each outsourcing category independently.

The "lean level" was estimated for survey participants based on the responses to two particular questions in the 2009 IW/MPI Manufacturing Study survey. One question was worded; "Please indicate which of the following improvement methodologies are followed at the plant," allowing for multiple responses. Six practices were listed, one of which was "Lean Manufacturing". A second question used for assessing "lean level" was worded; "What percentage of plant processes (across all departments) have been addressed with improvement methodology(ies)?" For the purposes of this study, any company that self-selected "Lean manufacturing" and more than 50% addressed was considered lean, and was then assigned the categorical variable "yes". All others were assumed to be "non-lean" and assigned the "no" variable.

RESULTS

Table 1 summarizes the number of "lean" and "non-lean" companies that responded under each level of outsourcing. For example, 23 companies that were categorized as "non-lean" also outsourced all of their fabrication work, while 52 outsourced some, and 83 none. The results of a cross tabulation and Chi-square analysis are shown in the table as well, and Pearson chi-square and P-values are listed for each outsourced function. In particular, we are interested in knowing if lean companies outsource these particular activities at different ratios than non-lean companies, referring to the null hypothesis that no association exists between lean level and outsourcing level. Using a significance level of alpha = 0.05, we reject the null hypothesis for the categories Maintenance (P-value = 0.000) and Transportation (P-value = 0.006). Of the companies that responded to the survey, lean companies outsource these activities at different levels than non-lean companies. For the other nine business activities, the results are less conclusive. In most cases, the differences between lean and non-lean companies were not statistically significant, at least not at the 0.05 alpha-level. At a higher alpha-level of 0.20, a moderate relationship between lean and outsourcing is suggested for the Assembly and Electrical activities.

Table 1
Summary of Analysis – Respondents for Each Function (Expected values in parenthesis)

Company Activities	Lean	Outsourcing Level and Number of Companies in Each				Pearson Chi-Square	P-Value
		All	None	Some	Total count		
Fabrication	No	23 (24.9)	52 (51.9)	83 (81.2)	158	0.399	0.819
	Yes	22 (20.1)	42 (42.1)	64 (65.8)	128		
Assembly	No	7 (4.4)	106 (108.9)	44 (43.7)	157	3.538	0.171
	Yes	1 (3.6)	91 (88.1)	35 (35.3)	127		
Electrical	No	33 (27.5)	71 (69.8)	51 (57.7)	155	4.239	0.120
	Yes	17 (22.5)	56 (57.2)	54 (47.3)	127		
Design/R&D	No	10 (8.2)	89 (92.0)	55 (53.9)	154	1.169	0.557
	Yes	5 (6.8)	80 (77.0)	44 (45.1)	129		
Maintenance	No	6 (4.3)	71 (86.8)	70 (55.9)	147	15.263	0.000
	Yes	2 (3.7)	92 (76.2)	35 (49.1)	129		
IT	No	16 (12.6)	73 (75.3)	67 (68.1)	156	2.180	0.336
	Yes	7 (10.4)	64 (61.8)	57 (55.9)	128		
Purchasing	No	3 (2.2)	145 (143.2)	5 (7.6)	153	2.712	0.258
	Yes	1 (1.8)	118 (119.8)	9 (6.4)	128		
Transportation	No	74 (81.8)	31 (21.8)	51 (52.4)	156	10.223	0.006
	Yes	76 (68.2)	9 (18.2)	45 (43.6)	130		
Customer Serv	No	3 (3.8)	138 (135.7)	11 (12.5)	152	0.881	0.644
	Yes	4 (3.2)	111 (113.3)	12 (10.5)	127		
HR	No	3 (2.2)	120 (124.7)	33 (29.1)	156	2.194	0.334
	Yes	1 (1.8)	107 (102.3)	20 (23.9)	128		
Sales/Marketing	No	2 (2.7)	111 (113.9)	40 (36.4)	153	1.376	0.503
	Yes	3 (2.3)	99 (96.1)	27 (30.7)	129		

When we look closer at Maintenance, we observe that lean companies outsourced this work less frequently than non-lean companies (71% versus 48%, respectively, of respondents falling into the "none" category). For Transportation, lean companies outsourced this function more frequently (58.5% versus 47.4% falling into the "All" category).

Overall, the results indicate a tendency for less outsourcing than initially expected. Indeed, glancing at the "None" column in Table 1 we observe many considerable percentages (as we concluded earlier, primarily independent of lean level). It is surprising how many companies do not outsource extensively, particularly when it comes to the following activities: Assembly (69.4% overall answered "none"), Design/R&D (59.7%), Maintenance (59%), Purchasing (93.6%), Customer Service (89%), Human Resources (79.9%), and Sales/Marketing (74.5%). The stacked bar graphs in Figure 1 illustrate and confirm these observations.

Figure 1
Summary of Survey Results
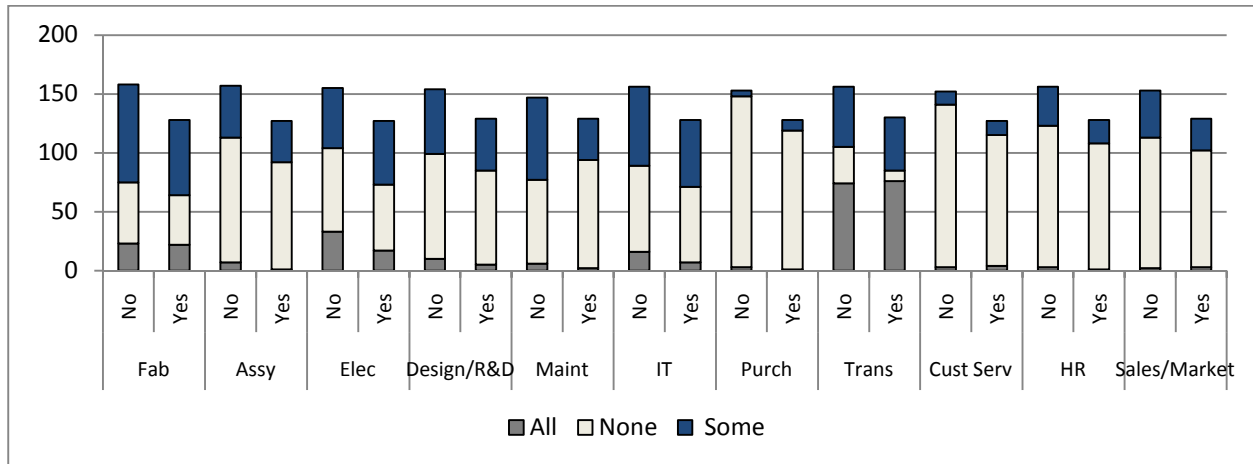
CONCLUSIONS

In this research, we investigated the degree to which outsourcing practice differs in lean and non-lean companies. We draw on plant-level data collected by the IW/MPI Census of Manufacturers survey, and analyze the data with a contingency analysis and chi-square test. The results suggest that of the eleven activities considered in the survey, the implementation of lean practices influences the decision to outsource transportation and maintenance, but none of the production activities (i.e. fabrication, assembly and electrical).

As such, this analysis provides limited evidence that implementing lean practices influences the outsourcing decision, suggesting a need for future research on factors that drive this result. Both contextual and moderating factors may influence the relationship between outsourcing and lean implementation. For example, whether or not a company considers production to be a core competency would influence whether the activity was improved through lean practices, or outsourced. Other factors that might influence the decision to outsource or implement lean would be plant size, age, the nature of the product and process, and product-process volume-variety level. Also, labor cost and content, unionization, customer locale, and supplier integration may be moderating factors. Future research may investigate these factors in the outsourcing decision process.

REFERENCES

[1]   Manufacturing Performance Institute, *Census of US Manufacturers*. 2008, Industry Week.

[2]   Inman, R.R., *Are You Implementing a Pull System by Putting the Cart Before the Horse?* Production and Inventory Management Journal, 1999. **40**(2): p. 67-71.

[3]   Arnheiter, E.D. and J. Maleyeff, *The Integration of Lean Management and Six Sigma.* The TQM Magazine, 2005. **17**(1).

[4]   Abelson, J., *Made in New Bedford: a suit designer retools; Abboud factory bucks US trend*, in *Boston Globe*. 2007. p. A.1.

[5]   Moules, J., *Body Shop spin-off stays alive*, in *Financial Times*. 2005. p. 20.

[6]   Langer, F., *Lean manufacturing challenges outsourcing*. Machine Design, 2007. **79**(4): p. 71.

[7]   Stafford, J., *Manufacturers 'getting lean' to help lower costs*, in *Daily Oklahoman, The (Oklahoma City, OK)*. 2004. p. 1B.

[8]   Cole, J., *D'Addario Company, Inc. Overhauls Company to Keep Manufacturing in the U.S.*, in *PR Newswire (USA)*. 2007.

[9]   Elewaut, T., P. Lindenboim, and D.L. Scokin, *Chile's lesson in lean banking*. The McKinsey Quarterly, 2003(3): p. 6.

[10]   Cooke, P. and Ehret, O., *Conceptualising aerospace outsourcing: Airbus UK and the lean supply approach*. International Journal of Technology Management, 2009 or 2010. **50**(3,4): p. 300.

[11]   Anon., *Executives Urged to Do 'Lean Math' to Calculate the True Cost of Offshoring*, in *PR Newswire*. 2007: New York.

[12]   Ward, P. and H. Zhou, *Impact of Information Technology Integration and Lean/Just-In-Time Practices on Lead-Time Performance**. Decision Sciences, 2006. **37**(2): p. 177.

[13]   Al-Ghamdi, S.M., T. Dolan, Koch, J.V., and Youssef, M.A., *Plant location, ISO 9000 certifications and quality management practices*. The International Journal of Quality & Reliability Management, 2006. **23**(8): p. 944.

[14]   Bardhan, I., Lin, S., and Mithas, S., *Performance Impacts of Strategy, Information Technology Applications, and Business Process Outsourcing in U.S. Manufacturing Plants*. Production and Operations Management, 2007. **16**(6): p. 747.

[15]   Shah, R. and Ward, P.T., *Lean manufacturing: context, practice bundles, and performance* Journal of Operations Management, 2003. **21**(2): p. 129-149.

SCHEDULING AND SEQUENCING SENSOR-EMBEDDED END-OF-LIFE PRODUCTS

Onder Ondemir, Northeastern Univ., Boston, MA, 02115, (617) 373-7635, ondemir.o@neu.edu
Surendra Gupta[1], Northeastern Univ., Boston, MA 02115, (617) 373-4846, gupta@neu.edu

ABSTRACT

Task scheduling and sequencing is known to be one of the domains of the hardest problems and has been studied for many years by many authors. However, scheduling in an end-of-life (EOL) management environment is considerably different than it is in a manufacturing or assembly situation. In this paper, a scheduling and sequencing problem is defined and investigated in an advanced refurbishment-to-order and disassembly-to-order (ARTODTO) setting with sensor embedded products. A hybrid-genetic algorithm is developed and shown to perform well on large-size problems.

For illustration purposes, a numerical example of water heater ARTODTO system is presented.

INTRODUCTION

With the increase of public awareness in environmental issues, manufacturers are forced to control their end-of-life products (EOLPs) by both regulations and customers. This phenomenon introduced the concept of product recovery. Product recovery, by decreasing the use of virgin resources, enables environment conscious manufacturing, thus allow companies to comply with the government regulations and customers' expectations. Product recovery also gives companies an opportunity to make use of the remaining value in EOLPs, instead of disposing of the value trapped in the product, which, in turn, provides economical benefits. Remaining value of an end-of-life product (EOLP) can be recovered through various operations, viz. disassembly, recycling and refurbishment. Recovered materials, components and products can be used to re-manufacture products or sold as is at lower costs as long as the cost of recovery is less than that of brand new (never-used) items. All these benefits create a new market for recycled materials, used components and products. Thus, traditional supply-demand balance problem is reproduced in a reverse setting. In fact, the main difference between the two is that, in the traditional manufacturing planning, preset processing operations and guaranteed availability (or release dates) of desired parts are considered. This certain environment is controlled by applying strict delivery rules on suppliers. In the reverse case, however, EOLPs are supplied by the users and companies have no control over their usage. Therefore, neither the rate nor the condition of returns is certain.

Condition of the returns may show high degrees of variability because, in an EOLP, (a) components may be non-functional, (b) replaced with newer versions, or (c) missing (removed). Moreover these cases are mostly up to the product user, just as the usage pattern is. In other words, every single EOLP has its own unique properties based on the circumstances (e.g., working temperature, vibrations, shocks, run cycles, maintenance, failures, etc.) that it has been

---

[1] **Corresponding author**. Laboratory for Responsible Manufacturing (LRM), Department of MIE, Northeastern University, 360 Huntington Avenue, Boston, MA, 02115 USA

through. To this end, generic disassembly plans applied to all EOLPs are, in fact, cause the disassembly of nonfunctional or unneeded components, which increases the cost of the system. If the condition of components can be known prior to the disassembly, unnecessary disassembly of a non-functional or unneeded component can be avoided. Sensors embedded in products make this information available prior to disassembly [7].

Advanced refurbishment-to-order and disassembly-to-order (ARTODTO) concept, involves disassembly and refurbishment of sensor embedded products (SEPs) in order to fulfill remaining-life-time based (sophisticated) component and product demands as well as material demands while some system criteria are minimized. Using sensor based information on the condition, availability and version of the components in EOLPs, optimal ARTODTO plans and schedules can be obtained.

In this paper, disassembly planning and scheduling problem is modeled as an integer non-linear program. As an alternative method, a hybrid-genetic algorithm is developed and shown to perform well on larger-size problems. For illustration purposes, a numerical example of water heater ARTODTO system is presented.

BACKGROUND

ARTODTO system can be considered as an extension to the traditional disassembly-to-order (DTO) system. DTO is a process, in which EOLPs are disassembled in order to fulfill the demand for materials and reusable components. This operation is usually performed on a disassembly line similar to traditional assembly line. Compared to the assembly situation, disassembly operations are performed with under the presence of uncertainty. The reader is referred to [2-4, 11-15, 22-25] for detailed information about disassembly processes, problems and solution methods. Emerging information technologies, such as sensors and radio-frequency identification (RFID), can help address the uncertainty-borne problems.

A sensor detects the changes in the value of various measures such as temperature, pressure, vibration and converts the value into a signal to have it recorded. A sensor-embedded product (SEP) is equipped with sensors that monitor the product during its life-cycle and record the product life-cycle data. By enabling life-cycle monitoring and facilitating real time usage data acquisition, sensors help predict component failures during the product usage [6] and enhance the end-of-life (EOL) management. Using the information collected by sensors, existence, types, conditions and remaining lives of components in an EOLP can be determined without any inspection and disassembly operations. The value of information in the context of a firm that faces uncertainty with respect to demand, product returns, recovery yield, and capacity utilization was explored by Ketzenberg [10].

Sensors may be used in conjunction with the radio frequency identification (RFID) technology. RFID technology has recently gained importance in closed loop supply chain operations, including reverse logistics, disassembly and remanufacturing, as a means of communication and data storage. With the ability to provide fully automatic and instantaneous item-level product information, RFID technology partially eliminates the lack of information in the optimization of product remanufacturing process [26]. RFID technology can be introduced as an enabler of product lifecycle management, by enhancing the traceability of the product throughout its value chain via automatic identification, and facilitating the integration of product lifecycle information and knowledge [1]. Parlikad and McFarlane [19] discussed how RFID based product

identification technologies can be employed to provide the necessary information and showed the positive impacts on product recovery decisions. When sensor and RFID technologies are combined, all static (bill of materials, sale date, shipping date, warranty terms, model number, etc.) and dynamic (usage patterns, working conditions, run lengths, maintenance history, removed, replaced or upgraded parts) information about an EOLP becomes readily available as soon as the EOLP reaches at the disassembly/refurbishment facility.

Availability of life-cycle information brings clarity to the EOL operations and can be used to determine the remaining life of the components [20]. Herzog *et al.* [5] proved the advantage of using condition-based data in remaining life prediction. Ondemir and Gupta [16] proposed a mathematical disassembly-to-order (DTO) model utilizing life-cycle data in order to fulfill remaining life time based (sophisticated) component demands assuming the availability of sensor based information. In its follow up papers [17-18], the authors extended the model in order to meet the sophisticated product demands by using repair option, and presented economic justification for establishing advanced disassembly-to-order systems in which SEPs are disassembled in order to fulfill sophisticated demands. Ilgin and Gupta [9] presented a quantitative assessment of the impact of SEPs on the various performance measures of a kanban-controlled disassembly line using simulation analysis. Authors showed that SEPs not only provide significant reductions in the total system cost, but also increase the revenue. Also in [8], they investigated a sensor-embedded air conditioner disassembly line with disassembly precedence relationships among components and obtained the same result.

PROBLEM DEFINITION

In the ARTODTO system, a number of sensor and RFID embedded EOLPs are either disassembled, recycled or refurbished. Refurbishment activities are carried out to meet the demand for products and may require spare parts. Necessary spare parts may come from two sources, viz. disassembled EOLPs and outside suppliers (procurement option). The purpose of disassembly is to recover a limited set of reusable components. Recovered components are (a) sold to satisfy the customers' component demand, (b) used in-house to assist refurbishment operations or (c) recycled to meet customers' material demands. It is assumed that cost of recovery is always less than outside procurement cost. Therefore, the problem is to find a set of EOLPs to cannibalize and refurbish and sequence them so that the remaining-life based demands are satisfied while some system criteria, viz. total disassembly cost, total refurbishment cost, total procurement cost, and total work-in-process holding (WIP) cost, are minimized.

Significance and difficulty of this particular problem is due to the very nature of disassembly and refurbishment operations. Refurbishment operations cannot be performed unless the necessary spare parts are available at the station. This situation forces the system to schedule all disassembly tasks before refurbishment tasks so that all necessary spare parts become available at the station. Disassembly, however, is an inventory-expanding operation and causes extra WIP holding costs. Therefore, it is important to find a sequence that minimizes the total WIP holding cost while making sure that necessary spare parts (reusable component) are extracted before attempting to refurbish an EOLP. Customers' remaining life requirements are also introduced to the system. This condition adds another constraint for the properties of necessary parts.

The ARTODTO planning and scheduling problem is modeled to minimize the sum of total disassembly cost, total refurbishment cost, total recycling cost, total procurement cost, total buffer cost and total completion time cost for a single machine case. Total completion time cost

is calculated as weighted completion cost $(1//\sum W_i C_i)$ where weights are the holding costs per product per unit time.

MATHEMATICAL MODEL

Preliminaries

There are j ($j \in J$) components on a product (core) and there exist a precedence relationship scheme among them.

An EOLP in the inventory can be disassembled for component recovery, refurbished for product recovery or left untouched. If an EOLP is disassembled, then every single component is taken out. In other words complete disassembly is considered. Each recovered component can be used to fulfill component demand, reused to repair other products for product demand fulfillment. All demands are assumed to occur based on remaining life time of the components and products. Therefore, recovered components and products are placed in different bins (life-bins) based on their remaining life time. In order to make the model more realistic, it is assumed that the demands occur for those products whose remaining life time is within a certain time range (e.g., less than a year). This is because the remaining life time is a continuous random variable (there would have been infinitely many life-bins to accommodate all components). Thus, life-bins are created considering time ranges mentioned above. It has to be noted that any component can be evaluated in any life bin as long as the remaining life time of the component is larger than the lower bound of the life bin's life time range.

An EOLP may contain both operable and non-operable (broken, zero-remaining-life) components. Broken parts require less attention in disassembly. Operable components are reused for refurbishment operations or sold as parts. These components have a remaining life and must be treated carefully. Thus, unit disassembly cost for an operable component is higher than that of a broken component. EOLPs may be short of some components. It is assumed that operable, non-operable and missing components are known prior to disassembly operations and the remaining life of the operable components can be calculated by means of sensor based life-cycle data.

Also, it is obvious that a product functions no longer than its component having the shortest remaining life. This means that reconfiguring an EOLP during repair is a way of remanufacturing certain a product that has a certain remaining life. Thus, depending on the target product life-bin, some components in an EOLP may be remaining-life-time deficit. The model considers replacing remaining-life-time deficit components with appropriate components. Index k indicates the order of any operation.

The Objective Function

The objective of ADTTOM is to minimize the total cost (Z), viz. the sum of total disassembly cost (*TDC*), total refurbishment cost (*TRPC*), total outside procurement cost (*TOPC*), total recycling cost (TRC) and total holding cost (THC). Therefore the total cost function can be given as follows:

$$Z = TDC + TPRC + TOPC + TRC + THC$$

Total Disassembly Cost (TDC)

TDC is incurred by completely disassembled EOLPs. Cost of disassembly activities in repair process will be covered in total repair cost. Then, *TDC* can be formulated as:

$$TDC = \sum_{\substack{i \in I \\ j \in J}} \bar{x}_{ik} \left(a_{ij} cd_j + f_{ij} cb \right)$$

Total Refurbishment Cost (TRPC)

Repair activity comprises the disassembly of broken and remaining-life-time deficit components, and assembly of required ones. Hence, TRPC is defined as the sum of the costs of these two activities (i.e., total repair disassembly cost and total repair assembly cost). One should also note that the disassembly of a component requires extraction of components listed in precedence relationship. Therefore, the related cost should reflect the cost of disassembling and reassembling all preceding components. Hence, this situation is taken into consideration by introducing binary variable rp_{ij} that indicates all components that have to be taken care of while repairing EOLP *i*. Therefore:

$$TRPC = \sum_{\substack{i \in I \\ j \in J}} rp_{ij} \left(a_{ij} \left(cd_j + ca_j \right) + f_{ij} \left(cb + ca_j \right) + mis_{ij} ca_j \right) \tag{1}$$

Total Outside Procurement Cost (TOPC)

TOPC is a function of unit purchase cost and the total number of procured component *j* whose remaining life is within the ranges of component life-bin *b*. Mathematical expression for *TOPC* is as follows:

$$TOPC = \sum_{\substack{j \in J \\ b \in B}} c_{jb} l_{jb} \tag{2}$$

Total Recycling Cost (TRC)

Material demand is met by recycled broken components. If there are not sufficient broken components then operable components from recovered component life-bins are recycled to fulfill the demand. *TRC* can be expressed as follows:

$$TRC = \sum_{j \in J} cr_j \left(\sum_{i \in I} f_{ij} \left(\sum_{k \in K} \left(\bar{x}_{ik} + \bar{y}_{ik} \right) \right) + \sum_{b \in B} r_{jb} \right) \tag{3}$$

Total Holding Cost (THC)

Total holding cost of an EOLP is calculated by multiplying the per-minute-inventory cost (ci_i) and the time spent in queue (w_{ik}).

$$THC = ci_i \sum_{\substack{i \in I \\ k \in K}} w_{ik} \left(x_{ik} + y_{ik} \right) \tag{4}$$

Constraints

The mathematical model is subject to the following constraints:

First of all, an EOLP in the inventory can be disassembled, repaired or left untouched.

$$\sum_{k \in K} \bar{x}_{ik} + \bar{y}_{ik} \leq 1, \quad \forall i \tag{5}$$

$$\sum_{i \in I} \bar{x}_{ik} + \bar{y}_{ik} \leq 1, \quad \forall k \tag{6}$$

In this study, complete disassembly is considered. Since a functional component can be placed in only one life-bin after disassembly, related constraints can be expressed as follows:

$$\sum_{b \in B} x_{ijbk} = \bar{x}_{ik} a_{ij}, \quad \forall i, j, k \tag{7}$$

Equation below assures that an EOLP is repaired to produce only one product and that product is evaluated in only one product life-bin.

$$\sum_{m \in M} y_{imk} = \bar{y}_{ik}, \quad \forall i, k \tag{8}$$

Sophisticated product demand is satisfied by repaired EOLPs. Thus, the number products in product life-bin m (which are produced by repairing EOLPs) must be equal to the corresponding product demand. Hence,

$$\sum_{\substack{i \in I \\ k \in K}} y_{imk} = dp_m, \quad \forall m \tag{9}$$

Component demand is satisfied by recovered plus the procured operable components that meet certain remaining life criteria. Recovered components are obtained from disassembled and refurbished EOLPs. During the refurbishment process, operable unnecessary components and remaining-life-time deficit components are taken out and placed in proper component life-bins. Therefore

$$\sum_{\substack{i \in I \\ k \in K}} \left(x_{ijbk} + def_{ijbk} \right) - \sum_{\substack{i \in I \\ m \in M \\ k \in K}} rep_{imjbk} - l_{jb} - r_{jb} \geq dc_{jb}, \quad \forall b, j \tag{10}$$

Non-functional, missing, and remaining-life-time deficit components must be filled in with components having a remaining life time that is eligible for producing a model t for product life-bin m. Therefore

$$\sum_{\{b \in B \mid b \geq m\}} rep_{imjbk} = y_{imk} \left(f_{ij} + mis_{ij} + dfc_{imj} \right), \quad \forall i, j, m \tag{11}$$

Replacement of a component can be taken from only one bin.

$$\sum_{\substack{b \in B \\ m \in M \\ k \in K}} rep_{imjbk} \le 1, \quad \forall i, j \tag{12}$$

Equation below forces the precedence relationship.

$$y_{imk}\left(f_{ij} + mis_{ij} + dfc_{imj}\right) \le rp_{ij'}, \quad \forall i, m, k, \left\{j, j' | j' \in P_j\right\} \tag{13}$$

Extracted remaining-life-time deficit components must be placed in only one bin.

$$\sum_{\{b \in B | cin_{ijb} = 1\}} def_{ijbk} = \sum_{m \in M} y_{imk} def_{imj}, \quad \forall i, j, k \tag{14}$$

$$\sum_{\{b \in B | cin_{ijb} = 0\}} def_{ijbk} = 0, \quad \forall i, j, k \tag{15}$$

Component inventory is calculated as follows;

$$q_{jbk} = \begin{cases} 0 & , k = 1 \\ q_{jb(k-1)} + diss_{jb(k-1)} - \sum_{\substack{i \in I \\ m \in M}} rep_{imjb(k-1)} & , k > 1 \end{cases} \tag{16}$$

Where;

$$diss_{jbk} = \sum_{i \in I}\left(x_{ijbk} + def_{ijbk}\right), \quad \forall b, j, k \tag{17}$$

The equation below makes sure that at any moment, component inventory is sufficient to assist a repair activity.

$$\sum_{\substack{i \in I \\ m \in M}} rep_{imjbk} \le q_{jbk}, \quad \forall b, j, k \tag{18}$$

Processing time of each EOLP is calculated as follows;

$$\tau_{ik} = \sum_{j \in J}\left(rp_{ijk}\left(a_{ij}\left(pd_j + pa_j\right) + f_{ij}\left(pb_j + pa_j\right) + mis_{ij} pa_j\right) + \bar{x}_{ik}\left(a_{ij} pd_j + f_{ij} pb_j\right)\right), \quad \forall i, k \tag{19}$$

Waiting time of each task is also calculated as follows.

$$w_k = \begin{cases} 0 & , \forall i, k = 1 \\ w_{(k-1)} + \sum_{i \in I} \tau_{i(k-1)} & , \forall i, \{k | k \ge 2\} \end{cases} \tag{20}$$

HYBRID GENETIC ALGORITHM

Genetic algorithm (GA) is a heuristic method commonly utilized to find good solutions to computationally difficult problems, such as NP-complete or non-linear problems. In many cases, GA performs more efficiently than exhaustive search. In disassembly sequencing and scheduling area, Seo *et al.* [21] proposed a GA to obtain optimal disassembly sequence considering both

economical and environmental constraints. Their algorithm however could visit infeasible solutions during the evolution steps, viz. crossover and mutation phases. GAs are generally not suitable for the system where precedence relations exists. Therefore, authors usually tailor and alter the algorithm for their needs.

In this paper, GA is combined with a heuristic sequencing algorithm. The overall approach consists of two steps. Firstly, an ARTODTO plans are generated using the GA. Secondly, the operation sequences are obtained using a weighted shortest processing time (WSPT) based heuristic algorithm. The total costs are calculated and saved as the fitness of the solutions, best of which is updated in each iteration. The proposed approach is depicted in figure 1.
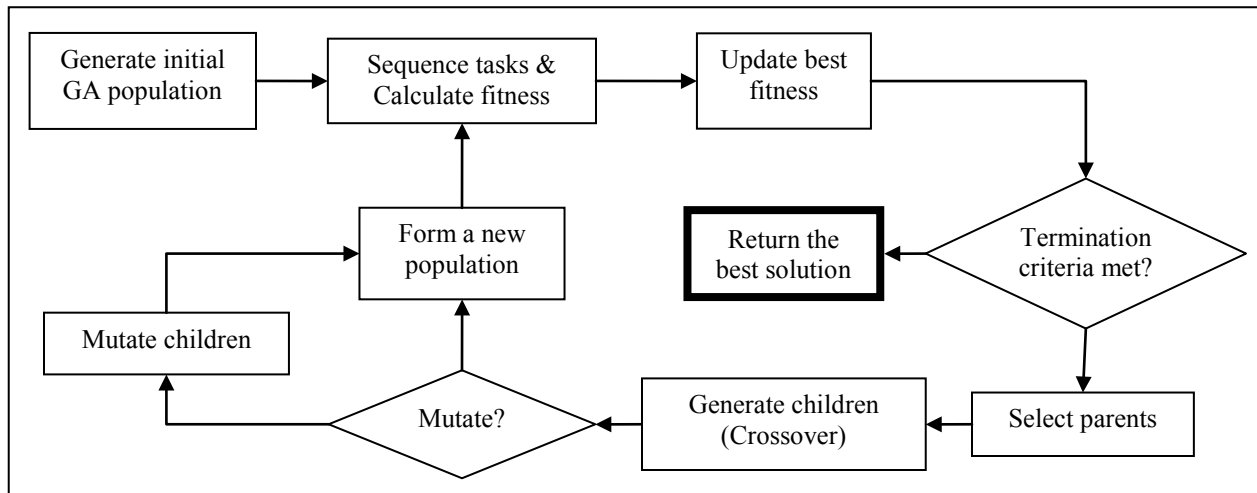
Figure 1: Flow chart of the proposed heuristic algorithm.

GA population is a group of chromosomes. A chromosome consists of a number of genes that hold the values of all variables. Each chromosome coincides with a solution point on the problem domain, thus leads to an objective function value. In the proposed approach, the number of genes is equal to the number of EOLPs and each gene takes the value 0 or an integer value less than or equal to 4. 0 indicates that the EOLP is left untouched. The values 1, 2, and 3 indicate that the EOLP is refurbished to satisfy a product demand in remaining-life-bins 1, 2, and 3, respectively. When the gene takes the value "4", the corresponding EOLP is disassembled for its components.

A feasibility preservative crossover method is also developed to increase the performance of the proposed method.

NUMERICAL EXAMPLE

To illustrate the methodology, an example involving device-embedded water heater is considered. There are 7 components that the model deals with. List of components, precedence relationships, procurement prices and demands are given in Table 1.

Three remaining-life-bins are defined for components. First life-bin holds those components having a remaining-life time of two years or less, second life-bin holds those components whose remaining-life are between two and three years. The last bin holds the other components (having 3 years or more remaining life). Same remaining life time ranges are used to define three product life-bins.
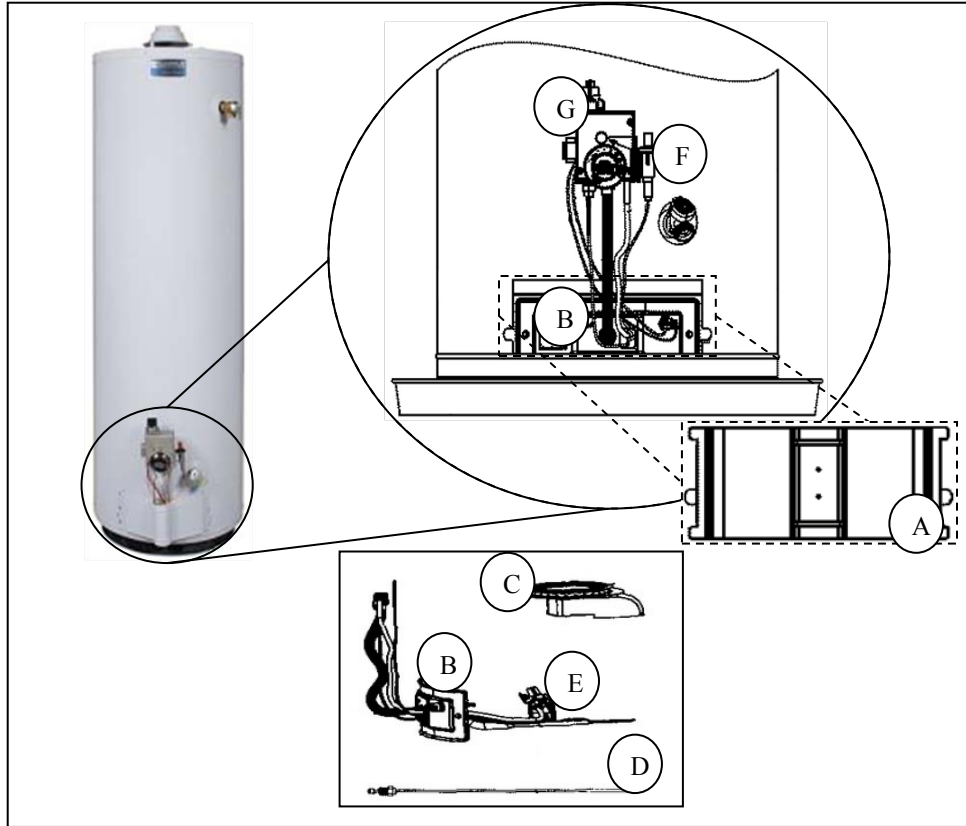
Figure 2: Components of the water heater.

Table 1: Components, precedence relationships and problem related data.

Part	Code	Predecessor	Disassembly Cost ($)	Assembly Cost ($)	Procurement Price ($)			Demand		
					Bin 1	Bin 2	Bin 3	Bin 1	Bin 2	Bin 3
Outer Door	A	-	0.50	0.50	13.99	13.99	13.99	0	0	0
Door +Manifold	B	-	2.00	2.00	20.99	20.99	20.99	0	0	0
Burner	C	A, B	0.15	0.15	12.99	27.49	41.99	0	6	6
Thermocouple	D	A, B, C	0.25	0.25	7.99	7.99	7.99	1	3	15
Pilot Assembly	E	A, B, C,D	1.00	1.00	9.99	17.99	25.99	3	6	12
Igniter	F	-	0.15	0.15	6.99	11.74	16.49	3	3	12
Gas Valve	G	A, B, F	1.00	1.00	74.99	134.49	193.99	0	3	6

Disassembly of a non-functional component is $0.10.

RESULTS

The mathematical model was programmed in LINGO programming language. LINGO 11.0 is used to evaluate the model. LINGO's nonlinear solver was unable to solve the problem optimally. In order to compare the proposed hybrid-genetic algorithm method to the nonlinear solver, several instances of the problem were considered. It has to be noted that, LINGO solver could not even reach to a local optimal solution for any instances. In those cases, solver was stopped after 10 minutes and best solution so far was taken. In fact, in most of the cases, no feasible solution was obtained. Those instances are marked with "None". Results are shown Table 2. Programs were run on a host computer featuring 3.20 GHz Intel Core i5 processor and 4GB memory.

Table 2: Computational results and comparison of the methods.

Instance #	Number of EOLPs	Product Demand			LINGO		Hybrid-GA	
		Bin 1	Bin 2	Bin 3	Total Cost ($)	Time (Sec)	Total Cost ($)	Time (Sec)
1	10	2	2	3	3090.74	600.00	3034.20	8.15
2	20	3	3	4	4606.59	600.00	2882.40	21.29
3	30	6	6	9	None	600.00	5907.40	31.58
4	50	10	10	15	None	600.00	11556.00	66.71
5	70	14	14	21	None	600.00	21244.00	184.82
6	100	20	20	30	None	600.00	38276.00	482.51

CONCLUSIONS

In this paper, a mixed integer non-linear model for a scheduling and sequencing problem under the advanced refurbish-to-order, disassembly-to-order (ARTODTO) system setting is presented and compared with a hybrid genetic algorithm (GA) developed for the same problem. Results show that the hybrid GA found significantly better solutions in all instances. As a future improvement, embedded weighted shortest processing time (WSPT) based sequencing logic may be replaced with another search algorithm.

As a conclusion, the hybrid GA developed for this problem performs well for the tested instances. However, improvements in the heuristic method can be done.

REFERENCES

[1] Cao, H., *et al.*, "RFID in Product Lifecycle Management: A Case in the Automotive Industry," *International Journal of Computer Integrated Manufacturing,* vol. 22, pp. 616 - 637, 2009.

[2] Gungor, A. and Gupta, S. M., "Disassembly Sequence Plan Generation Using a Branch-and-Bound Algorithm," *International Journal of Production Research,* vol. 39, pp. 481 - 509, 2001.

[3] Gungor, A. and Gupta, S. M., "Disassembly Line in Product Recovery," *International Journal of Production Research,* vol. 40, pp. 2569 - 2589, 2002.

[4] Gupta, S. M. and Taleb, K., "Scheduling Disassembly," *International Journal of Production Research,* vol. 32, pp. 1857-1866, 1994.

[5] Herzog, M. A., *et al.*, "Machine and Component Residual Life Estimation through the Application of Neural Networks," *Reliability Engineering & System Safety,* vol. 94, pp. 479-489, 2009.

[6] Ilgin, M. A. and Gupta, S. M., "Comparison of Economic Benefits of Sensor Embedded Products and Conventional Products in a Multi-Product Disassembly Line," *Computers & Industrial Engineering,* vol. 59, pp. 748-63, 2010.

[7] Ilgin, M. A. and Gupta, S. M., "Recovery of Sensor Embedded Washing Machines Using a Multi-Kanban Controlled Disassembly Line," *Robotics and Computer-Integrated Manufacturing,* vol. 27, pp. 318-334, 2011.

[8] Ilgin, M. A. and Gupta, S. M., "Evaluating the Impact of Sensor-Embedded Products on the Performance of an Air Conditioner Disassembly Line," pp. 1-18, in press.

[9] Ilgin, M. A. and Gupta, S. M., "Performance Improvement Potential of Sensor Embedded Products in Environmental Supply Chains," in press.

[10] Ketzenberg, M., "The Value of Information in a Capacitated Closed Loop Supply Chain," *European Journal of Operational Research,* vol. 198, pp. 491-503, 2009.

[11] Kim, H. J., *et al.*, "A Lagrangean Heuristic Algorithm for Disassembly Scheduling with Capacity Constraints," *Journal of the Operational Research Society,* vol. 57, pp. 1231-1240, 2006.

[12] Kongar, E. and Gupta, S. M., "A Multi-Criteria Decision Making Approach for Disassembly-to-Order Systems," *Journal of Electronics Manufacturing,* vol. 11, pp. 171-83, 2002.

[13] Kongar, E. and Gupta, S. M., "Disassembly to Order System under Uncertainty," *Omega,* vol. 34, pp. 550-561, 2006.

[14] Lambert, A. J. D., "Optimal Disassembly of Complex Products," *International Journal of Production Research,* vol. 35, pp. 2509-23, 1997.

[15] Meacham, A., *et al.*, "Optimal Disassembly Configurations for Single and Multiple Products," *Journal of Manufacturing Systems,* vol. 18, pp. 311-322, 1999.

[16] Ondemir, O. and Gupta, S. M., "End-of-Life Decisions Using Product Life Cycle Information," in *ASME International Mechanical Engineering Congress and Exposition (IMECE 2008)*, Boston, MA, 2008.

[17] Ondemir, O. and Gupta, S. M., "Cost-Benefit Analysis of Sensor Embedded Products Based Disassembly-to-Order System," presented at the Northeast Decision Science Institute 2009, Uncasville, CT, 2009.

[18] Ondemir, O., *et al.*, "Disassembly-to-Order Model for Sensor Embedded Products," in *the Eighth International Conference on EcoBalance*, Tokyo, Japan, 2008.

[19] Parlikad, A. K. and McFarlane, D., "RFID-Based Product Information in End-of-Life Decision Making," *Control Engineering Practice,* vol. 15, pp. 1348-1363, 2007.

[20] Rugrungruang, F., "An Integrated Methodology for Assessing Physical & Technological Life of Products for Reuse," Doctor of Philosophy Doctoral Dissertation, Life Cycle Engineering & Management Research Group, School of Mechanical and Manufacturing Engineering, The University of New South Wales, New South Wales, Australia, 2008.

[21] Seo, K. K., *et al.*, "Optimal Disassembly Sequence Using Genetic Algorithms Considering Economic and Environmental Aspects," *International Journal of Advanced Manufacturing Technology,* vol. 18, pp. 371-380, 2001.

[22] Taleb, K. and Gupta, S. M., "Disassembly of Multiple Product Structures," *Computers & Industrial Engineering,* vol. 32, pp. 949-961, 1997.

[23] Taleb, K., *et al.*, "Disassembly of Complex Products with Parts and Materials Commonality," *Production Planning and Control,* vol. 8, pp. 255-269, 1997.

[24] Tang, Y., *et al.*, "Disassembly Modeling, Planning, and Application," *Journal of Manufacturing Systems,* vol. 21, pp. 200-217, 2002.

[25] Veerakamolmal, P. and Gupta, S. M., "Analysis of Design Efficiency for the Disassembly of Modular Electronic Products," *Journal of Electronics Manufacturing,* vol. 9, pp. 79-95, 1999.

[26] Zhou, W. and Piramuthu, S., "Remanufacturing with RFID Item-Level Information," in *Exploring the Grand Challenges for Next Generation E-Business*, R. Sharman, *et al.*, Eds., 1st ed Berlin/Heidelberg: Springer 2011, pp. 96-104.

A STUDY OF THE RELATIONSHIP AMONG TEAM GOVERNANCE MECHANISMS, TEAM LEANING ACTIVITIES AND R&D PERFORMANCE

Lung-Far Hsieh, Dept. of Business Administration, Da-Yeh University, Taiwan, R.O.C., & Dept. of Business Administration, Chung Yuan Christian University, Taiwan, R.O.C.
Tel: 886-3-265-5113, E-mail: lungfar@cycu.edu.tw

Ping-Chuan Lee, Dept. of Business Administration, Chung Yuan Christian University, Taiwan, R.O.C.
E-mail: supercute.apple@gmail.com

ABSTRACT

By using a survey of 163 cross-functional R&D teams, this study explores the effects of three governance mechanisms (including team autonomy, span of integration, and professional training) on CFTs' leaning activities (consisting of discussions and knowledge sharing), and the effects of CFTs' leaning activities on the teams' R&D performance. The results find that: (1) All the three R&D governance mechanisms chosen by this study (Team autonomy, Span of integration, and Professional training) have significant impacts on CFTs' leaning activities (Discussion and Knowledge sharing); (2) Knowledge sharing has a positive impact on R&D performance; but (3) Discussions among CFT members have no effects on the teams' R&D performance. For R&D practice, the results implicate that R&D managers do have some sorts of R&D governance mechanisms that can facilitate certain types of team leaning activities. However, R&D managers should pursue the team learning activities that are truly beneficial for R&D performance and be wary of some superficial forms of team learning activities that have no effect on R&D performance.

Keywords: Cross-functional team (CFT), Team governance, Team learning activities, R&D performance

INTRODUCTION

Once, R&D was considered to be a unique, creative and unstructured process that was difficult, if not impossible, to control. The control techniques used in other business functions were considered inappropriate for the R&D function because control was supposed to harm creativity, and because of the uncertainty of R&D outcome. Therefore, in many companies, the governance of R&D was limited to setting budgets and periodical peer-reviews focusing on technological achievements (Roussel, Saad, and Erickson, 1991). But, in recent, there has been a major change in management's attitude toward the administration of R&D works. Although managers still assume that R&D processes have several characteristics that distinguish them from other business processes, they no longer consider that this means they are unmanageable. As a result, there is a growing acceptance of the need to manage R&D process and, as part of this, to measure R&D performance.

Furthermore, studies in organizational research have demonstrated that innovation in organizations is a collaborative process, and that groups have potential for innovation (Tjosvold et al., 2004). Teams, especially cross-functional teams (CFTs), have been used as the prime approach to facilitate the collaborative process of learning and knowledge creation that are vital in R&D practice. Since a cross-functional team consists of experts with different specialties, its members often bring with significant differences in perspectives and interests (Carlile, 2002). These knowledge barriers can stifle the exchange of different specialties among team members and, then, impede team learning within cross-functional teams. Therefore, the R&D governance mechanisms must be able to help

bring the transformation and combination of different knowledge domains of members of an R&D team, in other words, to promote team learning activities. Hence, in this study, we treat team members' learning activities (consisting of discussions and knowledge sharing) as mediate variables connecting a firm's R&D governance mechanisms and its R&D performance. Because R&D management governance mechanisms have a broad array of control arrangements, this study will focus on three governance mechanisms including team autonomy, span of integration, and professional training. We chose these mechanisms because they have been prevailing in R&D practice and are important for facilitating team members' interaction. The study's research framework is depicted in Figure 1.
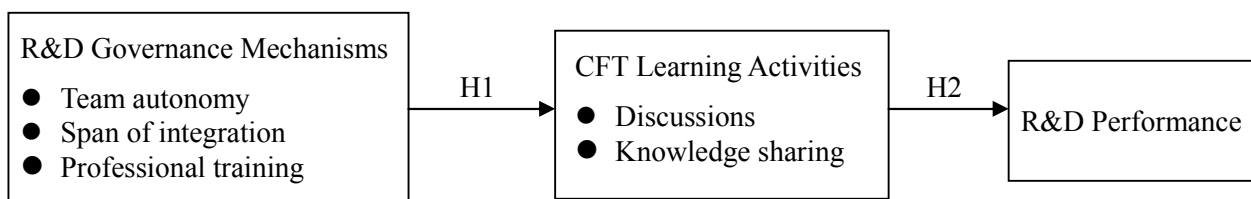
Figure 1   Research Framework

RESEARCH VARIABLES

Governance mechanisms of R&D

A firm's R&D governance management usually focuses on how the organization manages its R&D activities, and uses these aspects as variables of governance mechanisms of R&D. Since cross-functional teams (CFTs) are used as the main arrangement of implementing a specific R&D project in most firms, the R&D management practice is usually reduced to how to govern the firm's CFTs. Argyris (1992) has long argued that attributes of team governance style help establish an organizational environment necessary for deep (double-loop) learning to occur. Once team resources such as office space, technology, budget, and personnel have been assembled, many of these resources can be managed within the team through relatively well-understood administrative coordination mechanisms (Malone and Crowston, 1994). Nonetheless, in research of CFT management and R&D performance, there are few studies focusing on the influence of R&D governance mechanisms, therefore there is no commonly applicable dimensions of R&D governance mechanisms. This study selects some prevailing practice in R&D management from previous studies as the aspects of R&D's governance mechanisms, including autonomy, span of integration, and professional training

Cross-functional Team (CFT) learning activities

Learning is the cognitive process of acquiring new skills, knowledge, and worldviews (London et al., 2005) and teams carry out an increasing amount of work in an organization and are a basic venue of organizational learning (Edmondson, 1999). Senge (1990) also suggested that teams are the

fundamental learning unit in an organization. In this study, team learning is defined as a process that a team takes actions, obtains and reflects upon feedbacks, and makes changes to adapt or improve confronting circumstances (Edmondson, 2002). It is manifested when a team welcomes challenging assignments, is willing to take risks on new ideas, and works on tasks that require considerable skills and knowledge (Bunderson and Sutcliffe, 2003; Edmondson, 1999). In this research line, this study uses discussion and knowledge sharing as constructs of team learning.

R&D Performance

In R&D literature, there are many scholars strived to develop some forms of measuring R&D performance. There are roughly two types of R&D performance, one is focused on the efficiency of R&D process such as the implementation of listed tasks, accordance with time schedule, and/or aligned budget; the other is focused on the effectiveness of R&D results such as customers' satisfaction with the new products and/or services, accordance with pre-set goals, numbers of new patents.

In empirical research on R&D performance, Markham and Griffin's (1998) study provided a thorough view based on data from the 1995 PDMA study of the best practices in product development to investigate NPD performance at the program, firm, and project levels. The survey included four project-level measures of NPD cycle times and nine measures of overall performance. Factor analysis divided these eight overall performance measures into three factors (program performance, firm-Level success rates, and firm-level sales and profitability). The ninth measure (number of new products introduced) is statistically unrelated to any other measure and is thus treated separately. But, in this study the R&D performance is discussed at the project-level. We mainly follow a more parsimonious model provided by Griffin and Page's (1996) study that developed a set of performance measures for determining the project-level success of product development and the overall success of product development programs at firms.

HYPOTHESES

The relationship between R&D governance mechanisms and CFT learning activities

1. The relationship between Team autonomy and CFT learning activities

Team autonomy is defined as the group's level of collective control over critical decisions about its objectives, resources, design, and process, relative to others inside or outside the organization (Langfred, 2000). In research on NPD project teams, how much power such as authority for decision-making and priority for company resources is delegated from the organizational management to the project team is a major concern because it will confine the abilities and willingness of team members to carry out integrating works (Clark and Wheelwright, 1992). The decision-making autonomy offers team members a sense of control (Lawler, 2001), high level of autonomy motivates team members to make decisions that are in the best interests of the project even when these decisions may conflict with the agenda and interests of external stakeholders (such as functional departments of the organization). Recently, Daft (2007) argued that teams should have the freedom to think creatively and respond flexibly to new challenges that arise; and Hass (2006) revealed that it can enhance the processing and sense making capabilities of project teams by freeing the team members to focus more energy on these activities. Therefore, we hypothesize that

H1-1: Team autonomy has positive influences on CFT's learning activities.

H1-1-1: Team autonomy has a positive influence on discussions.

H1-1-2: Team autonomy has a positive influence on knowledge sharing

2. The relationship between Span of integration and CFT learning activities

The integration of an organization is always a major concern of organizational research. In R&D, monitoring is aided by the practice that researchers normally work in groups of peers (Hoecht, 2004) and physically work within the spatial boundaries of their organization. However, recent research on innovation (Verburg, Orrt, and Dicke, 2006) emphasizes the coordinated process of innovation in a network of partners, and that coordination is often attained by systemic integration (with key suppliers and customers) and parallel development (of components or modules of the innovation). Thus, whether an organization is disposed to expand or narrow the span of integration is an important aspect of the organization's governance mechanisms that will confine the team's learning scope.

H1-2: Span of integration has positive influences on CFT's learning activities.

H1-2-1: Span of integration has a positive influence on discussions.

H1-2-2: Span of integration has a positive influence on knowledge sharing.

3. The relationship between Professional Training and CFT learning activities

Senge (1990) argued that an organization's members have one learning agenda in common: to understand and optimize the learning capacity of an enterprise in order to enable a more optimal creation and application of new knowledge. Also, in R&D management practice, there is a growing emphasis on a more formal type of relationship between a senior member and a novice of an organization that facilitates learning processes in an organization (Young and Perrewe, 2000). Because organizations' members are typically with different experiences and professional expertise, how to transfer senior members' experience and professional expertise to junior members is an imperative of R&D management. Thus, the existence of a professional training mechanism to help junior members acquire professional expertise is the foundation of organizational learning.

H1-3: Professional training has positive influences on CFT's learning activities.

H1-3-1: Professional training has a positive influence on discussions.

H1-3-2: Professional training has a positive influence on knowledge sharing.

The relationship between CFT learning activities and R&D performance

1. The relationship between Discussion and R&D performance

Discussion of team-level is regarded as team members' group dialogue focusing on teamwork, which may be accompanied with objects such as documents, drawings, artifacts, etc. It is a ubiquitous, surface form of interpersonal face-to-face interaction among team members. Nevertheless, some scholars (Karau and Williams, 1993) had pointed out that there is a tendency for people to expend less effort when working collectively than when working individually, such as in some cases known as social loafing. Moreover, because there is a tacit behavioral norm that demands team members to engage in team discussion in NPD teams, team members may take part in discussion but never touch critical issues. Edmondson's (2002) empirical study reports that some teams do have a lot of

discussion but fail to reveal any reflection on the R&D's important issues. In sum, both the theoretical and empirical studies on discussion among team members are uncertain and mixture. In this study, we thought that because discussion is ubiquitous in most CFTs, it has became an inherent attribute of CFTs. Hence, it is difficult for merely using the existence of discussion among team members as a criterion to distinguish good CFTs from bad ones. The following hypothesis that depicts a positive influence is only for statistic convenience.

H2-1: Discussion has a positive influence on R&D performance

2. The relationship between Knowledge sharing and R&D performance

Knowledge sharing through interactions between team members from different functions helps they understand the content of others' works. It is through the process of sharing information and experiences with the group that the members learn from each other, and have an opportunity to develop themselves personally and professionally (Lave and Wenger 1991). London et al. (2005) pointed out that team members learn about each other's unique strengths and areas for potential contribution to the team and develop interpersonal congruence by sharing their own thoughts and feeling (Polzer et al., 2004). Lesser and Storck (2001) also found that group members shared the idea of common theme from different points of view whereby to stimulate the innovation of products and services. From a behavioral view, for conceiving that others can understand his/her perspective, a team member will be more willing to express his/her ideas. This process may trigger virtuous cycle to enhance a shared cognitive context based on common experience and mutual understanding (Dougherty, 2001). Conversely, difficulties in knowledge sharing and integration were identified as key factors hindering project outcomes (Walz et al., 1993), and information withholding within an organization can hinder that organization's ability to transfer best practices, learn from mistakes, stimulate innovation, or benefit from strategic alliances (Haas and Park, 2010). Therefore, knowledge sharing is fundamental to collaborative working in teams and has been shown to be associated with producing knowledge to create innovative products and resolving conflicts cooperatively (Parker and Axtel, 2001). In sum, it will promote R&D performance, therefore

H2-2: Knowledge sharing has a positive influence on R&D performance.

METHODS

The survey

A questionnaire survey was designed to explore the relationship among these research variables. The targeted respondents of the questionnaires are R&D members in companies belonging to technology industry (electronic components and computer, communication, and optical products manufacture), traditional manufacture industry (mechanical equipments and metallic material manufacture), Internet and information services, and some national research institutes. Their R&D works may encompass applied research, new product development, improvement of products, and new technique. In this study, we sent 330 copies of questionnaires by e-mail and post and received 185 returned questionnaires. Among them, 22 were abandoned for their incomplete answers. This resulted 163 valid questionnaires and a respondent rate of 49.4%.

Measures of research variables

In this section, we present the measure content of research variables.

A. Governance mechanisms of R&D

- Team autonomy

1. The members of R&D do tasks as a team partners and integrate their supplier and customers. Its measure items are:
2. The company provides sufficient resources for R&D teams.
3. The content of the R&D tasks is decided by teams.
4. The time schedule of the R&D tasks is decided by teams.
5. Most of the R&D decision-making is made by teams.

- Span of integration

This construct is to estimate the scope of involving stakeholders that may contribute their knowledge, information, and practice. Its measure items are:

1. R&D activities have the form of cross-functional cooperation (For instance, such as the department of technology manufacturing, finance, quality control, purchasing and other personnel involved in R & D).
2. In R&D activities, we integrate the opinions of other departments.
3. In R&D activities, we integrate the opinions of clients.
4. In R&D activities, we integrate the opinions of suppliers.

- Professional training

This construct is to estimate if there are training programs and mentoring system to help members acquire professional expertise. Its measure items are:

1. For new staffers of R&D, there are some learning systems and guidelines to help them.
2. For new staffers of R&D, there are specific experienced colleagues assigned to assist them to learn some knowledge about R&D tasks
3. In R&D activities, I learn much from experienced coworkers.

B. CFT learning activities

- Discussion:

This construct is to estimate CFT members' group dialogue that focuses on their teamwork, which may be accompanied with objects such as documents, drawings, artifacts, etc.

1. I often discuss the R&D content with team members.
2. We usually spend a lot of working time on routine meetings
3. Team members often conducted informal meetings to discuss work problems.
4. I will ask other members' opinions when I have problems.

- Knowledge sharing:

This construct is to estimate the exchange of task-related information and experiences between team members from different functions.

1. There is seamless sharing of R&D information for all of R&D members.
2. I speak out when I had different opinion.
3. The experiences of R&D members are instructive to me
4. I trust other members' expertise relating to their specific knowledge domains.

C. R&D performance

For estimate CFT's R&D performance, we choose indices from R&D research literature relating to satisfaction and project objectives as the indicators of R&D performance. They are:

1. I am satisfied with R&D results in past three years.
2. I think the clients are satisfied with our R&D output.
3. I think this R&D team had generated some valuable innovations
4. I think this R&D team has achieved the project's objectives
5. I think the R&D project's results will benefit our company financially

Statistical analyses

The data quality of each research variables is fist tested by using Cronbach's alpha reliability test and then by confirmative factory analysis (CFA) method. After that, multiple linear equations programming method is used to test the research model. AMOS 5.0 software was used to implement CFA tests for research variables and to test the research model and its hypotheses

RESULTS

Reliabilities and Validities

The reliability of each research variable was calculated using Cronbach's coefficient alpha. The values of Cronbach's coefficient alphas of all research variables are greater than 0.7 - the acceptable level, raging from 0.787 to 0.891. This suggests the reliabilities of all constructs in this study are acceptable.

We also used CFA (Confirmatory factor analysis) to test the discrimination validities of these research variables. According to the research framework, we conducted three CFA tests. The fist is to test the study's exogenous variables that are variables of R&D Governance mechanisms (Team autonomy, Span of integration, and Professional training); the second is to test variables of CFT leaning activities including Discussion and Knowledge sharing; and the final is to test R&D performance. All the statistic parameters of model fit for these three CFA models are within acceptable area and indicate that the data are accordant with the study's measure model of research variables. Furthermore, because the correlation coefficients between variables of R&D Governance mechanisms (Team autonomy, Span of integration, and Professional training) are 0.44, 0.47 and 0.50 (all below 0.7), this indicates that these variables have good discrimination validities. Similarly, because the correlation coefficient between variables of CFT learning activities (Discussion and Knowledge sharing) is 0.63 (bellows 0.7), this indicates these two variables have good discrimination validity.

Testing the hypothesized model (H1 and H2)

As illustrated in Figure 2, the parameters of model fit are in acceptable area and indicate the hypothesized model is accordant with data from the survey sample. And, because all the standardized coefficients of the causal paths between variables of R&D Governance mechanisms (Team autonomy, Span of integration, and Professional training) and variables of CFT learning activities (Discussion and Knowledge sharing) are positive and statistically significant, thus all the sub-hypotheses of hypothesis H1 (H1-1-1, H1-1-2, H1-2-1, H1-2-2, H1-3-1, and H1-3-2) are supported. Because the standardized coefficient of the causal path between Knowledge sharing and

R&D performance is positive and statistically significant, thus hypothesis H2-2 is supported; but he standardized coefficient of the causal path between Discussion and R&D performance is not statistically significant, thus hypothesis H2-1 is rejected. Therefore, hypothesis H2 is only partially supported.
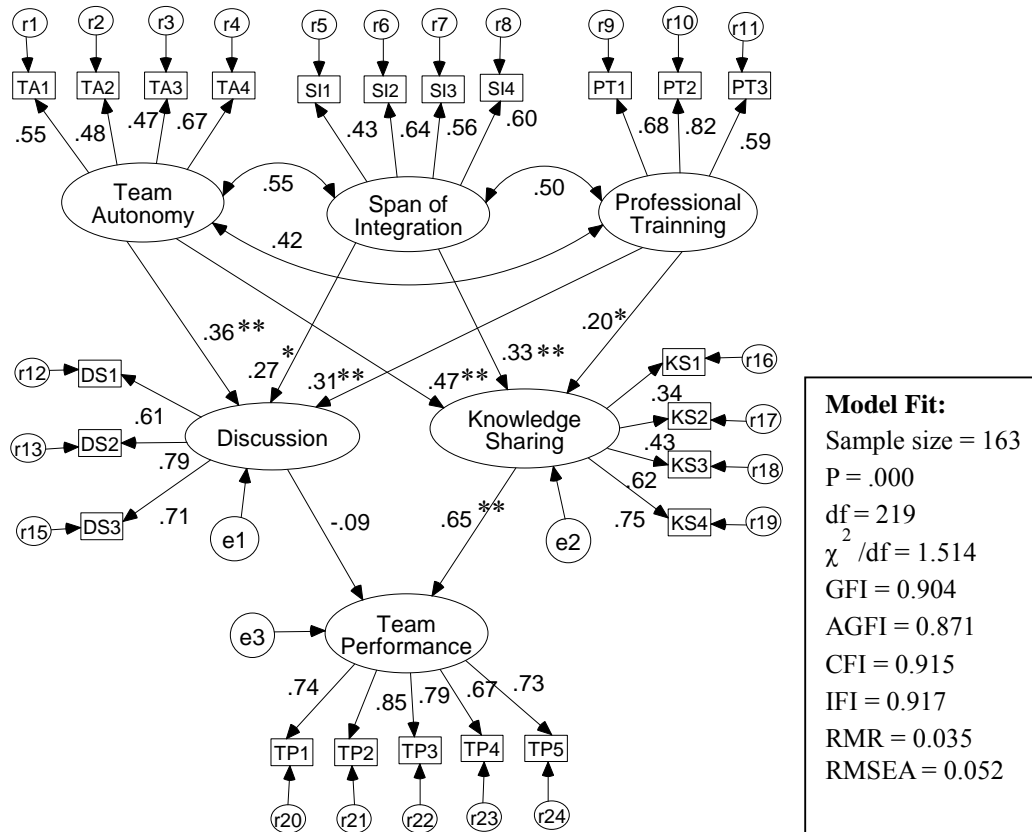
Figure 2   Results of Model Fit

CONCLUSIONS

1. The results of this study support our research hypothesis H1 that all the three R&D governance mechanisms chosen by this study (Team autonomy, Span of integration, and Professional training) have significant impacts on CFT's leaning activities (Discussion and Knowledge sharing). This consists with many previous studies (Daft, 2007; Verburg, Orrt, and Dicke, 2006; Young and Perrewe, 2000). Specifically, all the causal paths between three R&D governance mechanisms and two CFT learning activities are significantly positive adds a solid empirical evidence to this line of research that emphasizes the importance of team learning.

2. The rejection of H2-1 indicates that discussions among CFT members have no effects on the team's R&D performance. This result is consistent with some studies that revealed that the existence of discussion among team members could be only a superficial form of team members' communication (Edmondson, 2002) and in some situations may result in the area of social loafing

(Karau and Williams, 1993). But this result does not imply that we should avoid or discourage discussions among team members. It just that we should emphasize the collaborative aspects of team learning that underlie team members' discussions instead of pursuing superficial team learning activities, as Tsoukas (2009) revealed that dialogue is productive depending on the extent to which participants engage relationally with one another.

3. In contrast with the results of H2-1, the strongly support of H2-2 (the coefficient causal path is 0.65) emphasizes that the effects of team leaning stem from its deeper aspects, in this case the mutually sharing of members' information, experience, and idea. Basically, to uncover the genuine forms of team leaning activities is a common goal in many R&D studies (Dougherty, 2001; Parker and Axtel, 2001; and Haas and Park, 2010). In this study, we chose knowledge sharing to represent this form of genuine team learning. Although this research model seems some what rudimentary in comparison with some more sophisticated models in this line of research (Markham and Griffin, 1998); but by contrasting with a potentially superficial form of team learning (Discussion) within the research frame work, this study successfully disclosed what types of team learning are beneficial to promote R&D performance. For R&D practice, the results of this study can bring some pristine implications that R&D managers do have some sorts of R&D governance mechanisms that can facilitate certain types of team leaning activities; but R&D managers should focus on those team learning activities that are truly beneficial for R&D performance and be wary of some superficial forms of team learning activities that have no effect on R&D performance.

REFERENCES

[1] Argyris, C. 1993. *Knowledge for Action: A Guide to Overcoming Barriers to Organizational Change*, CA: Jossey-Bass.

[2] Bunderson, J.S. 2003. "Team member functional background and involvement in management teams: Direct effects and the moderating role of power centralization," *Academy of Management Journal*, 46(4): 458-474.

[3] Carlile, P. 2004. "Transferring, translating, and transforming: an integrative framework for managing knowledge across boundaries," Organization Science, 15(5): 555-568.

[4] Clark, K.B. and Wheelwright, S.C. 1992. "Organizing and Leading "Heavyweight" Development Teams," *California Management Review,* 34(3): 9-28.

[5] Dougherty, D. 2001. "Reimagining the Differentiation and Integration of Work for Sustained Product Innovation," *Organization Science*, 12(5): 612-631.

[6] Edmondson, A. 1999. "Psychological safety and learning behavior in work teams," *Administrative Science Quarterly*, 41(2): 350-383.

[7] Edmondson, A.C. 2002. "The local and variegated nature of learning in organizations: A group-level perspective," *Organization Science*, 13(2): 128-146.

[8] Haas, M.R. and Park, S. 2010. "To Share or Not to Share? Professional Norms, Reference Groups, and Information Withholding Among Life Scientists," *Organization Science*, 21(4): 873–891.

[9] Hass, M.R. 2006. "Acquiring and applying knowledge in transnational teams: The roles of cosmopolitans and locals, " *Organization Science*, 17(3): 367-384.

[10] Karau, S.J. and Williams, K.D. 1993. "Social loafing: A meta-analytic review and theoretical integration," *Journal of Personality and Social Psychology*, 65(4): 681-706.

[11] Langfred, C.W. 2000. "The paradox of self-management: Individual and group autonomy in work groups," *Journal of Organizational Behavior*, 21(5): 563-585.

[12] Lave, J.C. and Wenger, E. 1991. *Situated Learning: Legitimate Peripheral Participation.* New York: Cambridge University Press.

[13] Lawler, E.J. 2001. "An affect theory of social exchange," *American Journal of Sociology*, 107(2): 321-352.

[14] London, M.; Polzer, J.T. and Omoregie, H. 2005. "Interpersonal congruence, transactive memory, and feedback processes: An integrative model of group learning," *Human Resource Development Review*, 4(2): 114-135.

[15] Malone, T.W. and Crowston, K. 1994. "The interdisciplinary study of coordination," *Computer Surveys*, 26(1): 87-119.

[16] Parker, S.K. and Axtell, C.M. 2001. "Seeing another viewpoint: Antecedents and outcomes of employee perspective taking," *Academy of Management Journal*, 44(6): 1085-1100.

[17] Polzer, J.T. 2004. "How Subgroup Interests and Reputations Moderate the Effect of Organizational Identification on Cooperation," *Journal of Management*, 30(1): 71–96.

[18] Rentsch, J.R. and Klimoski, R. 2001. "Why do great minds think alike? Antecedents of team member schema agreement," *Journal of Organizational Behavior*, 22: 107–120.

[19] Rico, R.; Sanchez-Manzanares, M.; Gil, F. and Gibson, C. 2008. "Team implicit coordination process: A team knowledge-based approach," *Academy of Management Review*, 33(1): 163–184.

[20] Senge, P.M. 1990. *The Fifth Discipline: The art and practice of the learning organization*. New York: Doubleday.

[21] Tjosvold, D., Tang, M.M.L. and West, M. 2004. "Reflexivity for team innovation in China," *Group & Organization Management*, 29(5): 540-559.

[22] Tsoukas, H. 2009. "A Dialogical Approach to the Creation of New Knowledge in Organizations," *Organization Science*, 20(6): 941–957.

[23] Walz, D.B.; Elam, J.J. and Curtis, B. 1993. "Inside a software design team: Knowledge acquisition, sharing, and integration," *Communications of the Association for Computing Machinery*, 36: 63-77.

HOUSEHOLD INCOME EFFECT ON CONSUMER PRICE PERCEPTION OF FREQUENCY VERSUS DEPTH PRICE DISCOUNT STRATEGY

Sereikhuoch Eng, University of Rhode Island, USA

Send correspondence to Sereikhuoch Eng, College of Business Administration, University of Rhode Island, Ballentine Hall, Kingston, Rhode Island 02881, +1-209-782-7339, sereikhuoch_eng@my.uri.edu.

ABSTRACT

This paper proposes to examine household income effect on consumer's perception of frequency vs. depth price promotion. The results are expected to show that low income consumer group is more likely to invest time and effort in finding lower prices, resulting in a depth effect for a dichotomous price structure. On the other hand, a frequency effect is expected to be found under a non-dichotomous price structure among low income consumers due to the higher complexity which induces the use of frequency heuristic in price information processing. High income consumers, however, are less price sensitive and less likely to engage in price information search and processing, resulting in them being indifferent across all conditions.

KEYWORDS

Price discount strategy, Household income effects, frequency vs. depth pricing

INTRODUCTION

As recessionary pressures continue to affect consumers' spending, deals are crucial to winning traffic for many retailers, grocers and restaurant operators. Deals and price promotions have since made the central theme of many newspapers and television broadcast news. Consumers' relentless effort to find ways to save also reached its peak in March 2009 when the giant search engine Google reported that searches for the term "coupons" surpassed those for the pop princess "Britney Spears." Similarly searches for subjects such as "free" and "discount" have shown sharp trajectories [10].

A quick browse of the newspaper, and television commercial breaks alike, reveals the many hefty discounts and weekly price promotions happening at all levels from retail stores, to grocery chains, to restaurant and fast-food chains. Marketers implement a vast use of either hefty deep discounts or very frequent price cuts to lure consumers to their stores. Merchandise retailers have had to cut their luxury goods price up to 80 percent just to turn consumers' heads towards their stores [11]. For grocers, however, more frequent price promotions than deep price cuts are used to lure cash-strapped consumers to their stores [5]. While many experts speculate that price promotions of either form are not going to end soon, the key question for many marketers is whether to choose between frequent or deep price discount strategy, and, whether either strategy has a significantly bigger impact on consumers. The impact that marketers are hoping to see as a result from their choice of price discount strategy is that consumers will continue to favor their offers/stores as having cheaper price.

It is whether frequent or deep price discounts that lead to favorable consumer's price perception that is the topic of this research proposal. Extant research seeks to understand the relationship between price and consumer's choice. This body of research appears in various fields from economics to psychology, sociology and marketing (e.g., [6,7,9, 13]). Studies by Alba, Broniarczyk, Shimp, and Urbany [1], Alba, Mela, Shimp, and Urbany [2], and Lalwani and Monroe [8] examine consumer perception of frequent and deep price discounts. In their work, they explain the prevalence of one price discount strategy over another through a number of theories such as stimulus complexity, cognition and frequency heuristics. While this existing work offers valuable insights into the matter in question, household income has not been considered as a crucial factor interacting with the way consumers perceive price discount strategy. During tough economic times, it is of benefit to marketers to understand how different price discount strategy affects consumers at different income levels so that effective strategy could be chosen.

The objective of this research proposal is to answer the broad question of how household income affects consumer's price perception of frequent versus deep discount strategy. The proposal begins by reviewing the relevant body of literature, establishing the hypotheses and then

presenting planned experiments, planned analyses, and anticipated results. Finally, implications and directions for future research are offered.

CONCEPTUAL BACKGROUND

Existing research in the fields of economics and marketing suggests that household income has an effect on consumer's likelihood and willingness to engage in price search, processing and recall. This body of research supports the premise for this paper in looking at household income as a factor influencing consumer's perception of different price discount strategies.

The Economics of Information: Opportunity Cost, Price Sensitivity, and Household Income

The economic theory of opportunity cost has been widely used to explain why some consumers engage in certain activities and others do not. The basis of opportunity cost lies in the fact that an individual chooses to engage in an activity if he/she derives positive gain from such activity; that is the gain from engaging in such activity outweighs the cost of foregoing another competing activity. With regards to household income, economists posit that higher income consumers have higher opportunity cost than lower income consumers. And as higher income consumers are faced with higher opportunity cost, they are less price sensitive than lower income consumers.

George Stigler's [12] *The Economics of Information* suggests that lower income consumers are more apt to search for price promotion than higher income consumers. He suggests that consumers who engage in price promotion information search do so because they believe that the costs of the search will be more than offset by the expected savings on present and future purchases. Similarly, in the studies by Beatty and Smith [3] and Urbany, Dickson, and Kalapurakal [15], household income was found to influence consumer search for lower prices. Lower income consumer group was found to be more likely to initiate and conduct searches for better deals or lower prices than higher income consumer group.

Urbany's [14] study using grocery product category shows that household income influences consumer's price information encoding and processing. He reports that higher income consumers are less motivated to spend time encoding price information, which is perceived to result in limited savings for most grocery products. Similarly, Wakefield and Inman [16] find that lower income consumers are more likely to be more vigilant about prices than higher income consumers. They find that lower income consumers are more likely to engage in searching and encoding price information. In the same study, Wakefield and Inman [16] also find that lower income consumers are more able to recall prices accurately than higher income consumers.

Ziethaml [18] offers that higher income consumers may be more likely to give heavier weights on time, search and psychic costs than on price considerations. On the other hand, lower income consumers who have budget constraints may be more attentive to prices.

Establishing Frequency versus Depth Cues

Having established that household income levels do have influence on consumer's search, processing and recall of price information, we now turn to a review of the body of research done on how consumers perceive two different price discount strategies (i.e., frequency vs. depth) before drawing our hypotheses.

In a study by Alba et al. [1], the effect of frequent versus deep price promotions on consumer's perception of price was examined. A frequency cue is established when there are frequent but shallow price promotions. A depth cue is established, on the contrary, when there are less frequent but deep price cuts. In this study, they find that frequency cues have a dominating influence. They suggest that consumers are likely to simplify the task of processing price information and evaluation by adopting a heuristic that allows them to better manage the task. One heuristic suggested is the frequency heuristic, which allows substantial reduction in consumer's effort of price information processing. Alba et al. [1] further offer three phenomena under which the frequency heuristic dominates, two of which are relevant to the present study and are summarized here. First, frequency heuristic may occur due to the ease of encoding with which it requires. Second, consumers may pay little attention of and may not remember the magnitude of a price discount. Thus a mere awareness of price discount may be more influential than the magnitude itself.

In a follow-up study by Alba et al. [2], frequency cue was not found to dominate. A number of reasons were offered to explain this phenomenon, but are beyond the context of the present paper and are not discussed here. However, two concepts are of relevance to the current research proposal. First is the concept of brand-across-time context, which examines consumer price perception of a brand after a certain interval of promotional periods. Second is the dichotomous versus non-dichotomous price structure. A dichotomous price structure is a structure in which a brand has only two prices – the regular price across all non-promotional periods and the promotional price for all promotional periods. A non-dichotomous price structure is when the brand has fluctuating prices. Table 1 in the Appendix gives an example of these two price structures.

The findings from Alba et al.'s [1] study suggest that a non-dichotomous condition requires more cognitive resources from consumers in price information processing, thereby inducing the reliance on numerosity/frequency heuristic.

Establishing the Hypotheses

In connecting the findings by Alba et al. [1] with Stigler's [12] *The Economics of Information* concept and previous findings as established in the prior section, we can draw on the following:

For lower income consumers, in evaluating price discount information of two identical brands within the brand-across-time context under a dichotomous price structure, the depth brand will be perceived as being lower priced. This is because, in a dichotomous price structure, price information processing is less complex and is found to be less reliant on the frequency heuristic. In addition, because lower income consumers are more likely to engage in price information processing in search of lower prices, the depth effect might be more salient under a dichotomous price structure.

On the other hand, for lower income consumers, in evaluating price discount information of two identical brands within the brand-across-time context under a non-dichotomous price structure, the frequency brand will be perceived as being lower priced. Because a non-dichotomous price structure is more cognitively complex, the frequency heuristic is more likely to be used by consumers to ease their effort in price information processing, encoding and recall.

For higher income consumers, because they are less price sensitive and less likely to engage in price information search and processing, in evaluating price discount information of two identical brands within the brand-across-time context, they will be likely to rely on the frequency heuristic when asked to evaluate price discount information as this heuristic simplifies and requires little effort in processing price information. This will result in the frequency brand being perceived as lower priced under both dichotomous and non-dichotomous price structure, although the frequency and depth brand might not yield a significant (noticeable) difference in perceived price.

Therefore, our hypotheses can be depicted graphically as in Figure 1 and can simply be stated as follows:

   H1a: For lower income consumers: in evaluating two identical brands within the brand-across-time context under a non-dichotomous price structure, the frequency brand will be perceived as being lower priced.

   H1b: For lower income consumers: in evaluating two identical brands within the brand-across-time context under a dichotomous price structure, the depth brand will be perceived as being lower priced.

   H2: For higher income consumers: in evaluating two identical brands within the brand-across-time context, neither the depth nor the frequency brand will dominate as being lower priced under both dichotomous and non-dichotomous price structure. It is hypothesized that the

frequency brand might be perceived as lower priced than the depth brand, although the difference in perceived prices for the two brands might not be significant.

FIGURE 1: HOUSEHOLD INCOME EFFECT ON CONSUMER PRICE PERCEPTION
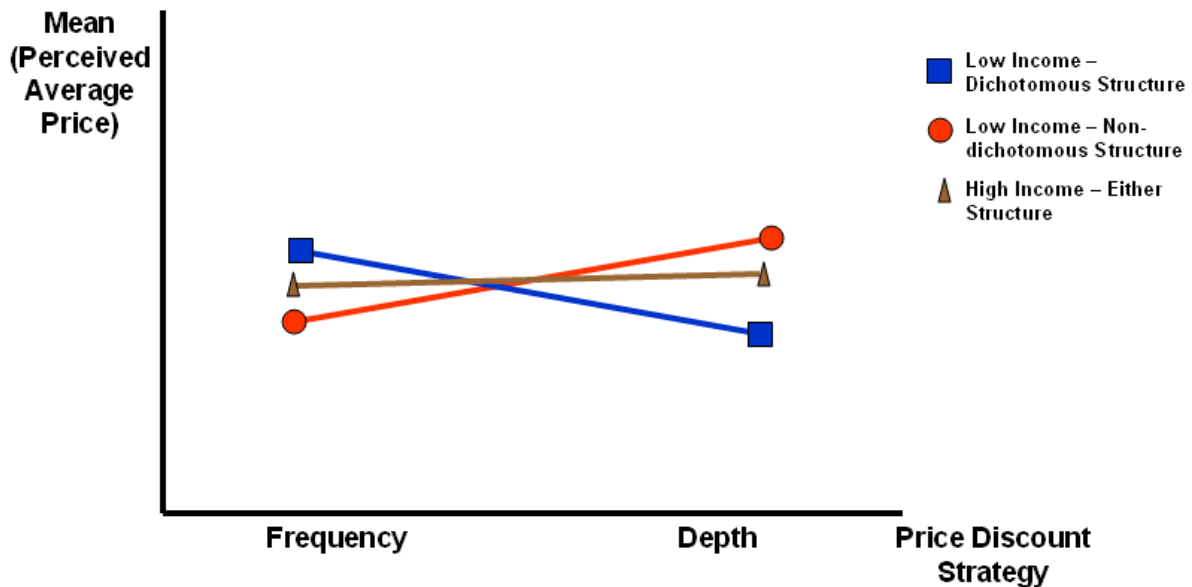
OVERVIEW OF PROPOSED STUDY

The purpose of this present research proposal is to test the proposition that household income levels affect consumer's price perception of frequent versus deep brands within the context of brand-across-time under two price structures. In the study, the frequency and depth of price promotions and price structures (non-dichotomous versus dichotomous) will be manipulated. Results from the study are expected to show that, under a non-dichotomous price structure, lower income consumers perceive the frequency brand as being lower priced while higher income consumers do not indicate any significant difference between frequency and depth brands. In addition, under a dichotomous price structure, lower income consumers are expected to perceive the depth brand as being lower priced, while higher income consumers remain indifferent.

In the study, we will adopt the use of paper-and-pencil task with two identical and comparable competing shampoo brands that run price promotions within the same store as in Alba et al.'s [2]

Study 5. The paper-and-pencil task will require subjects to review the prices of the condition they are assigned to and then answer a few questions afterwards regarding the price promotion scheme they have just reviewed – i.e., to indicate the average price of the brand they will have just reviewed. This task is chosen for our experiments for a number of reasons, similar to those that Alba et al. [2] stated. First, by looking at two identical brands within the same store, we are likely to factor out any existing bias or association subjects may have towards a store. Second, shampoo is chosen as a product for the studies because it is a product that students, who will be the subjects in this research, are likely to engage in the purchase process [4]. Further, we will adopt the price structures from Alba et al.'s [2] Study 5 for frequency and depth brands under dichotomous and non-dichotomous conditions. We will follow Alba et al.'s [2] Study 5's price structures for the key reason that they have found these price structures to be sufficient and complex enough for the purpose of this study. That is, under the non-dichotomous price structure, the prices vary enough that subjects are less likely to fall on the numerosity heuristic for their price evaluation. These price structures are presented in Table 1 in the Appendix.

In addition, the information from the U.S. Census Bureau 2008 report on household income will be used as the basis for determining high versus low income consumer groups. The report specifies that the median household income in 2008 was just above $50,000 per year. This median income along with the self-reported household income from subjects will be used to determine potentially three income groups – low, high and middle income groups. However, this study will focus on the low and high income groups only in the analyses.

STUDY

The study will test the effect of household income levels on consumer perception of frequency vs. depth price promotion within the context of brand-across-time and under two price structures (dichotomous vs. non-dichotomous).

Subjects and Design

The design will be a 2x2 between-subjects design. The between-subjects factors are the two levels of price structure (dichotomous vs. non-dichotomous), and the two price promotion strategies (frequent and shallow versus infrequent and deep). Both the price structure (dichotomous vs. non-dichotomous) and the price promotion strategies (frequent and shallow versus infrequent and deep) will be manipulated. Subjects will be randomly assigned to one of the four groups and household income level will be measured and analyzed as a covariate.

A minimum of 120 subjects will be recruited from the pool of undergraduate students at a large Northeast state university, who will participate either as part of a class requirement or for an

extra credit. Subjects will be recruited using a sign-up sheet where basic demographic information (i.e., name, age, gender, household size, and self-reported household income) will be collected. This demographic information will be relied upon to balance the sample of subjects between low household income and high household income groups. Table 2 in the Appendix provides the summary format of sample demographic compositions.

Subjects will be randomly assigned to one of the price structure scenarios, that is half of the subjects will see the dichotomous price promotion brand, and the other half will see the non-dichotomous price promotion brand. Crossed with this, half of the subjects will receive the frequency condition and the other half depth condition.

Materials and Procedure

The study will use a paper-and-pencil task exercise in which the subjects will be asked to review the price promotion scheme of the brand under the price structure they will be assigned to (i.e., dichotomous vs. non-dichotomous) and then they will be asked to answer a number of questions.

The price list for the study is given in Table 1 of the Appendix. Under the non-dichotomous price structure, for both frequency and depth brands there are 36 promotional periods, the regular selling price is $2.49 and the mean price is $2.39. The promotions for both brands are run in a non-overlapping fashion, that is – (i) the depth and frequency brand never promotes in the same period, (ii) at least one brand promotes in each of the 36 periods. The frequency brand carries a price advantage range of $0.12 to $0.22. The depth brand has a price advantage range of $0.24 to $0.36. Hence the depth brand has an advantage of 2X on average of the frequency brand. The frequency brand runs promotion with a ratio of 2:1 to the depth brand.

Under the dichotomous price structure, for both frequency and depth brands there are also 36 promotional periods and the regular selling price is $2.49 and the mean price for the 36 periods is $2.39. As in the non-dichotomous price structure, the promotions for both brands are run in a non-overlapping fashion. During the 36 periods, the frequency brand offers 24 price discounts of $0.15. The depth brand offers 12 price discounts of $0.30. Hence, the frequency brand has a 2:1 ratio in terms of promotion frequencies with the depth brand, while the depth brand has a 2X magnitude of discounts than the frequency brand.

Subjects in all groups will receive a booklet that contains *Consumer Reports* of a fictitious, shampoo brands. The booklet will inform the subjects that the brand is running price promotion at Store X for a period of 36 weeks. Subjects will then turn to see the price list for the respective price promotion and structure they will have been assigned to; that is the non-dichotomous group will get the non-dichotomous price list, and the dichotomous group will get the dichotomous

price list, and the frequency group will get the frequent brand while the depth group will get the depth brand.

Measures

Following inspection of the price list, subjects will be asked to provide answers to questions about average brand price (the dependent variable), and promotion frequencies. Subjects will also be asked to again indicate their age, household income, number of people in the household, and gender.

Planned Analyses and Expected Results

Manipulation Checks. Analyses of the manipulation checks are expected to show that both manipulations are successful. The frequency brand will be rated as having a higher number of promotion frequencies than the depth brand (M=XX; F(k-1, N-k-1)=YY; p<.01).

The dependent variable will be first analyzed using a 2 (price structures – non-dichotomous vs. dichotomous) x 2 (price promotion schemes – frequency vs. depth) ANCOVA with the two levels of household income as our main covariate along with other covariates such as age, number of people in the household and gender. The second group of covariates is expected to not show any significant effects on any of the analyses and will thus be dropped from further analyses.

All tests of significance will use p<.01. The expected means (i.e., the perceived average prices) for each promotional strategy by household income groups are presented in Figure 2 (non-dichotomous price structure) and Figure 3 (dichotomous price structure) below. Under the non-dichotomous price structure (Figure 2), for the low income group the frequency brand is expected to report a lower average price than the depth brand, and the difference between the two mean prices is expected to be significant at p<.01. This will support H1a that lower income consumers are likely to rely on frequency heuristic in evaluating price information and thus perceive the frequency brand as being lower priced than the depth brand under a non-dichotomous condition. However, for the higher income group the expected reported mean price between the frequency and depth brand is slightly different, however, the difference will not be found to be statistically significant at p<.01. This will support H2 that neither the frequency nor the depth cues will dominate among higher income consumers. The slightly lower mean price reported for the frequency brand among the high income group might be contributed by the fact they rely on the frequency heuristic when asked to evaluate price information between the two brands in the study.

Under the dichotomous price structure (Figure 3), for the lower income group the depth brand is expected to report a lower average price than the frequency brand, and the difference in the means of frequency vs. depth brand is expected to be significant at p<.01. This will support H1b that lower income consumers perceive the depth brand as being lower priced than the frequency brand under a dichotomous condition. A dichotomous condition presents less complexity allowing lower income consumers to be attentive to prices. Second, for the higher income group, the reported mean price for the frequency brand is expected to be slightly lower than the depth brand, however, the difference will not be significant at p<.01. This will support H2 that neither the frequency nor the depth cues dominate among higher income consumers.

FIGURE 2: MEAN PRICE BY INCOME GROUPS UNDER NON-DICHOTOMOUS PRICE CONDITION
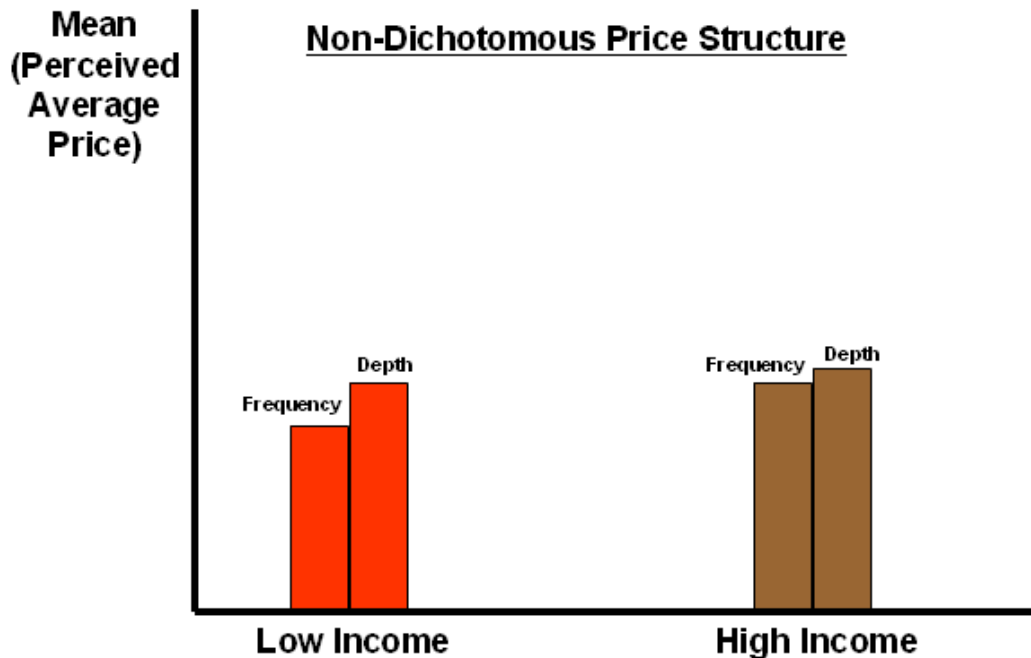
FIGURE 3: MEAN PRICE BY INCOME GROUPS UNDER DICHOTOMOUS PRICE CONDITION
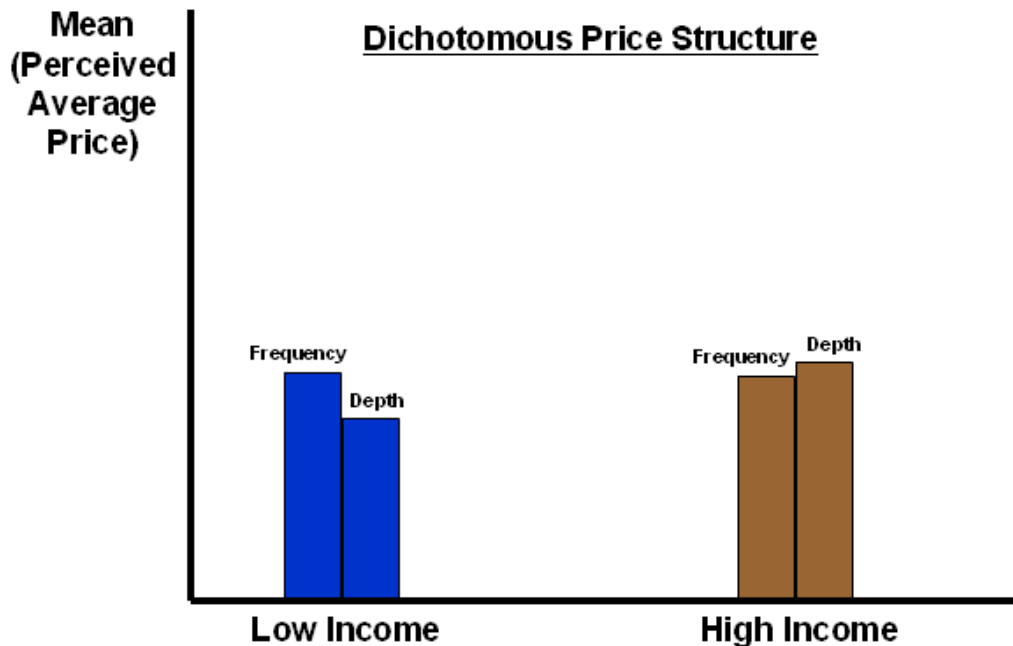
DISCUSSION

The anticipated findings in this paper extend the price information processing theory in a number of ways by offering a moderating influence of household income on price promotion perception. First, the knowledge of how different income groups, such like the dichotomous income variables used in this paper, perceive price promotion differently adds explanatory value to the price information processing theory. Second, we contribute methodologically by moving beyond grocery items in our examination of the topic matter. As Wakefield and Inman [17] suggest, numerous (almost 80 percent) experiments use some type of grocery items as the focal product in the experiments. As such, Wakefield and Inman [17] point out the need that academic research needs to move beyond the grocery lists.

Of importance to marketers, our expected findings will offer insights into consumer behavior characteristics with respect to household income levels. It is crucial for marketers to understand how the target consumer segments react to different price promotions. With the knowledge from our expected findings, marketers can more effectively tailor their price promotion strategies according to the income segments their targets belong to. In addition, marketers (especially grocery mangers) can more effectively use communication tools to emphasize price promotion to

different consumer groups in accordance to their household income level. For example, if the price promotion competition follows a non-dichotomous structure, a communication piece targeting a low income group might be more effective by emphasizing the products with deep discounts, but highlighting on the frequently discounted products for a dichotomous price structure.

Although the expected findings in this research proposal have valuable contributions and implications for managers and academicians, it should be cautioned that the product we intend to use in our study (shampoo) is perceived more as a functional product. Functional and experiential/hedonic products have different characteristics and oftentimes are found to have different relationship with consumer's purchase decision process. Hence the application of these expected research findings should not be extended to products that are more of an experiential/hedonic nature. A study by Wakefield and Inman [17] on the differences of price sensitivity with regards to hedonic vs. functional product suggests that there are situations under which low income group might "splurge" and as a result might behave inconsistently with the anticipated findings of the present research proposal. Hence, a direction for future research is to include product nature (functional vs. experiential) as an additional factor in order to further examine the interacting relationship of product nature and household income with consumer's price perception of frequency and depth price discount strategy.

Future research is also needed to further examine the effect of household income on price perception in relations to education and occupation as these factors are likely to be correlated with household income levels [17].

APPENDIX

Table 1: Non-dichotomous and Dichotomous Price Lists

Period (week)	Non-Dichotomous Frequency	Non-Dichotomous Depth	Dichotomous Frequency	Dichotomous Depth
1	2.49	2.17	2.49	2.19
2	2.33	2.49	2.34	2.49
3	2.29	2.49	2.34	2.49
4	2.49	2.19	2.49	2.19
5	2.35	2.49	2.34	2.49
6	2.33	2.49	2.34	2.49
7	2.35	2.49	2.34	2.49
8	2.49	2.25	2.49	2.19
9	2.37	2.49	2.34	2.49
10	2.49	2.19	2.49	2.19
11	2.34	2.49	2.34	2.49
12	2.37	2.49	2.34	2.49
13	2.49	2.21	2.49	2.19
14	2.35	2.49	2.34	2.49
15	2.33	2.49	2.34	2.49
16	2.27	2.49	2.34	2.49
17	2.37	2.49	2.34	2.49
18	2.49	2.19	2.49	2.19
19	2.49	2.13	2.49	2.19
20	2.33	2.49	2.34	2.49
21	2.37	2.49	2.34	2.49
22	2.35	2.49	2.34	2.49
23	2.49	2.25	2.49	2.19
24	2.31	2.49	2.34	2.49
25	2.35	2.49	2.34	2.49
26	2.29	2.49	2.34	2.49
27	2.49	2.13	2.49	2.19
28	2.34	2.49	2.34	2.49
29	2.49	2.19	2.49	2.19
30	2.31	2.49	2.34	2.49
31	2.34	2.49	2.34	2.49
32	2.35	2.49	2.34	2.49
33	2.49	2.17	2.49	2.19
34	2.33	2.49	2.34	2.49
35	2.37	2.49	2.34	2.49
36	2.49	2.14	2.49	2.19
Average Price	2.39	2.39	2.39	2.39
# Promotions	24	12	24	12

Table 2: Summary of Subjects' Demographic Composition

	Non-Dichotomous	Dichotomous
Gender (% female)	xx	yy
Age (years)		
Median	xx	yy
Range	xx	yy
Household Income (%)		
<$20K	xx	yy
$20K-$35K	xx	yy
$35K-$45K	xx	yy
$45K-$50K	xx	yy
$50K-$65K	xx	yy
>$65K	xx	yy
Sample Size	N1	N2

REFERENCES

[1] Alba, Joseph W., Susan M. Broniarczyk, Terence A. Shimp, and Joel E. Urbany (1994), "The Influence of Prior Beliefs, Frequency Cues, and Magnitude Cues on Consumers' Perceptions of Comparative Price Data," *Journal of Consumer Research*, 21 (September), 219-235.

[2] Alba, Joseph W., Carl F. Mela, Terence A. Shimp, and Joel E. Urbany (1999), "The Effect of Discount Frequency and Depth on Consumer Price Judgments," *Journal of Consumer Research*, 26 (September), 99-113.

[3] Beatty, Sharon E. and Scott M. Smith (1987), "External Search Effort: An Investigation Across Several Product Categories," *Journal of Consumer Research*, 14 (June), 83-95.

[4] Campbell, Margaret C. (2007), ""Says Who?!" How the Source of Price Information and Affect Influence Perceived Price (Un)fairness," *Journal of Marketing* Research, May 2007, 261-271.

[5] Cordeiro, Anjali, and Ziobro, Paul, "Grocers Add Promotions to Avoid Cutting Prices," *The Wall Street Journal*, Eastern Edition, April 01, 2009.

[6] Darke, Peter R. and Jonathan Freedman (1993), "Deciding Whether to Seek a Bargain: Effect of Both Amount and Percentage Off," *Journal of Applied Psychology*, 78 (6), 960-965.

[7] Kahneman, Daniel and Amos Tversky (1983), "Choices, Values, and Frames," *American Psychologist*, 39 (4), 341-350.

[8] Lalwani, Ashok K. and Kent B. Monroe (2005), "A Reexamination of Frequency-Depth Effects in Consumer Price Judgments," *Journal of Consumer Research*, 32 (December), 480-485.

[9] Monroe, Kent B. (1973), "Buyers' Subjective Perceptions of Price," *Journal of Marketing Research*, 10 (February), 70-80.

[10] Neff, Jack, "Deal or No Deal? Cheap Prices Can Maim Your Brand," *Advertising Age*, April 06, 2009, Vol. 80, Issue 12.

[11] Neuman, William, "Discounts Force Chains to Eat Their Own Lunch," *The New York Times*, June 24, 2009, pg. B1.

[12] Stigler, George J. (1961), "The Economics of Information," *Journal of Political Economy*, 49 (3), 213-225.

[13] Thaler, Richard (1985), "Mental Accounting and Consumer Choice," *Marketing Science*, 4 (3), 199-214.

[14] Urbany, Joel E. (1986), "An Experimental Examination of the Economics of Information," *Journal of Consumer Research*, 13 (September), 257-271.

[15] Urbany, Joel E., Peter R. Dickson, and Rosemary Kalapurakal (1996), "Price Search in the Retail Grocery Market," *Jounrnal of Marketing*, 60 (April), 91-104.

[16] Wakefield, Kirk L. and J. Jeffrey Inman (1993), "Who Are the Price Vigilantes? An Investigation of Differenitating Characteristics Influencing Price Information Processing," *Journal of Retailing*, 69 (2), 216-233.

[17] Wakefield, Kirk L. and J. Jeffrey Inman (2003), "Situational Price Sensitivity: The Role of Consumption Occasion, Social Context and Income," *Journal of Retailing*, 79 (September), 199-212.

[18] Zeithaml, Valarie A. (1988), "Consumer Perceptions of Price, Quality, and Value: A Means-End Model and Synthesis of Evidence," *Journal of Marketing*, 52 (July), 2-22.

ADOPTION PROCESS FOR VoIP: THE UTAUT MODEL

Eduardo Esteva-Armida, Instituto Tecnológico y de Estudios Superiores de Monterrey
Gral. Ramón Corona 2514, Zapopan, Jalisco, México 45120
Phone: (5233) 3669.3080
e-mail: eesteva@itesm.mx

Alberto Rubio-Sanchez, University of the Incarnate Word
4301 Broadway, San Antonio, Texas 78209
Phone: (210) 930.8766
e-mail: rubiosan@uiwtx.edu

In recent years, consumer research has used different models to find the variables that help consumers make decisions as to whether to adopt a new technology or not. Some of the more popular models used for this purpose are: Technology Acceptance Model (TAM), Theory of Reasoned Action (TRA), Theory of Planned Behavior (TPB), and Innovations Diffusion Theory, among others. Recently, Venkatesh, Morris, Davis, and Davis [16] tried to unify the different constructs of all these models in one unique model, and with this in mind developed the Unified Theory of Acceptance and Use of Technology (UTAUT). It is the objective of this study to test the usefulness of this model in the context of end user consumption, a task not specifically attempted by Venkatesh, et al. The setting for this test is Voice over Internet Protocol (VoIP) technology.

The Adoption Process

As with any new technology, consumers' adoption process of VoIP has been slow [8]. VoIP phones are not free of some disadvantages so suppliers need information about what factors are more important to customers. Suppliers can then develop strategies to increase the speed of the adoption process in order to maximize their profits. It was the intention of this study that as a by-product of testing the UTAUT model, information would be developed that is useful to the companies and end-users of VoIP.

The UTAUT Model

With the intention to formulate a comprehensive model that considered the variables included in previous theory aimed at explaining adoption behavior, Venkatesh, et al. [16], developed a way to test each of the constructs from eight pre-existing models: Theory of reasoned action [7], technology acceptance model [5], motivational model [4] [6] [17], theory of planned behavior [1], combined TAM and TPB [11] [13], model of PC utilization [14], innovation diffusion theory [12], and cognitive theory [3]. They presented a summary of prior model comparison studies and an empirical synthesis of the different models. Finally, with the variables that showed the biggest impact, they described a new model called Unified Theory or Acceptance and Use of Technology (UTAUT). According to the authors, this model accounted for up to 70 percent of the variance (adjusted R^2) in usage intention and it is a definitive model that summarizes what is

known and forms a basis for direct future research in this area. Considering a theoretical point of view, UTAUT gives a perspective on how the variables related to intention and behavior change over time. But the main contribution of UTAUT is by unifying the theoretical perspectives common in the adoption literature and incorporating moderators to consider dynamic impacts, namely organizational context, user experience, and demographic characteristics such as age and gender. Because most of the key relationships in the model are moderated, the study of these variables is an important added value of UTAUT [16].

UTAUT claims that three main factors (Performance Expectancy, effort expectancy, and Social Influence) determine the intention toward using a new technology while facilitating conditions and the Behavioral Intention toward using relate to the use behavior. At the same time, some variables moderate these relationships, namely gender, age, experience and Voluntariness of Use.

Based on the review of literature conducted, 4 hypotheses were developed:

> H1: Performance Expectancy will positively affect Behavioral Intention to adopt.

> H2: Effort Expectancy will positively affect Behavioral Intention to adopt.

> H3: Social Influence will positively affect Behavioral Intention to adopt.

> H4a: The impact of Performance Expectancy on Behavioral Intention will be moderated by Gender and Age.

> H4b: The impact of Effort Expectancy on Behavioral Intention will be moderated by Gender, Age, and Experience.

> H4c: The impact of Social Influence on Behavioral Intention will be moderated by Gender, Age, Experience, and Voluntariness of Use.

RESULTS

An online survey was used to test the above mentioned hypotheses. Participation was voluntary. Partial Least Squares (PLS) was used to evaluate the relationships in the model.

The sample for this study consisted of 475 respondents out of 2000 contacts. This represented a response rate of 23.8%. The instrument used to measure the variables was adapted from previous work by Venkatesh et al. and proved to be valid and reliable [16].

The results show that hypothesis 1 should be accepted. Performance Expectancy is significantly and positively related to Behavioral Intention to adopt. Hypothesis 2, however, did not find the necessary support and is rejected. Strong support was provided for Hypothesis 3. The relationship between Social Influence and Behavioral Intention is, as predicted, positive and significant.

The results show that the first order interactions for the four moderator variables with the main variables are not significant. Since the first order interactions are not significant, it is not expected than higher order interactions would be significant either. Hence, there is not support for hypotheses H4a, H4b or H4c.

DISCUSSION

The main objective of this study was to test the Unified Theory of Acceptance and Use of Technology (UTAUT) model in a reduced form in an end-use consumer context. The study worked with potential adopters of VoIP technology.

The results generated in this work using the UTAUT model are congruent with the original study made by Venkatesh et al. [16]. In both, Performance Expectancy has an influence in the Behavioral Intention to adopt a technology (called information system technology or VoIP technology). Both studies were able to account around for 50% of the variance in Behavioral Intention to adopt. In both studies, the significance of the relationship between Effort Expectancy and Behavioral Intention to adopt is not clear. The results in this study show a weak and non significant relationship.

The only difference between both studies is presented by the relationship between Social Influence and Behavioral Intention to adopt. In the original study, this relationship is not as strong and significant as that in the present research. One possible explanation can be related to the differences in characteristics of the sample. The study of Venkatesh et al. [15] was done with professionals in four organizations and the technology evaluated was something that is useful for their work. In comparison, the sample for this research was done with more heterogeneous respondents and the technology is something that they can use in daily life. The respondents' disposition to accept other person's influence in the decision to adopt these technologies may be different and stronger than in the sample used in the original study.

The results in this research provide support for almost all of the relationships specified in the model. Future research will be necessary to validate the relationship between Effort Expectancy and Behavioral Intention to adopt. Questions could be addressed to the sample, the model or to the scales used to measure one or both of these variables. This opens possibilities for future research.

Another noteworthy aspect of testing the model is the effect of the moderator variables. In the past, other researchers had found it difficult to evaluate and provide support for moderators in the model [1] [9] [10]. This study found the same problem. Even though the moderator variables Gender, Age, Experience and Voluntariness of Use show some relationship with other variables in the model, none of them was statistically significant. Academics will need to be cautious in future research regarding these relationships.

REFERENCES

[1] Ajzen, I. *The Theory of Planned Behavior*. Organizational Behavior and Human Decision Processes, 1991, 50(2), 179-211.

[2] Anderson, J. E., Schwager, P. H., & Kerns, R. L. *The Drivers for Acceptance of Tablet PCs by Faculty in a College of Business*. Journal of Information Systems Education, 2006, 17(4), 429-440.

[3] Bandura, A. *Self-Efficacy: Toward a Unifying Theory of Behavioral Change*. Psychological Review, 1977, 84(2), 191-215.

[4] Calder, B. J. & Staw, B. M. *Self-Perception of Intrinsic and Extrinsic Motivation*. Journal of Personality and Social Psychology, 1975, 31(4), 599-605.

[5] Davis, F. D. *Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology*. MIS Quarterly, 1989,13(3), 319-340.

[6] Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. *Extrinsic and Intrinsic Motivation to Use Computers in the Workplace*. Journal of Applied Social Psychology, 1992, 22(14), 1111-1132.

[7] Fishbein, M. & Ajzen, I. Belief, Attitude, Intention and Behavior: An Introduction to Theory and Research. Reading, MA: Addison-Wesley, 1975.

[8] Harbert, T. *VoIP for the Masses*. Electronic Business, 2005, 31(5), 26-28.

[9] Li, J. P. & Kishore, R. *How Robust is the UTAUT Instrument? A Multigroup Invariance Analysis in the Context of Acceptance and Use of Online Community Weblog Systems*. Proceedings of the 2006 ACM SIGMIS CPR Conference on Computer Personnel Research, 2006, 183-189.

[10] Marchewka, J. T., Liu, C., & Kostiwa, K. *An Application of the UTAUT Model for Understanding Student Perceptions Using Course Management Software*. Communications of the IIMA, 2007, 7(2), 93-104.

[11] Mathieson, K. *Predicting User Intentions: Comparing the Technology Acceptance Model with the Theory of Planned Behavior*. Information Research Systems, *1991,* 2(3), 173-191.

[12] Rogers, E.M. *Diffusion of Innovations.* New York, NY: The Free Press, 1983.

[13] Taylor, S. & Todd, P. *Assessing IT Usage: The Role of Prior Experience.* MIS Quarterly, *1995,* 19(4), 561-570.

[14] Thompson, R. L., Higgins, C. A., & Howell, J. M. *Personal Computing: Toward a Conceptual Model of Utilization.* MIS Quarterly, 1991*, 15(1), 125-143.

[15] Venkatesh, V. *Determinants of Perceived Ease if Use: Integrating Perceived Behavioral Control, Computer Anxiety and Enjoyment into the Technology Acceptance Model.* Information Systems Research, 2000, 11(4), 342-365.

[16] Venkatesh, V., Morris, M., Davis, G., Davis, F. *User Acceptance of Information Technology: Toward a Unified View.* MIS Quarterly, 2003, 27 (3), 425-478.

[17] Venkatesh, V. & Speir, Ch. *Computer Technology Training in the Workplace: A Longitudinal Investigation of the Effect of Mood.* Organizational Behavior and Human Decision Processes, 1999, 79(1), 1-28.

EWOM via MAVENS, BUZZ AGENTS, AND FOLLOWERS

Kathleen R. Ferris-Costa
Visiting Assistant Professor of Marketing
Gabelli School of Business, Roger Williams University, Bristol, RI
Dr. Ruby R. Dholakia
Director, RITIM
Professor of Marketing
College of Business Administration
University of Rhode Island, Kingston, RI.

ABSTRACT

With the growth of eWOM, both firm-sponsored and consumer initiated, the opportunities to seek and receive marketplace information from personal sources are increasing. Parker (2005) reports that eWOM (both seeking and giving) is pervasive and growing. Consumers in this study were approximately 16% more likely to be influenced by eWOM than by traditional advertising media (radio, TV, and newspapers). The plethora of platforms for communication using the Internet also suggests that the size and nature of social networks as well as the speed and variety of feedback mechanisms are significantly different from the offline world where face-to-face interactions among close ties are common.

Most of the available evidence on the influence of market mavens and buzz agents come from the offline world. The objective of this paper is to extend the current literature on market maven and buzz agent communication behaviors to the online world. With the growing use of consumers as communication agents, the research focuses on the motivation and communication processes of the buzz agent and compares them to the market maven; the aim is to describe and determine how these two types of influentials differ in their motivation and communication processes.

Drawing upon Resource Exchange Theory (Foa 1993, 1976, Foa & Foa, 1974), eWOM participation will be described as an exchange process where participants engage in eWOM practices for specific resources. Motivation theory (Deci 1971; Deci and Ryan 1991) informs the need for feedback and the importance of the source from which consumers seek rewards. Finally, communication theories pertaining to source credibility and trustworthiness (Ahuja, Michels, Walker, and Weissbuch, 2007; Carl, 2006; Milne, Rohm, and Bahl, 2009; Nyilasy, 2006) inform proposals of relationships between tie strength and the effect of disclosure on influence processes.

The research method consists of one online survey and one online experiment. Members of a social networking site and agents from a word of mouth marketing agency will be recruited to

participate in the study. This procedure increases the external validity of the research since the study sample will be composed of consumers who are active online as receivers, seekers and/or providers of marketplace information.

The paper hopes to contribute to managers' understanding of the online inter-personal networks that influentials use to share marketplace information with others. Knowing what motivates these individuals, in addition to the communication channels and social networks that they use to share marketplace information, would be useful to marketers with the purpose of utilizing their resources more efficiently.

ONLINE CONSUMER'S PERCEIVED VALUE: REVISITING THE PERCEIVED VALUE (PERVAL) SCALE WITH ONLINE CONSUMERS

Sereikhuoch Eng, University of Rhode Island, USA

ABSTRACT

Online consumers are found to be more utilitarian oriented than hedonistic compared to their offline counterparts. Convenience and timelessness are the principal advantages of online shopping. Hence, consumers who shop online may place value differently than those who shop offline. Despite these indications of differences between online and offline consumer behavior, there has not been research conducted to understand whether the Perceived Value (PERVAL) scale adequately measures online consumer's perceived value. The current research proposes to address this gap; it specifically aims to answer two questions: (i) Does the PERVAL scale measure similarly for online consumers? (ii) Do the means differ for female and male consumers on any of the PERVAL factors?

KEYWORDS

Online consumers, Perceived value, PERVAL scale

INTRODUCTION

The U.S. online shopping spending was reported to be $32.1 billion for the third quarter of 2010, a solid nine percent increase compared to the same period last year [18]. With the growing focus and activities of online shopping, managers and academicians in the field of marketing need to have a better understanding of online consumer behavior in order to effectively create appealing and relevant values. Research in online consumer behavior clearly indicates that consumers behave differently in an online environment, and variables understood in an offline environment do not necessarily or sufficiently apply to an online environment (e.g., [1, 7, 11, 12, 14, 24]).

Consumer's perceived value literature indicates that perceived value is an imperative variable that leads consumers to the purchase path. With increased time pressure among consumers, the convenience and timelessness of online shopping have been cited as the principal advantages for online consumers [24]. In addition, with the convenient online tools (e.g., Google or BizRate to compare offerings) to ease choice making within an online shopping environment, consumer's expectations and experience in online differ from offline [17]. Bridges and Florsheim [6] find that online consumers view their online shopping as utilitarian than hedonic. However, the existing four-factor, 19-item PERVAL scale of consumer's perceived value [19] was developed within the physical retail environment, and does not consider these unique characteristics of an online shopping environment. Given the variability of online consumers, the current research proposes to address the need to understand whether the PERVAL scale equally explains online consumer's perceived value. Specifically, the current research proposes to address two questions:

1. Does the PERVAL scale measure similarly for online consumers?

2. Do the means for female and male consumers differ on any of the PERVAL factors?

CONCEPTUAL BACKGROUND

Perceived value can be regarded as a "consumer's overall assessment of the utility of a product (or service) based on perceptions of what is received and what is given" [23, p. 14]. Informed by extant research on perceived value (e.g., [2, 19]), Sweeney and Soutar [20] developed a four-factor, 19-item scale for measuring perceived value (PERVAL). PERVAL was developed as a measure of consumer's perceived value in the retail literature and industry. This scale was originally developed and tested within the traditional retailing concept – that is, it involves the idea of tangible and physical assessment of products before making a purchase decision. The PERVAL scale is presented in Table 1.

TABLE 1: PERVAL SCALE (ADAPTED FROM SWEENEY & SOUTAR, 2001)

PERVAL Factors	Items
Quality/Performance (functional value): The utility derived from the perceived quality and expected performance of the product	has consistent quality is well made has an acceptable standard of quality has poor workmanship (*) would *not* last a long time (*) would perform consistently
Price/value for money (functional value): The utility derived from the product due to the reduction of its perceived short term and longer term costs	is reasonably priced offers value for money is a good product for the price would be economical
Social value (enhancement of social self-concept): The utility derived from the product's ability to enhance social self-concept	would help me to feel acceptable would improve the way I am perceived would make a good impression on other people would give its owner social approval
Emotional value: The utility derived from the feelings or affective states that a product generates	is one that I would enjoy would make me want to use it is one that I would feel relaxed about using would make me feel good would give me pleasure

(*) reverse scored.

METHOD

The 19-item PERVAL scale uses seven-point Likert scale from "1-Strongly Disagree" to "7-Strongly Agree." This study will adopt an Internet questionnaire using an online survey tool (surveymonkey.com) in collecting data. The questionnaire link will be posted on relevant College of Business Sakai course sites to invite participation. The survey link will be posted on social networking site (e.g., facebook.com), and email invitations will be sent out to solicit participation from non-student adult populations. All participants will be required to have online shopping experience within the last 3-6 months. The questionnaire will ask participants to think of their recent online shopping experience in completing the survey. The questionnaire will screen participants for their recency of online shopping experience. Participants who pass this screening stage will continue to answer questions on the PERVAL scale and demographic information (e.g., sex, age, race, education level, income, state/city of residence). Those who do not satisfy the qualifying criterion of having engaged in past online shopping in the last 3-6 months will be screened out to the demographic page. The questionnaire will have 26 items and

take approximately 20 minutes to complete. The study expects to collect N>260, with a reasonably balanced split between male and female samples.

PLANNED ANALYSES

The total sample used in the analyses will be N>260 (ratio of cases to free parameters 10:1; [15]), about half of which are expected to be females. Participants who do not have online shopping experience in the past 3-6 months will be excluded from the analyses (actual sample size excluded from analysis will be reported).

Confirmatory factor analysis will be used to examine the four-factor, 19-item PERVAL scale on the *overall* online consumer sample. Consistent with Sweeney and Soutar [20], the four factors will be allowed to correlate (e.g., an attractive rug would have a favorable response on social value). In order to test for mean differences between female and male samples on the PERVAL scale, multiple-sample CFA with mean structure analysis will be executed. Both analyses will be run using structural equation modeling program EQS 6.1 [5].

EXPECTED RESULTS

Descriptive statistics of the data will be examined first to detect any violation of general linear model assumptions of linearity, normality and homoscedasticity [13]. The data are expected to meet these assumptions.

Reliability and validity. Internal consistency (or reliability) of the current study is expected to be indicated by relatively high Cronbach's alpha (all >.70; [10]). As the scale was adopted from the original PERVAL scale, the questionnaire in this study has content validity.

CFA results. CFA analysis is performed to categorize online shoppers' perceived value of their purchase decisions. Consistent with Sweeney and Soutar [20], a forced four-factor analysis is imposed on the data. CFA analysis is expected to indicate that three factors of the PERVAL scale converge from the data. Items in each of these three factors are expected to have relatively high loadings of >.60 and relatively small measurement errors with standardized residuals <|.20| [16]. However, the Emotional value scale is not expected to be robust (all loadings are expected to be <.60 with a low coefficient alpha <.50; [8]). The absence of strong association of Emotional value among online consumers may be attributed to online purchases being facilitated by utilitarian motivations [6].

Multiple-sample CFA results. A multiple-sample CFA with mean structure analysis will be run with the three retained factors from the previous planned analysis. Multiple-sample CFA analysis will allow detection of any mean differences between the two gender-groups of the sample on

the retained three factors of the PERVAL scale. First, the evaluation of measurement invariance is expected to indicate that all three retained factors (and indicators) measure similarly for both female and male samples. Second, two factors, Price and Quality, are expected to have mean differences that differentiate the two gender groups. Specifically, male online consumers are expected to be motivated by Quality value significantly more than their female counterparts. On the other hand, female online consumers are expected to be dominated by Price value significantly more than male consumers. One possible explanation to such expected finding is that females are likely to be more involved in household tasks (e.g., childcare, house chores) that they have less time for shopping [17] and may be more likely to use comparison sites (e.g., Google and BizRate to compare offerings) in making purchase decisions.

DISCUSSION

The Sweeney and Soutar's PERVAL scale for measuring consumer's perceived value in an online shopping environment is expected to reveal three robust factors that online shoppers perceive significant in their online purchases: Quality value, Price value, and Social value. The Emotional value from the original scale is not expected to load significantly among online shoppers. This may be due to the fact that online shoppers normally view online shopping as a utilitarian experience than hedonistic [6]. Also, consumers who engage in online shopping have been found to have thought about their purchases before going online, and not think of online shopping as a recreational activity [24]. This expected discrepancy suggests that there is a need to exercise caution before applying a theory or construct developed for the physical retail environment to the online environment. It also indicates that a need exists for the development of perceived value scale specifically for online shopping context.

Expected findings in the current study also indicate that potential differences in perceived value exist between male and female groups. In particular, male online consumers are expected to be motivated by Quality value significantly more than their female counterparts. On the other hand, female online consumers are expected to be dominated by Price value significantly more than the male group. The availability of comparison sites (e.g., Google and BizRate) has been found to facilitate the increase in online shoppers' focus on price value [17]. This expected finding benefits managers of online shopping sites in their decisions on marketing strategy variables from positioning, pricing, product strategy, to attribute-focused communication to these specific groups of online consumers.

Finally, while the current research will contribute to the consumer perceived value literature and the understanding of online consumer behavior, it adopts a generic online shopping scenario thereby is insufficient at explaining specific behaviors. Extant research in consumer decision-making suggests that consumer's level of involvement, information processing, and cost-benefit analysis differ across product categories and situational conditions (e.g., [3, 4, 9, 21, 22]). Future research may add additional value by testing the perceived value scale within the online context

using different product categories and situational factors (e.g., different value placed on going to the movie with friends than with a date).

REFERENCES

[1] Al-maghrabi, T., & Dennis, C. (2010). Driving online shopping: Spending and behavioral differences among women in Saudi Arabia. *International Journal of Business Science and Applied Management*, 5(1), 30-47.

[2] Babin, B.J., Darden, W.R., & Griffin, M. (1994). Work and/or fun: Measuring hedonic and utilitarian shopping value. *Journal of Consumer Research*, 20(March), 644-656.

[3] Beatty, S.E., & Smith, S.M. (1987). External search effort: An investigation across several product categories. *Journal of Consumer Research*, 14(June), 83-95.

[4] Belk, R.W. (1975). Situational variables and consumer behavior. *Journal of Consumer Research*, 2, 157-164.

[5] Bentler, P.M. (2004). *EQS 6.1 Structural equations program manual*. Encino, CA: Multivariate Software, Inc.

[6] Bridges, E., & Florsheim, R. (2008). Hedonic and utilitarian shopping goals: The online experience. *Journal of Business Research*, 6, 309-314.

[7] Chan, K.W., & Li, S.Y. (2010). Understanding consumer-to-consumer interactions in virtual communities: The salience of reciprocity. *Journal of Business Research*, 63, 1033-1040.

[8] Churchill, G.A. (1979). A paradigm for developing better measures of marketing constructs. *Journal of Marketing*, 16(February), 64-73.

[9] Clarke, K., & Belk, R.W. (1978). "The effects of product involvement and task definition on anticipative consumer effort," in *Advances in Consumer Research*, Vol. 5, ed. H. Keith Hunt, Ann Arbor, MI: Association for Consumer Research, 313-318.

[10] Cronbach, L.J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297-334.

[11] Garbarino E., & Maxwell, S. (2010). Consumer response to norm breaking pricing events in e-commerce. *Journal of Business Research*, 63(9), 1066-1072.

[12] Gupta, P., & Harris, J. (2010). How e-WOM recommendations influence product consideration and quality of choice: A motivation to process information perspective. *Journal of Business Research*, 63, 1041-1049.

[13] Harlow, L.L. (2005). *The essence of multivariate thinking: Basic themes and methods*. Mahwah, NJ: Erlbaum.

[14] Ho, S.S., & McLeod, D.M. (2008). Social-psychological influences on opinion expression in face-to-face and computer-mediated communication. *Communication Research*, 35(2), 190-207.

[15] Jackson, D.L. (2003). Revisiting sample size and number of parameter estimates: Some support for the N:q hypothesis. *Structural Equation Modeling*, 10, 128-141.

[16] Kline, R.B. (2010). *Principles and practice of structural equation modeling*. New York, NY: The Gilford Press.

[17] Papatla, P., & Liu, F. (2009). Google or BizRate? How search engines and comparison sites affect unplanned choices of online retailers. *Journal of Business Research*, 62(11), 1039-1045.

[18] Sage, A., & Wallace, J. (2010, November 01). U.S. online spending rose 9 percent in Q3: comScore. *Reuters*. Retrieved November 06, 2010, from http://www.reuters.com/article/idUSTRE6A03LF20101101

[19] Sheth, J.N., Newman, B.I., & Gross, B.L. (1991). Why we buy what we buy: A theory of consumption values. *Journal of Business Research*, 22(March), 159-170.

[20] Sweeney, J.C., & Soutar, G.N. (2001). Consumer perceived value: The development of a multiple item scale. *Journal of Retailing*, 77, 203-220.

[21] Urbany, J.E., Dickson, P.R., & Kalapurakal, R. (1996). Price search in the retail grocery market. *Journal of Marketing*, 60(April), 91-104.

[22] Wakefield, K.L., & Inman, J.J. (2003). Situational price sensitivity: The role of consumption occasion, social context and income. *Journal of Retailing*, 79(September), 199-212.

[23] Ziethaml, V.A. (1988). Consumer perceptions of price, quality and value: A means-end model and synthesis of evidence. *Journal of Marketing*, 52(July), 2-22.

[24] Yang, C., & Wu, C. (2010). Gender and Internet consumer's decision-making. *Cyberpsychology and Behavior*, 10(1), 86-91.

CO-EVOLVING BUSINESS MODELS FOR SERVICE QUALITY, INNOVATION STRATEGY, AND MARKETING: A CASE STUDY OF ORGANIZATIONAL CHANGE IN CHUNGHWA TELECOM

Wei-shang Fan, Nan Hua University, (886)911-111243,ws.fan@msa.hinet.net

Ruei-Shiuan Chang, Nan Hua University, (886)988-368837, rschang@cht.com.tw

Yung-Chang Li, Nan Hua University, (886)937-636689,ycl0726@cht.com.tw

Yi-Hsuan Lin, Nan Hua University, (886)928-789797, shelly90425@yahoo.com.tw

Yu-Hsuan Ou, Nan Hua University, (886)918-275660, a901090@yahoo.com.tw

ABSTRACT

This paper employs a single case study methodology to examine changes in service quality, innovation strategy, and marketing during organizational transformation of the case company. Since its corporatization in 1996, the case company has passed the first phase of market testing for developing wireless communications. Following privatization, the case company has actively enhanced service quality, ensured its technological advancement using various innovation strategies, and provided sophisticated services to create a competitive advantage. By continuously strengthening its marketing at all levels and establishing a common view on "all-staff marketing," the company has transformed to face competition during the liberalization of telecommunications, playing a leading role in telecommunications in Taiwan. This study concludes by drawing on the transformation process of the case company to create a business model for developing service quality, innovation strategy, and marketing. Lastly, the study provides recommendations on planning strategies for sustaining competitive advantage after transition that can serve as references for real-world applications.

Keywords: All-staff marketing, Service quality, Innovation strategy, Case study, Chunghwa Telecom

INTRODUCTION

Mao (2004) believed that any organization has its expected functions and objectives for its existence. When material changes occur in the external environment or when degeneration happens within an organizational system that makes the organization unable to perform its functions or to achieve its expected objectives, then the organization needs a reform. From this definition, reorganizations and restructuring in China in the past hundred years are undoubtedly a form of organizational transformation. Some of these were successful; some were not. However, aside from discussions from a historical aspect, few of the factors for these successes and failures have been analyzed from a management aspect.

The early telecommunications market in Taiwan was a monopoly with the Directorate General of Telecommunications (DGT) under the Ministry of Transportation and Communications (MOTC) in charge of operating and managing various telecommunications businesses. With rapid changes in international economy and trade, state-owned enterprises with their conservative approach have successively and vigorously adopted response measures under the impact of trends from liberalization and internationalization. For this

reason, the period for diversification in operations for state enterprises has arrived, and it is imperative that transformation, consolidation, or privatization be enforced and not be hindered.

As state enterprises exist as government organizations and whose development is insufficient to meet the demand, everything is oriented towards production and weak on marketing. Its primary task is to implement policies of the central government, mainly establishing advanced and practical telecommunications equipment and applications to bring greater convenience to the people. Its service to its customers emphasizes on providing the basic communications functions to the people and meeting the move of the government of promoting the development of telecommunications while avoiding as much as possible any public discontent towards customer service. However, since development of the telecommunications towards diverse services, the traditional monopoly model has been unable to meet consumer demands. Telecommunication agencies around the world such as AT&T in the United States, BT in England, and NTT in Japan have been privatized one after another. To comply with the WTO agreement on liberalizing basic telecommunications, Taiwan formally separated in 1996 the DGT and CHT responsible for regulating and operating telecommunications respectively and has opened various telecommunications services year by year, strengthening the service quality and competitiveness of the telecommunications market in Taiwan.

In an open competitive market transformed from a monopoly, the example of Chunghwa Telecom (CHT) sustaining a stable position as Taiwan's leading telecommunications company and firmly maintaining profitability is a successful case of transformation. Thus, the objective of this study is to construct the transformation process of the case company and business models for the equally evolving service quality, innovation strategy, and marketing while at the same time, providing recommendations on devising strategies for sustaining competitive advantage after the transition to serve as a reference for real-world applications.

LITERATURE REVIEW

For enterprises, organization is an instrument for seeking survival and development in the market and industry systems. When changes occur in the subjective and objective environments internally and externally, the organization, as an instrument of survival for enterprises, must then also adjust and adapt, thereby facilitating organizational transformation (Mao, 2005). Thus, the period when enterprises only stress profits has already changed. Future competition in the market rests not on money; only enterprises that emphasize innovation and imagination can stand out. Innovation and imagination then must be the trends for development for business operations in the future. Particularly in the future international market where emerging markets play a critical role, businesses must learn to adjust operations strategies, establish closer interactive ties with consumers, and develop products that are customer-oriented. This is because operating strategies for business in the e-generation center on strengthening service quality and creating value.

Engel, Blackwell, and Miniard (2001) stated that marketing is creating an exchange that meets individual and organizational objectives through pricing, promotion, and distribution. Customers exchange anything of value with marketing personnel and are willing to pay the price, in the process satisfying individual needs and organizational objectives at the same time. Jay (2003) showed that marketing capability also can influence innovation abilities of a company and the sustainability of its competitive advantage. This research contributes to the development of marketing strategy theories and applications in practice. The above discussions indicate that the sustainable development of business competitive advantage is

closely related to service quality, marketing capability, and organizational innovation strategies. These are discussed as follows.

Service Quality

Service quality is an operating strategy of businesses for satisfying customer needs, creating customer value, and increasing company profits. Generally, service quality is interpreted from the standpoints of businesses and consumers. Several scholars believed that service quality is the perceived difference between the expectations of consumers and the performance of service. Research into the sustainability and subsequent impact of service, including the effect of service quality on customer satisfaction and loyalty (Yu, Chang & Huang, 2006; Olorunniwo, Hsu & Udo, 2006a; Olorunniwo et al., 2006b; Bell, Auh & Smalley, 2005) and the effect of service value on satisfaction (Sinclair, Fleming, Radwinsky, David, & Jill, 2002) have also gained acceptance. Thus, service quality is a critical factor in the service industry. The increase in service quality can increase customer satisfaction.

Wakefield (2001) suggested that service quality influences both service attitude and location. Customers will usually evaluate service quality based on the tangible and intangible clues observed during the interaction with the business. Thus, good service quality brings more new customers for the business, increases transaction with existing customers, and reduces turnover of customers, making it harder to be affected by price competition (Huang and Lin ,2005; Brady, Robertson & Cronin, 2001). After the opening of the telecommunications market in Taiwan, the case company demonstrated more the importance of its competitive advantage in service quality amidst competitions using low costs to attract customers.

Telecommunications service has four characteristics: (1) intangibility: indicating that service cannot be seen, tasted, touched, listened to, or smelled before being purchased; it is intangible (Lovelock, 1981). (2) Inseparability: illustrating that regardless of whether the provider is human, equipment, or machine, the service and its provider are inseparable (Regan, 1963). (3) Perishability: indicating that service cannot be stored, retained, resold or returned for future sales or usage. (4) Variability: referring to the heterogeneity of service; there will be difference for the same service when provided by a different person at a different time or location (Parasuraman, Zeithaml & Berry, 1985).

Innovation strategy

Innovation strategy indicates that in addressing an innovative subject, innovative objectives are devised by organizations and appropriate resources are allocated to innovative activities able to achieve innovative objectives. Kotler (2004) pointed out eight stages in the development process of innovative products: idea generation, idea screening, concept development and testing, marketing strategy development, business analysis, product development, test marketing, and commercialization. Crowe and Brennan (2007) also pointed out that the importance of innovation for a company lies in the myriad of policies and strategies which help the company nurture and reap its profits.

Breakthrough product innovation can be classified into technology breakthrough and market breakthrough innovation (Benner & Tushman, 2003), whereas Zheng Zhou, Yim, & Tse (2005) have modified them as technology- and market-based product innovation. Compared with the product or process perspective, scholars have observed that many studies emphasize the technology innovation of businesses while ignoring management innovation, thus beginning to view management innovation (including system, policies, programs, and service) as a manifestation of organizational innovation. Robbins (2001) also emphasized that innovation is a new concept that can be used to set in motion or promote any product, process, or service; innovation includes at the same time product innovation, new production process technology, new structure and management system, and new plans.

From the above discussion on innovation types, it can be observed that measuring innovation involves several constructs. But generally, it can be categorized into management and technology innovation. Management innovation includes innovation in management, marketing, and market, while technology innovation includes product, production, and process innovation.

Marketing

The key for marketing to achieve a company's objectives is to investigate the demands and desires of the target market and to meet consumer demands more effectively and more efficiently than its competitors. Under a profitable situation, businesses provide service that is enough to satisfy customers; customers bring in profit and money returns to the company. Social marketing must be able to equally consider both the welfare of consumers and society. It requires companies to likewise consider the balance between company profits, consumer desires, and social benefits when deciding on marketing policies.

Engel, Blackwell, and Miniard (2001) indicated that marketing is creating an exchange that meets individual and organizational objectives through pricing, promotion, and distribution. Customers exchange anything of value with marketing personnel and are willing to pay the price, in the process satisfying individual needs and organizational objectives at the same time. "Marketing is creating, communicating, and having anything of value to customers (Kotler, 2003)." In comparing with production, human resources, research and development, and finance among the business functions, marketing is the closest to customer relationship. At present, many of the strategies used by companies do not only attract new customers and create transactions. More than this, they also employ customer relationship to retain existing customers and establish long-term and profitable relationship with them. From this new perspective, marketing then is a science and art of pursuing, maintaining, and cultivating profitable customers (Kotler, 2004)

The telecommunications products provided by telecommunications companies refer to the sum of services that market activities satisfy customer communication demands and benefits and other related additional services (Xu, 2005). Specifically, telecommunications products are comprised of three levels: (1) Core products: referring to the utility and benefits customers can expect to receive when purchasing any telecommunications products, which is the real purpose of consumers purchasing products. (2) Formal products: which are the form of the product. Telecommunications products are presented by name, kind, brand, features, and performance quality. Formal telecommunications products are the core benefits of customers as reflected externally on product form and characteristics. (3) Additional products: which are the sum of additional services and benefits customers can receive or expect to receive when purchasing telecommunications products. These are added to the core products to allow customers to experience their "value." To provide good telecommunications products and services does not only mean having customers receive additional benefits and symbolic value, it also involves strengthening the popularity of and satisfaction towards telecommunication products. Furthermore, on enhancing and stressing core and formal products, additional products gain customer attention and strengthen efforts of convincing the customers to purchase.

RESEARCH DESIGN

This study employs the longitudinal case analysis methodology on a single case company to examine the dynamic process of its growth. Eisenhardt and Graebner (2007) believed that case studies can emphasize the background of the phenomenon and can fully and truly illustrate it. Hinings (1997) considered qualitative research methodology relatively

appropriate for analyzing dynamic processes, especially when these processes need a particular explanation. Even though case studies concentrate on a specific situation, employ multiple data collection methods, and provide an in-depth perspective in analyzing a situation, they are often being criticized in the areas of generalization, accuracy, objectiveness, strictness, and other biases. Pettigrew (1990) considered that longitudinal research is appropriate for examining business development and transformation and is able to be used to observe dynamic changes of a phenomenon at different times and the details of these actual operations.

Yin (1984) proposed that when research questions are "why" or "how," case studies are always a better research strategy. Because the objective of this study is to examine "why" service quality, innovation strategy, and marketing need to change during the transformation of the case company and "how" these three interact with one another, an in-depth case study can provide a substantial explanation for the subject for this study. This is the primary research method employed for this study

Case selection

The objective of this study is to understand the changing business models for service quality, innovation strategy, and marketing during the process of transformation of a telecommunications company. From being a government organization to a state corporation to a private enterprise, CHT has significant research value due to its different roles, differences in background, and changes in its responsibilities. It is the only integrated telecommunications company in the Taiwanese market, and was selected as the case company for this study.

Data collection

This study adopts the case study methodology. Using a single case study, the co-evolving relationships of service quality, innovation strategy, and marketing of a company during transformation is examined through data and interview collection. The observations and in-depth interviews for a single case can help understand the evolving process of a company in the transition of its entire marketing.

In-depth interviews were conducted with high-level management. The total number of people interviewed was 6, while the total number of times they were interviewed was 12. Information on the interviewees including the number of times of their interviews and their position at the time of the interviews are shown in Table 3-1.

Table 3-1 Information on the interviews

Position/title of interviewee	Times of interview
Labor director	3
Branch associate	1
Assistant manager of business unit	2
Assistant manager of business unit	2
Assistant manager of business unit	2
Branch deputy director	2
Total 6 people	12 times

Interview procedures

Before the interviews, questions were first drafted and sent by electronic mail to the respondents. Confirmation and arrangement of time for interview followed. Interview processes were conducted in a non-structured manner according to the subject of the research. With the consent of the respondents, the entire content of the interviews was recorded to have

a complete documentation of the interviews. After each interview, discussions with the interviewees were facilitated to share observations that would be used as main points for the next interview and as topics for the research.

THREE STAGES OF CO-EVOLVING MODELS FOR SERVICE QUALITY, INNOVATION STRATEGY, AND MARKETING

I. Development and introduction of case company

Chunghwa Telecom Co., Ltd. is a state-owned enterprise that was separated on September 1, 1996 from the DGT under the MOTC in accordance to the three telecom laws of Taiwan. It is specifically responsible for the operations of telecommunications under the MOTC and functions as a Type 1 and Type 2 telecommunications carrier. On August 2005, its privatization was completed (Chunghwa Telecom global website, 2009).

The principal line of services offered by CHT includes fixed-line communications, mobile communications, and data communications. It provides voice services, leased lines, Internet, broadband Internet access, intelligent networks, virtual networks, e-commerce, enterprise integrated services, and other types of value-added services. It is the most experienced and largest integrated telecommunications company in Taiwan. As of the end of December 2008, its market share of local telephone customers in fixed line services was 97.3%, market share of domestic long distance telephone services was 85.2%, and market share of international telephone services was 59.5%. Its market share of cellular customers was 35.2% and in terms of Internet and data communications, its market share of subscribers in HiNet was 68.1%. The number of its broadband customers accounted for 83.8% of the overall broadband market (Chunghwa Telecom 2008 Annual Report).

The consolidated revenues of CHT in 2008 was NT$ 201.67 billion. Its profit was NT$ 45.01 billion and net profit rate was 22.3%. Fixed-line communications accounted for 31% of the total revenues, mobile communications 38.7%, data communications 26.1%, and other areas of business 4.2%. The total amount of its capital was NT$ 96.968 billion and the total number of its employees was 24,551 (Chunghwa Telecom 2008 Annual Report).

II. First stage: Monopoly stage (Before 1996)

Because of the time period and government needs in the past, public utilities had to achieve many policy-related tasks and objectives and in turn, they had certainly achieved a number of policy objectives. At the monopoly stage, the case company was a government agency dealing with the responsibility of constructing advanced and practical telecommunications equipment and applications to offer greater convenience to the people. In the 1950's, general transactions and emergency connections were mainly done by telegram. At that time, the primary responsibility of the telecommunications board was to send this kind of information accurately and quickly to the people.

Along with the rise of the economy, demand for local calls had also greatly increased. A huge number of local call networks were expanded to meet the demands of the people and achieve the government policy of having telephones for all villages. At that time, the case company primarily handled the policy objectives of Taiwan, drove its economic growth, met the communications demands of the people, and strengthened the competitiveness of Taiwan in telecommunications equipment. It mainly focused on the local market and first-time users. Mode of thinking was done from the standpoint of administrative government. Its service to its customers emphasized providing the basic functions in communications for the people and fulfilling the move of the government of promoting the development of telecommunications while avoiding as much as possible any public discontent towards customer service.

Another key point at this stage was the nurturing of local communications personnel, for example, by establishing collaborations with schools. Telecommunications personnel at the case company, MOTC, DGT, and National Communications Commission were developed at this stage.
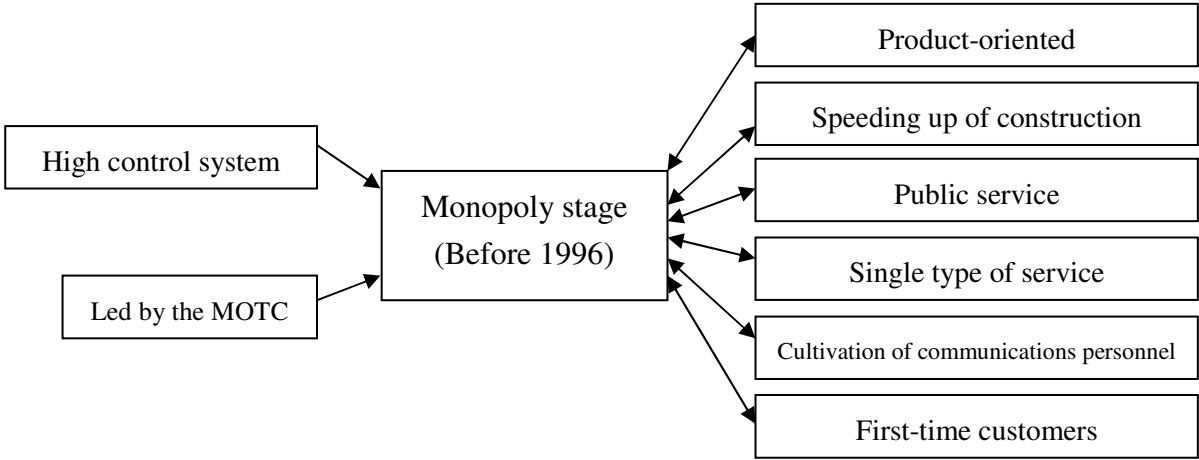
Figure 4-1 First stage of the case company: Monopoly stage

III. Second stage: Stage of major change (1996-2005)

In 1996, the regulating and operating functions of telecommunications, DGT and CHT respectively, were formally separated, while the telecommunications market in Taiwan was opened to various telecommunications businesses in the following years. At this point, the telecommunications market in Taiwan formally entered the period of total liberation. At the beginning of its corporatization and liberalization, its main task was to prioritize the completion network infrastructure, such as fiber optic cabling and wiring, and to construct 3G mobile infrastructure and infrastructure for IP information network in data communications. On the client side, on the other hand, the case company had aggressively driven the market share of ADSL and mobile customers. However, with the rigidity of the budget and procurement system, it was unable to meet the demands of the market. When the mobile services was opened in 1997, new operators easily seized a major part of the market and CHT placed third for a time in terms of market share.

After the opening of fixed-line network services in 2001, competition in the telecommunications market became fiercer. At this stage, aside from expanding devices, establishing base stations, improving the quality of communications and meeting the demands of customers in mobile services, undertaking the challenge of new operators also involved meeting the demands of customers, optimizing, especially in making the fiber optic transmission backbone, and expanding the transmission rate within the fixed-line network services. Using advancements in telecommunications technology, developing value-added services, making good use of human resources and driving "all-staff marketing," establishing various service centers to respond to open competition in the market, and most importantly, vigorously promoting the privatization of CHT were done to enhance its service quality and competitiveness in the telecommunications services market.

Kotler (1998) believed marketing is pursuing customer satisfaction through target marketing, customer orientation, and marketing coordination to achieve organizational objectives. Competition in the market, opening of mobile services, and decline in market share have changed marketing. For example, the transformation of organizational structure has changed business units to marketing units and the pricing of local calls and mobile telephones have also evolved to include various pricing strategies for customers to choose

from. The concept of 4P marketing mix has also gradually seeped into the organization. The role of the public works department has slowly changed from main to a supporting one, sustaining activities and demands of marketing, while the supporting role of marketing department has slowly changed to that of a principal one, satisfying customer needs and value and becoming the main axis of the company, changing the organization from being product-oriented to marketing-oriented.
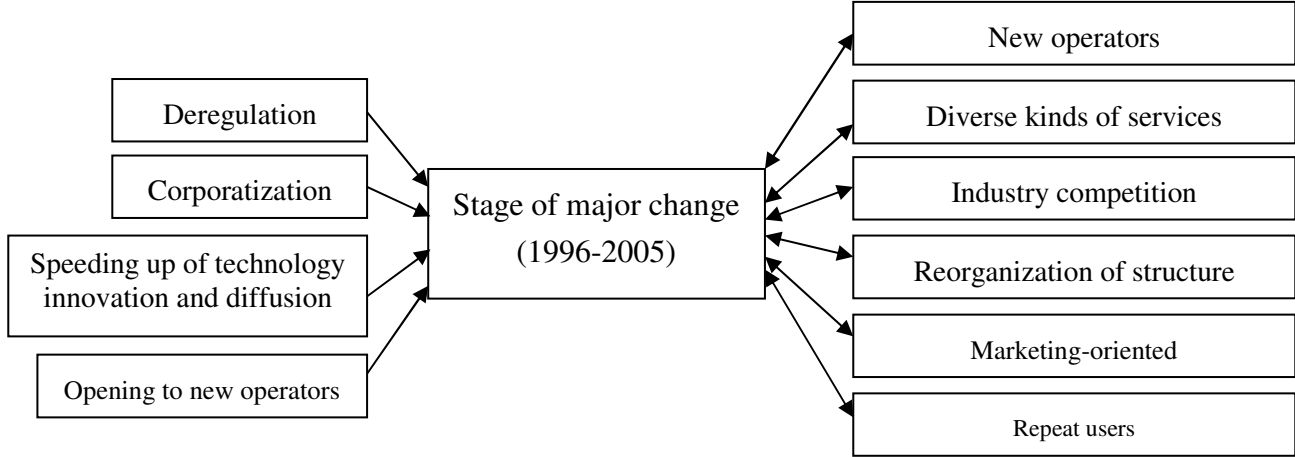
Figure 4-2 Second stage of the case company: Stage of major change

IV. Third stage: Competitive Stage (2005-present)

Continuously innovating is an important path for businesses in gaining a competitive position. In his work "National Competitive Advantages," the American strategic management expert Porter (1990) conducted a competitiveness study on ten countries that achieved success in different industries. He concluded that the most important factor for their success is continuous innovation. In encouraging its employees to innovate, the case company facilitated establishing innovation teams among its various branch offices and business units, placing individuals of different fields of expertise together and using various forms of competitions, awards, and trainings, along with sending them to participate in trainings for creativity facilitators under the Industrial Technology Research Institute of Taiwan. This allows them to drive innovation, to be familiar with an innovative environment, and to cultivate innovative habits in order to achieve the company's objective of innovation within and outside of the industry and assure its sustainable management.

In terms of service quality, the case company has the country's largest service center, providing various telecommunications services, accounts, technology, and consulting services round of the clock. There are also 389 other service centers distributed around Taiwan, providing counter service for customers. Aside from surveys on counter service satisfaction conducted by its international branches to know its service performance, CHT has also introduced and received certification on international service quality from SGS and ITA in 2008 and has contracted a management consulting firm to conduct mystery customer surveys and "service quality reengineering: heart-touching services" with the aim of enhancing service quality and responding to an intensive competitive environment.

In terms of enhancing its marketing, the case company, after its privatization, regularly holds a marketing consensus meeting to strengthen its overall marketing capability, setting objectives, strategies, and methods for its central focus of the business for the year. It introduces a performance appraisal system using key performance indicators (KPI) and

marketing programs to allow everyone to form a consensus and have clear objectives and also to regularly perform reviews. In this manner, it is able to reach its mission. Especially under the wish for integrity in its operations from the chairman, it concentrates on how to enhance the brand image of the company to facilitate trust of the customers to the company, thus becoming the niche of its marketing.

Corporate social responsibility is of great importance and influence on the long-term development and prosperity of businesses as it demonstrates the human aspect of businesses. Gerald Levin, president and CEO of Time Warner Inc., has once said, "Our position as the world's leading media and entertainment company could not have been reached – and could not have been sustained – solely from business success. It rests equally on our tradition of social responsibility and community involvement." Simply put, social responsibility is the decision made by business managers that includes moral and ethical considerations under the law and market economy. Since 2005, the case company has been a member of the World Business Council for Sustainable Development. With the purpose of world sustainable development, it has driven electronic-billing, usage of water recovery systems to save energy and reduce carbon footprint, voluntary participation of all staff in charity activities, assistance to disadvantaged groups, cooperation with the government in promoting broadband services, and lessening of the urban-rural digital divide.
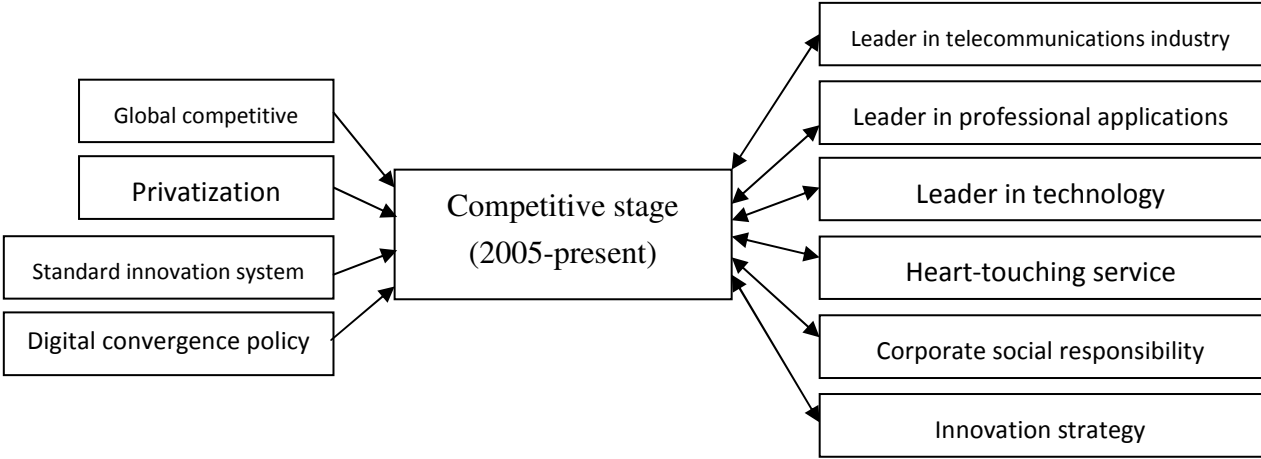
Figure 4-3 Third stage of the case company: Competitive stage

CONCLUSION AND RECOMMENDATIONS

From being a government organization at its early age to being a state-owned enterprise, the case company is the most experienced and largest integrated telecommunications company and is one of the most important partners for international telecommunications cooperation with distinctive achievements in promoting global real-time communication, improvement social life, and enhancing economic efficiency.

In the competitive environment after the complete liberalization of the telecommunications market, the case company can still be able to lead in telecommunications, professional applications, and technology in Taiwan by possessing a business culture of continuous learning and growth. As the Chairman of the case company said, "The long-standing legacy and success of Chunghwa Telecom have been built on principles of good corporate governance, sound ethics, strong integrity, and respect. Maintaining these highest standards of governance, integrity, responsibility, and accountability is important to its continued long-term growth and success."

Service quality, innovation strategy, and marketing have each long been followed with interest by both the academe and practitioners and have also been regarded as the source for achieving sustainable competitive advantage for businesses. However, studies delving into the relationships among the three have been scarce with even fewer qualitative researchers. Thus, the contribution of this study to the academe is to examine its evolution by using qualitative analysis and bringing in "marketing" within the quantitative research on the relationship between service quality and innovation strategy. It is known that after privatization, state-owned enterprises can only be established in the market and achieve the objective of sustainable management through enhancement of service quality, innovation of technology and service, and proper marketing in facing intensive competition in the market.

The case company selected for this study is a company that has been relatively successful in its transformation and one that possesses international competitiveness. Through observations, interviews, and collection and analysis of secondary data, this study has constructed the transformation process of the case company and the business models for service quality, innovation strategy, and marketing that have equally evolved (as seen in Figure 5-1) which can be used as reference for real-world applications. After its corporatization in 1996, the case company has already passed the first phase of market testing for liberalizing mobile communications services. Following its privatization, it has aggressively improved its service quality and ensured that it is leading in technology by employing various innovation strategies, while providing the most advanced services to create an innovation competitive advantage. In continuously strengthening marketing at all levels and establishing a common view on "all-staff marketing," it has smoothly undergone transformation to face competition during the liberalization of telecommunications, in effect playing a leading role in telecommunications in Taiwan.

Stages in the development of the telecommunications industry	**Before 1996** High control system	**1996-2000** Opening of mobile communications services	**2001-present** Opening of fixed-line network services
Stages in the development of the case company	**Before 1996** Monopoly stage	**1996-2005** Stage of major change	**2006-present** Competitive stage

Service quality	**Before 1996** • Provided basic communications functions for the people • Refrained from having public discontent	**1996-2005** Met the demands and value of customers	**2006-present** • Seek customer satisfaction • Focus on customers
Innovation strategy	**Before 1996** Product innovation	**1996-2005** Product/service innovation	**2006-present** • Technology innovation • Business model innovation • Service type innovation
Marketing	**Before 1996** • Production-oriented • First-time customers • Local market	**1996-2005** Growth of marketing	**2006-present** • Market-oriented • Social marketing

Figure 5-1 Co-evolving business models for service quality, innovation strategy, and marketing of the case company

After undergoing trial for transformation, an enterprise must seek channels to expand room for continuous development through innovation and begin attempting to undertake challenges in the new market to rise as valuable multinational enterprises and to enhance competitiveness (Seuring & Muller, 2008; Bhagwati, 2004; Hart, 2005; Prahalad, 2005). From innovation in the industry and repositioning as development for a green technology, effective use of resources must be maintained consistently and satisfy the base of the economic pyramid to contribute to society (Hart, 2005). From the present development of the case company in line with internal and external situations, this study provides strategic planning for the future from the perspective of innovation and competitive advantage sustainability (as seen in Figure 5-2). It is hoped that this can be used as a reference for the case company in providing a faster development and value creation for sustainable management.

Tomorrow
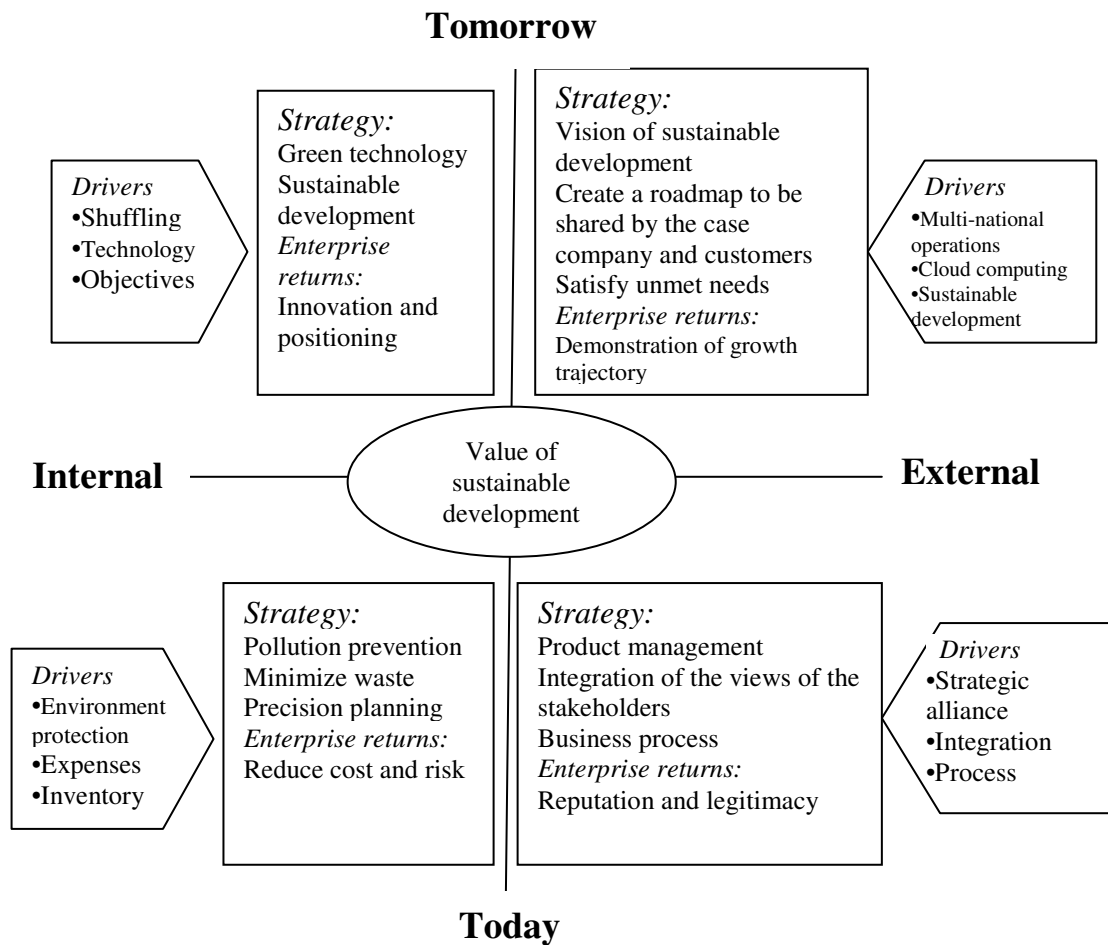
Figure 5-2 Strategic planning for sustaining competitive advantage after transition

Since this study only employed a single case study, we feared that it lacks representativeness and might be unable to generalize the observed phenomenon. Also, this study mainly conducted case interviews for its collection of primary data. This method might be constrained by the cognitive judgment of the respondents based on memory and subjective opinions that will inevitably result in biases and would likely affect the findings. This is the main limitation of this study. For subsequent researchers, this study suggests to include the subject regarding the dynamic ability of sustaining competitive advantage and with particular emphasis on the cumulative process of ability in each stage.

REFERENCES

1. Chunghwa Telecom, http://www.cht.com.tw .

2. Chi-Kuo Mao (2004). "Change must be understood before regulating, knowing change can facilitate long-term existence." Revolution. Taipei: Baoding.

3. Chi-Kuo Mao (2005). "Principles of Organization Change," *Leadership in New Society*. Taipei: Baoding.

4. Maw-Liann Shyu & Hui-Ling Shih, (2007). "A Study on Organizational Innovation System: The Case of P&G." *Chaoyang Business and Management Review*. Vol. 6 No. S, pp. 47-48.

5. Syue-yue Xu , (2005). "Telecom Marketing Management." Beijing: Beijing University of Posts and Telecommunications Press.

6. National Communications Commission, http://www.ncc.gov.tw

7. Directorate General of Telecommunications, http://www.motc.gov.tw

8. Bell , S. J., Auh, S. & Smalley, K., （2005）, "Customer Relationship Dynamics: Service Quality and Customer Loyalty in the Context of Varying Levels of Customer Expertise and Switching Costs," *The Journal of Academy of Marketing Science*, Vol.33 No.2, pp. 169-183.

9. Benner, M. J. & Tushman, M. L. , （2003）, "Explotation, exploration, and process management: The productivity dilemma revisited." *Academy of Management Review*, Vol. 28 No.2, pp.238-256.

10. Brady, M. K., Robertson C. J. & Cronin J. J., （2001）, "Managing behavioral intentions in diverse cultural environments: an investigation of service quality, service value, and satisfaction for American and Ecuadorian fast-food customers," *Journal of International Management,* Vol.7 No.2, pp. 129-149.

11. Bhagwati , J.N., （2004）, *"In defense of globalization."* New York: Oxford University Press.

12. Crowe, D. & Brennan, L. , （2007）, "Environmental Considerations within Manufacturing Strategy,an International Study." *Business Strategy and the Environment*, Vol.16 No.4, pp.266-289.

13. Engel, J. F., Blackwell, R. D., & Miniard, P. W., （2001）, *"Consumer Behavior."* (10th ed.), New York: The Dryden Press.

14. Eisenhardt, K. M. & Graebner, M. E.,（2007）,"Theory Building From Cases: Opportunities and Challenges," *Academy of Management Journal,* Vol.50 No.1, pp. 25–32.

15. Gilbert, J. T., （1994）, "Choosing an Innovation Strategy: Theory and Practice." *Business Horizons*, Vol.37 No.6, pp.16-22.

16. Gobeli, D. H. & Brown, D. J., （1987）,"Analyzing Product Innovations," *Research Management*, Vol.30 No.4, pp.25-31.

17. Hart, S., （2005）, *"Capitalism at the crossroads: the unlimited business opportunities in solving the world's most difficult problems."* Indianapolis, IN:Wharton.

18. Hinings, C. R. , （1997）, "Reflections on Processual Research," *Scandinavian Journal of Management*, Vol.13 No.4, pp. 493-503.

19. Huang Jen-Hung & Lin Chia-Yen,（2005）, "The explanation effects on consumer perceived justice, satisfaction and loyalty improvement: an exploratory study," *Journal of American Academy of Business*, Vol.7 No.2, pp. 212-218.

20.Kotler, P.,（2004）, Marketing management (11th ed.). NJ:Prentice-Hall.

21.Lewis, R. C. & Booms, B. H.,（1983）, "The Marketing Aspects of Service Quality," in Emerging Perspectives on Services Marketing, Edited by Berry, L. L.,Shostack, G. L. and Upah, G. D., AMA,pp.99-107.

22.Olorunniwo, F. & Hsu, M. K.,（2006a）, "A typology analysis of service quality, customer satisfaction and behavioral intentions in mass services," *Managing Service Quality,* Vol.16 No.2, pp. 106-123.

23.Olorunniwo, F., Hsu, M. K. & Udo, G. J.,（2006b）, "Service quality, customer satisfaction, and behavioral intentions in the service factory," *The Journal of Services Marketing,* Vol.20 No.1, pp.59-72.

24.Parasuraman, A., Berry, L. L. & Zeithaml, V. A.,（1985）, "Refinement and Reassessment of The SERVQUAL Scale," *Journal of Retailing,* Vol.67 No.4,pp. 420-450.

25.Parasuraman, A., Zeithaml, V. A. & Berry, L. L.,（1985）, "A Conceptual Model of Service Quality and Its Implication for Future Research," *Journal of Marketing*, Vol.49 No.4,pp. 41-50.

26.Pettigrew, A.,（1990）, "Longitudinal Field Research on Change: Theory and Practice," *Organizational Science*, Vol.1 No.3, pp. 267-292.

27.Prahalad, C.K., （2005）, "The fortune at the bottom of the pyramid: eradicating poverty through profits: enabling dignity and choice through markets." Upper Saddle River, NJ: Wharton School Pub.

28.Porter, M. E.,（1990）, "The Competitive Advantage of Nations." The Free Press, New York.

29.Robbins, S. P. (2001). "Organizational behavior."   (9th ed.). NJ: Prentice-Hall.

30.Sehumpeter, J. A.,（1934）, "The Theory of Economic Development," Harvard University Press, Cambridge, MA.

31.Seuring, S. & Muller, M.,（2008）, "From a literature review to a conceptual framework for sustainable supply chain management." *Journal of Cleaner Production* .

32.Sinclair, R. R., Fleming, W. D., Radwinsky, R., David,C. R., Jill, C.H.,（2002）, "Understanding Patients' Reactions to Services: The Role of Personal Service Values in Heart Patient Satisfaction and Wellness," *Journal of Applied Social Psychology*, Vol.32 No.2, pp. 424-442.

33.Wakefield, R. L.,（2001）, "Measuring Service Quality：A Reexamination and Extension," *The CPA Journal*, August, pp.55-68.

34.Weerawardena, J.,（2003）,"The role of marketing capability in innovation-based ompetitive strategy" *Journal of Strategic Marketing* , Vol.11No. 1, pp.15-35.

35.Yin, R.,（1984）, *"Case study research: Design and methods ,"* (1st ed.). Beverly Hills, CA: Sage Publishing.

36. Yu Chang-Hsi, Chang Hsiu-Chen, Huang Gow-Liang,（2006）, "A Study of Service quality, Customer Satisfaction and Loyalty in Taiwanese Leisure Industry," *Journal of American Academy of Business,* Vol.9 No.1, pp. 126-132.

37. Zheng Zhou, K., Yim, C. K., & Tse, D. K.,（2005）, "The effects of strategic orientations on technologe and market-based breakthrough innovations." *Journal of Marketing,* Vol.69 No.2, pp.42-60.

THE OPTIMAL SHAPE OF TRIANGULAR YIELD DISTRIBUTION IN AN EOQ MODEL WITH PLANNED SHORTAGES

Farrokh Nasri, Frank G. Zarb School of Business, Hofstra University, Hempstead, NY 11590, 516-463-4244, Farrokh.Nasri@hofstra.edu

Javad Paknejad, Frank G. Zarb School of Business, Hofstra University, Hempstead, NY 11590, 516-463-4244, M.J.Paknejad@hofstra.edu

John F. Affisco, Frank G. Zarb School of Business, Hofstra University, Hempstead, NY 11590, 516-463-4244, John.F.Affisco@hofstra.edu

ABSTRACT

This paper extends the results of a series of recent papers which focused on the analytical investigation of the economic trade-offs associated with investment decisions aimed at changing the parameters of yield distribution, and thereby improving process quality, in an EOQ model with planned shortages and random production yield. Production yield is defined as the percentage of perfect quality items in a purchased or manufactured lot. Assuming that the production yield follows a triangular probability density function, the paper presents explicit results for the optimal values of shape parameter, lot size, backorder level and the expected total annual cost when the shape parameter of the yield distribution follows a logarithmic investment function.

Keywords: Inventory Theory, Quality

1. INTRODUCTION

The main purpose of any inventory system is to harmonize, at a minimum cost, the mismatches between supply and demand processes that do not fit neatly together. One of the most challenging of these mismatches is when unsatisfied demand or shortage is realized as a result of imperfect quality of purchased or manufactured items. However, classical inventory models generally overlook the effect of quality on lot size and shortage level. The reality of zero inventory systems [3] provoked researchers to start studying the possible relationship between lot size, backorder level, and quality when developing new inventory paradigms. At the outset, Porteus [12] and Rosenblatt and Lee [13] investigated the effect of process quality on lot sizes in the traditional Economic Order Quantity (EOQ) and Economic Manufacturing Quantity (EMQ) models, respectively. Moinzadeh and Lee [4] investigated the effect of defective items on the order quantity and reorder point of a continuous-review inventory model with Poisson demand and constant lead time. Paknejad, Nasri, and Affisco [11] extend this work to consider stochastic demand and constant lead time in a continuous- review (s,Q) model. Paknejad Nasri, and Affisco

[10] develop a quality-adjusted EOQ model with planned shortages and study the impact of efforts aimed at improving both quality and flexibility on the optimal lot size and backorder level.

San-Jose, Garcia-Laguna, and Sicilia [14] study an EOQ model with partial backlogging where shortage cost is a non-decreasing function of customer waiting time for the item.

The above authors mainly assume that the manufacturer operates a process that is in statistical control. Hence, the production yield, defined as the percentage of perfect items in each lot, is assumed to be known and constant. This assumption implies that the number of non-defective items in each lot is binomial. Such an assumption is also made in Affisco, Paknejad, and Nasri [1] for the case of Economic Order Quantity (EOQ) model and Affisco, Paknejad, and Nasri [2] for the case of Joint Economic Lot Size (JELS) model.

In two relatively new papers, Nasri, Paknejad, and Affisco [6,7] gave careful consideration to the relationship between lot size and quality for processes that are in the initial stages of implementing a quality program and, hence, have not yet achieved the state of statistical control. Specifically, the authors assumed that the production yield is random, rather than constant, and adjusted the EMQ and EOQ with planned shortages models for the quality factor. In addition to the general relationships, the authors provided closed form expressions for a few special cases of production yield distribution. In both [6] and [7], the shortage cost is levied based on the number of shortages per unit of time. Paknejad, Nasri, and Affisco [9] extended the results in [7] by combining two distinct shortage costs: (1) based on the number of shortages per unit time, and (2) based on the average number of shortages, irrespective of duration of shortage.

The analytical results developed in [6,7,9] reinforce the prevalent understanding that the relationships between the production yield, lot size, and backorder level are vitally important to all manufacturing firms. However, these models simply overlook the reality that most manufacturing organizations routinely invest in yield improvement programs and, implicitly, assume that all the parameters of the production yield distribution are known, constant, and fixed. Nasri, Paknejad, and Affisco [5] relaxed this assumption and treated the parameters of yield distribution as decision variables, rather than constants, and studied the impact of investment efforts aimed at altering these parameters on the optimal values of policy variables in the context of an EOQ model with planned backorders, random yield, and time dependent shortage cost. Paknejad, Nasri, and Affisco [8] extended the results in [5] by combining two alternative assumptions commonly made about the cost of shortage in their analysis. In both [5] and [8] explicit expressions for the optimal values of yield location parameter, yield scale parameter, lot size, and backorder level when yield follows a uniform probability density function.

The main objective of this paper is to extend the results in [5] to the case of triangular yield distribution. We assume that the location and scale parameters of yield distribution are 0 and 1, respectively. Hence, the shape parameter, C, floats between 0 and 1 (i.e. $0 \leq C \leq 1$). Using these assumptions, the paper considers the shape parameter of triangular yield distribution , C, as a decision variable and studies the impact of investment efforts

aimed at improving quality, by altering C, on lot size and backorder level. Closed form expressions for the optimal values of policy variables are provided for the case of logarithmic investment function.

2. MODEL AND ASSUMPTIONS

The basic model considered in this paper is the undiscounted Economic Order Quantity (EOQ) model that allows two distinct types of shortage with the following total annual cost function

$$C_{Trad}(Q,S) = \frac{D}{Q}K + \frac{(Q-S)^2}{2Q}C_h + \frac{S^2}{2Q}C_b \tag{1}$$

where

D = Annual Demand in units,
Q = Lot size per order,
S = Number of units backordered,
K = Ordering cost per order,
c_h= Holding cost per unit per year,
c_b= Backordering cost per unit per year.

Upon using traditional optimization techniques, we find the well-known expressions for the optimal values for the lot size, Q^*_{Trad} , units backordered, S^*_{Trad} , and the annual cost, C^*_{Trad}

$$Q^*_{Trad} = \sqrt{\frac{2DK}{c_h}\left(\frac{c_h+c_b}{c_b}\right)} \tag{2}$$

$$S^*_{Trad} = \sqrt{\frac{2DK}{c_b}\left(\frac{c_h}{c_h+c_b}\right)} \tag{3}$$

and

$$C^*_{Trad} = \sqrt{2DK\,c_h\left(\frac{c_b}{c_h+c_b}\right)} \tag{4}$$

One of the implicit assumptions in the above derivations is that all units produced by the vendor, in response to the purchaser's order, are of acceptable quality. Now, assume that this is not the case and that the proportion of non-defective items in each lot is a random variable. Further assume that the purchaser inspects the entire lot upon receipt. We assume that the purchaser's inspection process is perfect and that all rejected items are returned to the vendor at no cost to the purchaser. In addition, the inspection cost is paid by the vendor. Based on this scenario, we adjust the traditional EOQ with backorders and two alternative shortage costs for the quality factor as follows:

Let

λ = Yield, being defined as the proportion of non-defective items in an order lot,

$\quad$ $\lambda \, \varepsilon \, [0, 1]$, a continuous random variable,

$f(\lambda)$ = Probability density function of λ,

μ = Mean of λ,

σ^2 = Variance of λ,

$E(.)$ = Mathematical expectation,

$y = \lambda Q$ = Number of non-defective items in a lot,

$c(y)$ = Total cost per cycle given that there are y non-defective items in the lot of size Q,

$T = y/D$ = Cycle time, time between two successive placement of orders,

$C_{Adj}(Q,S)$ = Expected total cost per year.

The total cost per cycle is

$$c(y) = K + \frac{(y-S)^2}{2D} c_h + \frac{S^2}{2D} c_b = K + \frac{\overline{\lambda Q - S}^2}{2D} c_h + \frac{S^2}{2D} c_b \tag{5}$$

The average cycle time and cycle cost are

$$E(T) = \frac{E(y)}{D} = \frac{E\left[\lambda Q\right]}{D} = \frac{Q}{D} \mu \tag{6}$$

and

$$E(c) = K + \frac{Q}{2D}\left[\sigma^2 + \mu^2\,Q - 2\,\mu S\right]C_h + \left(\frac{c_h + c_b}{2D}\right)S^2 \tag{7}$$

The expected total annual cost is

$$EAC_{adj}(Q,S) = \frac{D\,K}{\mu Q} + \left[\frac{\sigma^2 + \mu^2\,Q}{2\mu} - S\right]c_h + \frac{(c_h + c_b)S^2}{2\mu Q} \tag{8}$$

In what follows we assume that the probability density function of λ is triangular with location parameter = 0, scale parameter = 1, and shape parameter = C. That is,

$$f(\lambda) = \begin{cases} \dfrac{2\lambda}{C} & for\ \ 0 \le \lambda \le C \\[2mm] \dfrac{2(1-\lambda)}{1-C} & for\ \ C \le \lambda \le 1 \end{cases}, \tag{9}$$

In this case

$$\mu = \frac{1+C}{3} \quad , \tag{10}$$

and

$$\sigma^2 = \frac{1+C^2-C}{18} \quad . \tag{11}$$

Substituting (10) and (11) into (8) and using classical optimization techniques, the optimal values for the order quantity, $Q^*_{adj,t}$, units backordered, $S^*_{adj,t}$, and expected total annual cost, $EAC^*_{adj,t}(S,Q)$, are easily found as follows:

$$Q^*_{adj,t} = \left(\frac{3}{1+C}\right) \sqrt{\frac{2DK}{c_h\left[\left(\frac{1}{2}\right)\left(\frac{1-C+C^2}{1+2C+C^2}\right)+\left(\frac{c_b}{c_h+c_b}\right)\right]}}, \tag{12}$$

$$S^*_{adj,t} = \sqrt{\frac{2DK\left(\frac{c_h}{c_h+c_b}\right)}{c_b\left[\left(\frac{1}{2}\right)\left(\frac{1-C+C^2}{1+2C+C^2}\right)\left(\frac{c_h+c_b}{c_b}\right)+1\right]}}, \tag{13}$$

and

$$EAC^*_{adj,t} = \sqrt{2DKc_h\left[\left(\frac{1}{2}\right)\left(\frac{1-C+C^2}{1+2C+C^2}\right)+\frac{c_b}{c_h+c_b}\right]}. \tag{14}$$

3. THE OPTIMAL YIELD PARAMETER MODEL

The policy variables in the model of section 2 are the order quantity, Q, and backorder level, S, for a fixed shape parameter of yield distribution, C. The value of this parameter determines the values of both mean and standard deviation of yield distribution. As C approaches one from 0, yield rate increases but yield variability first decreases to reach its minimum at $C=.5$ and then increases. In this paper, we consider the option of investing to change the location parameter, C. In order to evaluate the economic trade-offs associated with this investment option, we introduce a companion yield parameter, Ω, as follows:

$$\Omega = \frac{1-C+C^2}{1+2C+C^2} \qquad \text{for} \qquad 1/4 \le \Omega \le 1 \ , \tag{15}$$

Please note that as C increases from 0 to 1, Ω decreases from 1 to 1/4. Thus, reducing Ω implies increasing C and, hence, improving quality.

Now, we consider Ω to be a decision variable and seek to minimize the average annual cost composed of, investment cost to change Ω, ordering, shortage, and holding costs. Specifically, we seek to minimize

$$C(Q,S,\Omega) = i a_\Omega(\Omega) + EAC_{adj,t}(Q,S) \tag{16}$$

subject to

$$1/4 \le \Omega \le 1, \tag{17}$$

where i is the cost of capital, $a_\Omega(\Omega)$ is a convex and strictly decreasing function of Ω representing the cost of reducing the yield parameter to the level Ω, $EAC_{adj,t}(Q,S)$ is the sum of all inventory related costs given in equation (8) for the case of triangular yield distribution, and Ω_0 is the original yield parameter.

In order to solve this optimization problem, we use a sequential approach similar to the one suggested by Porteus [12]. In this case, we hold Ω fixed, optimize over Q and S to obtain $Q_{adj}^*(\Omega)$ and $S_{adj}^*(\Omega)$, given by equations (12) and (13) with $\dfrac{1-C+C^2}{1+2C+C^2} = \Omega$, and then optimize over Ω. That is, we seek to minimize

$$w(\Omega) = i.a_\Omega(\Omega) + EAC_{adj,t}(\Omega) \tag{18}$$

where $EAC_{adj,t}\Omega)$ is given by equation (14) modified for Ω as follows:

$$EAC_{adj,t}^*(\Omega) = \sqrt{2DKC_h \left[\frac{\Omega}{2} + \frac{c_b}{c_h + c_b} \right]}. \tag{19}$$

Please note that if the optimal Ω obtained in this way fails to satisfy restriction (17), then we should not make any investment and the results of the quality adjusted EOQ model with backorders hold. Of course it may not always be possible to carry out the minimization outlined above except for some special cases of $a_\Omega(\Omega)$. In what follows we consider the case where the investment function is of logarithmic form.

4. THE LOGARITHMIC INVESTMENT FUNCTION

Because of its mathematical tractability and practical appeal the logarithmic investment function is frequently used in research dealing with quality improvement as well as flexibility improvement through setup cost reduction [5, 10, 11, 12]. Its use here is based on the idea that yield improvement should exhibit decreasing marginal return. In this case

the yield parameter, Ω, declines exponentially as the investment amount, a_Ω, is increased. That is

$$a_\Omega(\Omega) = \frac{1}{\Gamma} Ln \frac{\Omega_0}{\Omega} \qquad for \qquad 1/4 \leq \Omega \leq \Omega_0 \leq 1, \tag{20}$$

where Γ is the percentage decrease in Ω per dollar increase in $a_\Omega(\Omega)$. Here our main objective is to minimize $w(\Omega)$ after substituting (19) and (20) into (18).

Theorem: If $1/4 \leq \Omega_0 \leq 1$ and Γ is strictly positive, then the following hold:

a) The optimal value of the yield parameter is given by

$$\Omega^{**} = \min \ \Omega_0, \Omega_{imp,t} \ , \tag{21}$$

where

Ω_0 = the original yield parameter and

$$\Omega_{imp,t} = \left(\frac{i}{\Gamma}\right)^2 \left[\frac{1 + \sqrt{1 + 2DKC_h \left(\frac{2\Gamma}{i}\right)^2 \left(\frac{C_b}{C_h + C_b}\right)}}{2DKC_h} \right]. \tag{22}$$

(b) The resulting optimal shape parameter of triangular yield distribution is

$$C_t^{**} = \max \ C_0, C_{imp,t} \ , \tag{23}$$

where C_0 is the original shape parameter, and

$$C_{imp,t} = \frac{1 + 2\Omega_{imp,t} - \sqrt{12\Omega - 3}}{2 \ 1 - \Omega_{imp,t}}, \tag{24}$$

(c) The optimal values for the order quantity, $Q_{adj,t}^{**}$, units backordered, $S_{adj,t}^{**}$, and expected total annual cost, $EAC_{adj,t}^{**}$, are as follows

$$Q_{adj,t}^{**} = \begin{cases} Q_{adj,t}^* & if \ \Omega_{imp,t} \geq \Omega_0 \\ Q_{imp,t} & if \ \Omega_{imp,t} < \Omega_0 \end{cases}, \tag{25}$$

$$S_{adj,t}^{**} = \min(S_{adj,t}^*, S_{imp,t}) = \begin{cases} S_{adj,t}^* & if \ \Omega_{imp,t} \geq \Omega_0 \\ S_{imp,t} & if \ \Omega_{imp,t} < \Omega_0 \end{cases}, \tag{26}$$

$$EAC_{adj,t}^{**} = \min(EAC_{adj,t}^*, EAC_{imp,t}) = \begin{cases} EAC_{adj,t}^* & if \ \Omega_{imp,t} \geq \Omega_0 \\ EAC_{imp,t} & if \ \Omega_{imp,t} < \Omega_0 \end{cases}, \tag{27}$$

where $Q^*_{adj,t}$, $S^*_{adj,t}$, and $EAC^*_{adj,t}$ are given by (12), (13), and (14), respectively, with $C = C_0$,

$$Q_{imp,t} = \left[\frac{1 + \Omega_{imp,t} - \sqrt{3\left(\Omega_{imp,t} - \frac{1}{4}\right)}}{1 - \Omega_{imp,t}} \right] \sqrt{\frac{2DK}{c_h\left[\left(\frac{\Omega_{imp,t}}{2}\right) + \left(\frac{c_b}{c_h + c_b}\right)\right]}}, \qquad (28)$$

$$S_{imp,t} = \sqrt{\frac{2DK\left(\frac{c_h}{c_h + c_b}\right)}{c_b\left[\left(\frac{\Omega_{imp,t}}{2}\right)\left(\frac{c_h + c_b}{c_b}\right) + 1\right]}}, \qquad (29)$$

and

$$EAC_{imp,t} = \sqrt{2DKC_h\left[\left(\frac{\Omega_{imp,t}}{2}\right) + \frac{c_b}{c_h + c_b}\right]}. \qquad (30)$$

Careful examination of the above results reveal that $\Omega_{imp,t}$, $C_{imp,t}$, $Q_{imp,t}$, $S_{imp,t}$, and $EAC_{imp,t}$ do not depend on the initial yield parameter, Ω_0. Furthermore, when $\Omega_{imp,t} \geq \Omega_0$, then no investment is made and the results of the quality adjusted model with triangular yield distribution hold. In such case Ω_0 will be used in place of $\Omega_{imp,t}$ and equations (28), (29), and (30) will be replaced by equations (12) through (14). Details of the proofs are omitted.

5. CONCLUSION

This paper extended the analytical results of a series of recently developed models pertaining the impact of changes in the parameters yield distribution on the optimal lot size and backorder level. The paper studied the economic trade-offs associated with investment efforts aimed at improving the production yield through changes in the shape parameter of yield distribution. Assuming that the yield probability density function is triangular, the paper presented explicit results for the optimal values of shape parameter, lot size, backorder level, and expected total annual cost where the shape parameter of triangular yield distribution follows a logarithmic investment function.

REFERENCES

[1]   Affisco, J.F., Paknejad, M.J., and Nasri, F. "Quality improvement in the economic order quantity model," *Proceedings of the Northeast Decision Sciences Institute*, pp. 256-258, Atlantic City, New Jersey, 2000.

[2]   Affisco, J.F., Paknejad, M.J., and Nasri, F. "Quality improvement and setup reduction in the joint economic lot size model," *European Journal of Operational Research*," 2002, 142, 497-508.

[3]   Hall, R.W. *Zero Inventories.* Dow Jones-Irwin, Homewood, IL, 1983.

[4]   Moinzadeh, K., and H.L., Lee, "A continuous review inventory model with constant resupply time and defective items," *Naval Research Logistics*, 1987, 34, 457-467.

[5]   Nasri, F., Paknejad, J., & Affisco, J. F., "The Optimal Parameters of Yield Distribution in An EOQ Model with Planned Shortages," *Western Decision Sciences Institute Proceedings,* pp. 4571-4577, April 2010.

[6]   Nasri, F., Paknejad, J., & Affisco, J. F., "An EMQ Model with Planned Shortages and Random Defective Units", *Decision Sciences Institute Proceedings,* pp. 10541-10546, November 2009.

[7]   Nasri, F., Paknejad, J., & Affisco, J. F., "Lot Size Determination in an EOQ Model with Planned Shortage and Random Defective Units", *Northeast Decision Sciences Institute Proceedings*, pp. 403-408, on CD-ROM, April 2009.

[8]   Paknejad, J., Nasri, F. & Affisco, J.F., "The Optimal Parameters of Yield distribution in An EOQ Model with Two Types of Shortage," *Decision Sciences Institute Proceedings*, Forthcoming.

[9]   Paknejad, M.J., Nasri, F. & Affisco, J.F., "An EOQ Model with Two Types of Shortage and Random Defective Units," *2010 Northeast Decision Sciences Institute Proceedings*, pp. 439-444, 2010.

[10]   Paknejad, M.J., Nasri, F. & Affisco, J.F., "Flexibility and Quality Improvement in the EOQ with Planned Shortage Model," *Northeast Decision Sciences Institute Proceedings*, pp. 257-259, 2001.

[11]   Paknejad, M.J., Nasri, F. & Affisco, J.F., "Defective Units in a Continuous Review (s,Q) System," *International Journal of Production Research*, 1995, 33(10), 2767-2777.

[12]   Porteus, E., "Optimal Lot Sizing, Process Quality Improvement and Setup Cost Reduction," *Operations Research*, 1986, 34(1), 137-144.

[13]   Rosenblatt, M.J. and Lee, H.L., "Economic Production Cycles with Imperfect Production Processes," *IIE Transactions*, 18(3), 1986, 48-55.

[14]   San-Jose, L. A., Garcia-Laguna, J., and Sicilia, J. "An Economic Order Quantity Model with Partial Backlogging Under General Backorder Cost Function," TOP, 17(2), 2009, 366-384.

AN AGGREGATE INVENTORY-BASED MODEL FOR PREDICTING REDEMPTION AND LIABILITY IN LOYALTY REWARD PROGRAMS INDUSTRY

Aaron Luntala NSAKANDA
Sprott School of Business
Carleton University
Ottawa, ON, Canada,
K1S5B6
aaron_nsakanda@carleton.ca

Moustapha DIABY
School of Business Administration,
University of Connecticut
Storrs, CT 0629-2041, USA
Moustapha.diaby@business.uconn.edu

Yuheng CAO
Sprott School of Business
Carleton University
Ottawa, ON, Canada,
K1S5B6
ycao2@connect.carleton.ca

ABSTRACT

Loyalty Reward programs (LRPs) are marketing programs aimed at rewarding customers for repeat purchasing of a product or a service. Although different types of LRPs exist today across a spectrum of industries (travel, hotel, retail, telecommunication, banking, gasoline, etc.), most of the modern systems have their roots from AAdvantage®, the loyalty program introduced by American Airlines in 1981 (also called frequent flyer program in the airline industry). Consumers in these systems are given incentives or rewards for repeat business, which in turn serve as motivation for them to continue buying a product or service. In general, these systems involve at minimum a promotional currency (e.g. points or miles); single or multiple reward tier(s); a comprehensive database of individual consumers, demographics and detailed transaction information; and an advanced technology to manage the program (e.g. redeem rewards directly or through internet), operate the contact center, and to analyze the members' database. Hence broadly, in a typical LRP, customers become members of the program, earn points (based on some specified "accumulation scheme") on their purchases of products or services throughout the network of LRP's commercial partners. These points can be redeemed based on a "reward chart" pre-established by the firm that owns the LRP (i.e., host firm). Points that are not redeemed are saved in the customer's account (under some conditions, e.g., being

active) and constitute the LRP outstanding balance ("liability"). Points earned by customers during a given period (e.g., a year) constitute the LRP issued points ("accumulation"), whereas points redeemed by customers for rewards during a given period (e.g., a year) constitute the LRP redeemed points ("redemption").

Loyalty reward programs have increasingly become prevalent in recent years. For instance, in the airline industry alone, more than 130 companies currently have a LRP, and 163 million people throughout the world are enrolled in their programs. Geographically, LRPs have been quite popular in the United States, United Kingdom, Canada, and a host of other countries. Some studies show that 90% of Americans and 92% of UK consumers are members of at least one LRP. In the Canadian market settings, a special article featuring loyalty programs reports that according to Visa Canada Association, more than 25 million VISA cards are in circulation in Canada and about 78% of all card holders belong to one or more LRPs. A study from ACNielsen reports that 95% of Canadians belonged to LRPs of department stores, mass retailers, general merchandisers, or warehouse clubs.

Despite the prevalence of LRPs and the increased complexities in their management and control, there are few academic models that specifically deal with LRPs to support planning and operational decision–making. Most of existing work focuses on the leverage of the individual consumer information accumulated in the LRP databases to improve marketing and sales decisions. One of the challenges faced by LRPs is that of developing aggregate and disaggregate predictions of redemption, liability, and accumulation to support short, medium, and long term planning and operational decision-making. LRP managers rely on good predictions of redemption and liability to plan for rewards supply, set program budget, maintain a balance between customer service level and overall costs of rewarding customers, and to assess the

growth of the program and the risk level associated with this growth. The lack of availability of rewards at the time of redemption results in a poor service level and/or an increase of the reward supply costs to meet customers' demands, since the LRP's host firm will have to acquire the additional rewards at a higher cost. On the other hand, too much availability will result in a higher cost as well (although the level of customer service would be high in that case). The unused availability will result in a penalty whenever the LRP's host firm decides to reduce or cancel their reservation of rewards or return unused rewards to reward suppliers (i.e., LRP partners). In effect, in setting up long term contracts with partners, LRP managers must decide the volume of rewards to purchase in advance. This results in the needs for good predictions of redemption. Good forecasts of redemption are also required in establishing proper budgeting plans or forward financial statements. Moreover, good predictions of redemption provide LRPs managers with the ability to develop promotion plans that seek for better management of redemption demand between peak and off-peak periods.

Liability of LRPs is widely recognized in the industry as a risk indicator for firms' future LRP operations. It represents the value of future redemption obligation of points earned by LRP members. LRP organizations will face a higher risk and corresponding challenges in LRP operations when the liability level is too high. Therefore, a good prediction of LRP liability (i.e. with a greater degree of accuracy) provides managers with the ability to anticipate the growth of the program as well as the liability associated risks (e.g. hyperinflation and devaluation of points). Furthermore, risk mitigation plans can be developed which may include strategies such as revision of reward scheme, reward pricing, changes in management policies, etc.

We propose a predictive model of redemption and liability to support short, medium, and long term planning and operational decision-making in Loyalty Reward Programs (LRPs). The

proposed approach is an aggregate inventory model in which the liability of points is modeled as a stochastic process. An illustrative example is discussed as well as a real-life implementation of the approach to facilitate use and deployment considerations in the context of a frequent flyer program, an airline industry based LRP.

UNPACED SERIAL PRODUCTION LINES WITH TWO SIMULTANEOUS SOURCES OF IMBALANCE - THEIR OPERATING CHARACTERISTICS

Sabry Shaaban[1], Tom McNamara[2] and Ahmed Atil[2]

1 Department of Economics, Strategy and Organization
ESC La Rochelle
102 Rue de Coureilles
17024 La Rochelle, France
Email: shaabans@esc-larochelle.fr

2 Department of Finance and Operations
ESC Rennes School of Business
2 Rue Robert d'Arbrissel
35065 Rennes, France
Email: tom.mcnamara@esc-rennes.fr
Email: ahmed.atil@esc-rennes.fr

ABSTRACT

This paper investigates the behaviour of unpaced serial production lines with two concurrent imbalance sources, namely, lines that are simultaneously unbalanced with respect to either their work time means (MTs) and variability - as depicted by the coefficients of variation of processing times (CVs), their MTs and buffer capacities (BCs), or their CVs and BCs. The lines were simulated under stable operating conditions with different values of line length, buffer capacity, coefficient of variation, degree of imbalance, as well as a variety of imbalance patterns. Making use of a number of statistical methods, idle time and average buffer level output data were analysed and compared to a balanced line counterpart. Conclusions concerning the relationships among the exogenous and endogenous variables were made. For the MT and CV imbalance, the configuration providing the lowest idle time turned out to be an inverted bowl pattern for MT, along with a bowl-shaped arrangement for CV, whereas the best configuration that generally generates low average buffer levels was found to be an MT decreasing order, together with a CV bowl allocation. As for unbalanced MTs and BCs, it was found that in terms of idle time the best unbalanced pattern is an MT bowl configuration, coupled with a distribution of buffer capacity as evenly as possible. On the other hand, for average buffer level the best pattern was a monotone decreasing MT order, together with an ascending buffer size order. As regards CV and BC imbalance, the most superior pattern with respect to idle time was identified as a bowl-shaped CV configuration, combined with a buffer placement whereby more buffer capacity is concentrated at the beginning of the line, but for average buffer level the most favourable pattern was a bowl-shaped CV assignment, together with a buffer arrangement under which the buffers are concentrated towards the end of the line.

1

1. INTRODUCTION

Unpaced sequential production lines are treated as queuing systems in tandem. In the interests of "decoupling" the line and limiting the effects of worker variability from one station to another, inter-station buffers are provided to store partially finished work pieces. A typical flow line is depicted in Figure 1, where the stations and buffers are respectively denoted by circles and triangles.
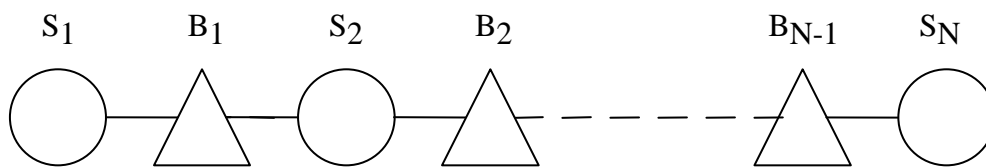
Figure 1. A serial production line

In the design of reliable unpaced production lines, there are a number of issues to be considered if performance is to be improved. The operators at each station along the line work at different speeds or mean processing times (MTs) from each other; therefore, where to place these operators along the line is an important consideration.

The time taken by an operator to complete a task may naturally vary quite considerably. People in general cannot perform a task or a series of tasks again and again at exactly the same speed over a length of time, due to natural differences in their work time variation patterns and complexity and specificity differences in the work elements making up each individual task. This variation can commonly be measured in relative terms using the coefficient of variation (CV).

Another influence on the performance is the buffer size and placement. In theory, the best way to allocate buffer space along the line is evenly. However, this is not always possible for technical reasons which can restrict the availability of buffer space in the line, as a result of which a manager may have to distribute buffer capacity (BC) unevenly.

As a balanced line has generally been considered to be an 'ideal' line that does not exist in real life operations, it is quite interesting to see the effects on line performance resulting from distributing operators with different speeds or variations in operating times into carefully selected positions and placing buffer capacity with care along the line. This sort of research might lead to results which at worst are an improvement on just randomly placing workers anywhere, and at best will outperform the balanced line. The aim of this paper is to study the operating characteristics and efficiency of unpaced serial production lines with two simultaneous sources of imbalance, i.e. unequal mean processing times, variability, or buffer capacities, and compare them to those obtainable by a balanced line counterpart.

This paper is organised as follows. In Section 2 the relevant literature is reviewed. Section 3 discusses the research objectives, methodology and experimental design aspects. The

2

simulation output data are exhibited and analysed in Section 4. Section 5 compares the performance of balanced and unbalanced lines. A summary and discussion of the findings, along with a set of conclusions are presented in Sections 6 and 7.

2. PREVIOUS LITERATURE

Unpaced serial production lines play a critical role in the production of many goods and services. Due to their fundamental importance in many economies, a considerable number of studies have been done so as to gain a better understanding of their operating behaviour and find ways for improving their efficiency.

2.1 Joint Imbalance of Both Service Time Means and Variability

[1] [2] examined a zero-buffered two-station line with one station having exponential service times (CV = 1) and the other having deterministic service times (CV = 0). He showed that a slight increase in TR of nearly 0.26% is achieved when the highly variable exponential station has a slightly faster processing time. Increasing the buffer size, however, quickly diminished this improvement, and for a buffer size of 2, a balanced line was found to be the best.

Later on, [3] extended the analysis to a three-station line. Six different patterns of imbalance were examined, with either one deterministic and two exponential stations, or two deterministic and one exponential station. In all situations, maximum TR was achieved when assigning lower MT to the exponential station(s). He reasoned that when both MTs and CVs are unbalanced, two opposing effects come into play; the bowl phenomenon for MT imbalance and the "variability imbalance" for CV imbalance. Which of these two effects dominates will depend on station CVs and their particular service time distributions. For example, in a 3-station line where a deterministic station is sandwiched between two exponential stations, the optimal pattern of their MTs is an inverted bowl. In this specific case the bowl phenomenon is outweighed by the variability imbalance.

[4] examined the same three–station line as Rao, with the single exception that the middle station was Erlangian. They established CV ranges for the middle station over which either the bowl phenomenon, or the variability imbalance prevails. They found the two effects to neutralise each other when the central station's CV is exactly $1/\sqrt{3}$. In another paper, [5] showed that the above guidelines are violated in lines consisting of a hyper-exponential station.

A number of researchers (see for example [6] [7] [8]) have developed approximation or optimization methods for computing the performance of various unbalanced lines (principally in terms of TR). These algorithms were applied to lines having simultaneous imbalances of both MT and variability as expressed by the CV, squared CV, or standard deviation (SD), under various operating conditions.

[9] conducted a simulation study of an unbalanced three-station line. He observed that when both MTs and SDs differ, the MT imbalance has a stronger impact than SD imbalance, so that the station having high MT is given preference in buffering over that with a high SD, unless there is a severe degree of the imbalances. This finding has since been confirmed for even longer lines by other authors, including [10] [11] [12].

A relatively recent investigation into joint MT and CV unbalanced lines was undertaken by

3

[12]. They simulated single and multiple bottleneck lines having 9 and 15 stations, lognormal service time distribution, and several MT and CV bottleneck location configurations. The results showed that a variance bottleneck station will attract more buffer units than an MT bottleneck station, only when there is an extremely high variance imbalance. Further, if both MT and variance bottlenecks are located at the same position in the line, the bottleneck will have a stronger impact on buffer allocation.

2.2 Combined Imbalance of Both Operation Time Means and Buffer Capacities

[13] found that in terms of production output and efficiency, buffers were more important in balanced lines than unbalanced lines. They also concluded that due to the bowl phenomenon, symmetrical unbalanced lines should be buffered symmetrically and that it is better to place buffers near a bottleneck station.

[14] observed that a bottleneck station draws inventory units towards it, with most units accumulating in the slot preceding the bottleneck station and the buffer space succeeding the bottleneck station having an extremely low degree of utilization. But, while this output buffer is nearly always empty, it is still critical in order to avoid blocking. As the bottleneck becomes more severe, the amount of buffer space needed to maintain a given throughput is actually reduced.

[9] investigated a three station line to find the maximum throughput for an optimal placement of buffers. They put forward the 'Alternation Rule', which states that in a line with only two buffer locations, namely B1 and B2, the optimal placement of the first buffer is into the location in between the two stations having the highest difference in mean service time.

[15] suggested that it is best to give more work to stations with higher buffer capacities and that slow stations should be placed towards both ends of the line.

[16] found that in lines with a single bottleneck station, the best buffer allocation is to use the middle stations as a starting point and then to move outward in the direction of the end stations, with a balanced buffer arrangement being the best. When two bottleneck stations are symmetrically placed, the optimal buffer pattern is still a balanced arrangement. Also, a large degree of mean operation time imbalance is needed to shift the optimal buffer assignment away from a balanced arrangement.

[17] developed a heuristic method for the optimal allocation of buffer and concluded that placing the bottleneck at the first work station results in the best output for the least amount of work in process. The worst results are obtained when locating the bottleneck at the last station.

[10] developed an algorithm to find the optimal or near optimal allocation of buffer capacity in the presence of mean operation time imbalance. [18] used a simulated annealing approach to solve the optimal buffer allocation problem for unbalanced lines.

[19] indicated that at a low degree of imbalance, an inverted bowl pattern for buffer capacity is not optimal. However, as the mean service time of the bottleneck becomes larger, more buffer space will be drawn to the line centre. At 30% imbalance, buffers start to become ineffective and so their numbers should be reduced significantly.

4

[11] showed that a nearly balanced line benefits the most from large buffers when inventory holding cost is low. As mean operation time imbalance goes up, buffers soon lose their effectiveness. They also studied a line with a single constraining station and observed that simultaneously increasing both mean processing time imbalance and buffer sizes has a positive effect on throughput, but with declining marginal benefits. They further stated that whether the buffer units are placed before or after the bottleneck is not a critical issue.

[12] studied both single and multiple bottleneck lines, combining a heuristic buffer allocation algorithm with simulation. They used a number of unbalanced mean operation time configurations under various conditions and found that the higher the severity of the bottleneck, the more buffers are pulled toward it, though the attraction rate diminishes when a certain coefficient of variation threshold is arrived at.

[20] attempted to find the optimal allocation of mean processing times and buffer capacities that result in the maximum revenue per unit of throughput for a minimum work in process holding cost. They concluded that when the cost of buffer space / work in process is high, then a balanced buffer allocation in conjunction with a bowl allocation for workload would most likely result in the best overall performance. However, if the cost of buffer space / work in process is low, then a good rule of thumb is to allocate a few percentage points more buffer to the centre of the line.

2.3 Simultaneous Imbalance of Both Variability and Buffer Sizes

One of the earliest reported studies in this area was that of [21], who simulated a three-station line with normal service times , CVs of 0.06, 0.12, 0.17, and total buffer capacities of 4, 8, 12 units. They considered six CV allocation patterns and three buffer distribution configurations, namely 25–75%, 50–50% and 75–25%. Some of the main findings of this study were:
- The lowest mean throughput (highest idle time) occurs when the biggest CV is positioned at the middle station.
- A balanced buffer allocation (50–50%) is superior in throughput to all other arrangements.
- Assigning higher BC around stations with larger CV is beneficial in terms of throughput. The greater the difference in CV among the stations, the more the gain in throughput.

[22] simulated a five-station line having four steady stations (each with CV of 0.1) and one relatively variable station (CV = 0.3). Four patterns of joint CV and BC imbalance were considered. They concluded that placing BC evenly around variable stations is the best with respect to idle time, followed by the configuration of positioning higher BC around stations with high CVs.

[13] studied six-station lines with uniform and exponential distributions. They stated that larger buffers should be allocated to stations with high service time variability for both input and output. Furthermore, they found that the total number of buffer spaces needed to reach some output target rate increases as the variability of one station is increased.

[23] examined the impact of CV on the optimal buffer assignment in lines with two-stage, Coxian-type distributions. Their results showed that in lines with higher variability, more BC should support the stations near the centre. [24] recommended that in serial lines buffers should be positioned in the vicinity of the highest variability station, rather than being close to

5

the slowest station.

[9] considered three-station lines where two or all three stations differ in standard deviation from the base level of 0.5. He reached the conclusion that to optimise buffer allocation, the first buffer should be placed closest to the high variance end station, except when the imbalance between the standard deviations is extremely big. He also demonstrated that unbalanced mean operation times have more effect on optimal buffer allocation than unequal variances.

In a follow up investigation, [16] studied the same serial line with a single bottleneck station and moderate variability. Their findings again point to the contention that differences in mean work times are more prominent than differences in variability.

[8] proposed a heuristic approximation method to compute the throughput rate of unbalanced lines and tested it for a number of cases, including one involving a six-station line with simultaneously unbalanced CVs and BCs.

[25] implemented a dynamic programming algorithm to determine the minimum-total-buffer allocation needed to achieve a desired throughput in lines with phase-type operation times. They provided numerical examples, including some concerning lines with three and five stations where both the squared CVs and BCs were imbalanced.

[19] studied lines with exponential and Erlang processing times and found that the variability of work times exerts only a small influence on the buffer allocation pattern. Generally speaking, an inverted bowl configuration for buffer assignment is optimal, but with larger total buffer space, a bowl arrangement becomes more definite.

[11] did a study of five-station lines in which one bottleneck station has either a mean service time or a variance larger than the other stations. They indicated that if the degree of imbalance is not substantial, more units of buffer are pulled toward a station with high mean processing time than another with severe CV or standard deviation.

3. RESEARCH OBJECTIVES, METHODOLGY AND EXPERIMENTAL DESIGN

The aim of this research study is to examine the characteristics and effectiveness of lines having two concurrent sources of imbalance; namely service time means, coefficients of variation and buffer sizes, taken two at a time.

The main objectives of this study are:
- To evaluate the merits of various possible joint patterns of imbalance and identify the ones with the most potential.
- To compare the efficiency of the unbalanced lines investigated to that of a corresponding balanced line and find out if improvement in performance is attainable.
- To examine the effects of line design factors (line length, buffer capacity / mean buffer capacity and degree of imbalance) on the dependent variables - idle time and average buffer level.
- To determine the most significant factors affecting the dependent performance measures.

Computer simulation was viewed as the most suitable tool for this study, since no

6

mathematical method can currently assess the more realistic serial flow lines, typically reported with positively skewed operation times.

3.1 Factorial Design

The most efficient and powerful of the many experimental designs is the complete factorial design. This design has therefore been chosen for the current investigation. In the context of the particular lines being studied, the independent variables are:

- Total number of stations in the line (N).
- Total amount of buffer capacity for the line (TB).
- Buffer capacity, BC / mean buffer capacity (MB), where MB = TB divided by the number of buffers.
- Degree of unbalanced service time means (DI).
- CV value range
- Pattern of mean work time (MT) imbalance.
- Pattern of coefficients of variation (CV) imbalance.
- Pattern of buffer capacity (BC) imbalance.

3.2 Performance Measures and Work Times Distribution

There are different ways of measuring how efficiently a line is working. The two measures used in this study were the total idle time (IT) % i.e. the fraction of time that the line is not working to total working time, and the average buffer level (ABL) for the whole line. Evidently the goal is to reduce idle time and make sure the line is functioning as smoothly as possible, and also to reduce average buffer levels, keeping the quantity of unfinished items in the buffers as low as possible.

A detailed study of published histograms of work times experienced in practice was conducted by [26]. He concluded that the work time distribution is positively skewed and follows a Weibull distribution, with a CV value varying between 0.08 and 0.50 and averaging around 0.274. This probability distribution was applied in this study

3.3 Simulation Run Parameters and Model Assumptions

The following parameter values were employed:
- Initial conditions: start the simulation run with all the buffers being nearly half-full.
- Length of the transient (non-steady state) period: discard all accumulated statistics produced during the 5,000 time units (TU) start-up period.
- Length of the simulation run and number of observations of performance measures: use a steady-state run length of 30,000 TU, divided into 12 sub-runs (blocks, or batches) of 2,500 TU each, i.e. the mean dependent variable values are recorded every 2,500 TU and then the grand mean, representing the average of these 12 mean values, is computed. This will reduce the autocorrelation to a negligible value.

The basic operating assumptions for the asynchronous flow line studied were as follows:

- The last station is never blocked and the first station is never starved.
- No breakdowns occur.
- Only one type of product flows in the system, with no changeovers and no defective

7

parts being produced.
- Time to move the work units in and out of the storage buffers is very tiny, hence ignored.

3.4 Unbalanced Lines Investigated

Three types of unbalanced lines with two sources of imbalance were studied. The designs are summarized in Table 1 below:

	Source of Imbalance	MT	CV	BC	Line Length (N)
TWO SOURCE	**a. MT + CV Investigation**	**Degree of Imbalance: 2%, 5% and 12%**	**CV = 0.08; 0.27 and 0.50**	1,2 and 6 units equal for each station	5 and 8 stations
	b. MT+ BC Investigation	**Degree of Imbalance: 2%, 5% and 12%**	CV = 0.274	**MB 2 and 6 allocated unevenly**	5 and 8 stations
	c. CV + BC Investigation	10 units/station	**CV = 0.08; 0.27 and 0.50**	**MB 2 and 6 allocated unevenly**	5 and 8 stations

Table1. Line designs studied (source of variability **in bold**)

In order to study the influence of two sources of imbalance concurrently, the two independent variables under study were varied in different combined patterns, and the third variable was kept constant in each of the three investigations. Idle times and average buffer levels were calculated for each of the patterns simulated along with the corresponding transient size. In all three investigations, the simulations were run for line lengths of N = 5 and N = 8.
The lines with two sources of imbalance are as follows:

a. Joint imbalance of both **mean service times and CVs (MT&CV investigation)**: each station has equal buffer sizes, whereas both the means and CVs vary. Buffer capacity values of 1, 2 and 6 were decided on. The degree of imbalance (DI) of the mean times was set at 2%, (very low imbalance), 5% (relatively low imbalance), and 12% (higher imbalance). CV imbalance patterns used values going from steady (CV = 0.08), through medium (CV = 0.274), to variable (CV = 0.50).

b. Combined imbalance of both **means and buffer sizes (MT&BC investigation):** CVs of 0.274 are utilized for each station. However, both the means and the buffer capacities are allowed to be unequal. As in the previous investigations, DI was 2%, 5% or 12%, and mean buffer (MB) was set at 2 and 6 units.

c. Simultaneous imbalance of both **CVs and buffer capacities (CV&BC investigation)**: all operation time means are equal to 10 time units, but the buffer sizes and CVs are unbalanced. The mean buffer is set at MB = 2 and 6 allocated unevenly between stations, and CV ranges through the three values (CV = 0.08, 0.274, 0.5), from steady to variable.

8

3.5 Patterns of Imbalance

In this section we will go into more detail with respect to the specific patterns of imbalance that are studied here.

In the studies where mean operation times are imbalanced, four patterns of imbalance are considered:
- A monotone increasing order (/).
- A monotone decreasing order (\).
- A bowl arrangement (V).
- An inverted bowl shape (^).

When coefficients of variability were imbalanced, four CV imbalance policies were simulated; these were as follows:
- Separating the variable stations from one another by steadier stations (patterns P1- P3 portray this policy).
- Assigning steadier stations to the line centre, i.e. a bowl arrangement (patterns P4 and P5 depict this policy).
- The stations with medium variability are allocated to the middle of the line. This policy represents both a decreasing order (pattern P6) and an increasing sequence (pattern P7) of CVs along the line.
- The most variable stations are assigned to the centre of the line centre - an inverted bowl arrangement (pattern P8).

These CV imbalance policies are illustrated in Table 2 below:

Pattern (Pi) of Unbalanced CVs	Line Length (N)	
	5	8
P1	MSVMS	MSVMSVMS
P2	VMSVM	VMSVMSVM
P3	SMVSM	SMVSMVSM
P4	MSSSV	MMSSSSVV
P5	MSSSV	MMSSSSVV
P6	VMMMS	VVMMMMSS
P7	SMMMV	SSMMMMVV
P8	MVVVS	MMVVVVSS

Table 2: Unbalanced CV patterns S: (CV = 0.08), M (CV = 0.27), V: (CV = 0.50)

In the case of buffer capacity imbalance, four policies were also explored for total buffer capacity allocation – these can be described as:
- Concentrating available capacity closer to the end of the line. This policy displays an increasing order of BC (pattern A).
- Concentrating buffer capacity nearer the middle of the line. This policy portrays an inverted bowl BC sequence (pattern B).
- Concentrating capacity towards the beginning of the line. This policy shows a decreasing order of BC (pattern C).
- No concentration. This policy is broken into three main sub-policies:

9

- General (pattern D1).
- Alternating BC between high and low along the line (pattern D2).
- Positioning smaller BC towards the centre - a bowl shape (pattern D3).

These policies are displayed below in Table 3 below:

Line Length (N)		5		8	
Mean Buffer Size (MB)		2	6	2	6
P1	A	1,1,1,5	3,3,3,15	1,1,1,1,6,2,2	3,3,3,3,18,6,6
P2	B	1,1,5,1	3,3,15,3	1,1,6,2,2,1,1	3,3,18,6,6,3,3
P3	C	5,1,1,1	15,3,3,3,	6,2,2,1,1,1,1	18,6,6,3,3,3,3
P4	D1	2,2,3,1	6,6,9,3	2,2,2,3,3,1,1	6,6,6,9,9,3,3
	D2	2,3,2,1	6,9,6,3	2,2,3,3,2,1,1	6,6,9,9,6,3,3
	D3	2,1,3,2	6,3,9,6	2,2,1,1,3,3,2	6,6,3,3,9,9,6
Total Buffer Capacity (TB)		8	24	14	42

Table 3. Unequal buffer size patterns (Pi = policy of buffer capacity imbalance)

4. RESULTS

For brevity only IT and ABL data for the best, second best, some good and the worst patterns will be shown.

4.1 Idle Time Results

4.1.1 IT Data

Tables 4-8 exhibit IT data for unbalanced and balanced line configurations for the MT&CV, MT&BC and CV& BC investigations:

Buffer Size		1			2			6		
% Degree of Means Imbalance)		2	5	12	2	5	12	2	5	12
Pattern of CV Imbalance	P4	8.255	8.457	9.100	4.505	7.555	7.732	2.146	2.891	7.765
	P5	**7.549**	**7.563**	**9.096**	**4.181**	**4.976**	**6.305**	**1.992**	**2.488**	**5.117**
	P8	18.451	19.201	20.352	11.833	13.286	15.436	4.790	7.075	9.718
Balanced Line			9.522			4.985			2.066	

Table 4. MT&CV investigation: IT data for MT pattern (^) & CV patterns P4, P5, P8, and the balanced line (line length = 5).

10

Buffer Size		1			2			6		
% Degree of Means Imbalance)		2	5	12	2	5	12	2	5	12
Pattern of CV Imbalance	P4	13.012	12.430	12.621	7.690	8.052	8.133	3.290	3.638	5.529
	P5	**10.911**	**10.996**	**11.263**	**7.144**	**7.236**	**7.756**	**2.523**	**2.775**	**5.113**
	P8	20.690	21.238	22.868	13.146	14.708	16.730	5.818	7.277	9.590
Balanced Line						11.522			5.935	

Table 5. MT&CV investigation: IT data for MT pattern (^) & CV patterns P4, P5, P8, and the balanced line (line length = 8)

Line Length			5						8					
Mean Buffer Capacity			2			6			2			6		
% Degree of Means Imbalance)			2	5	12	2	5	12	2	5	12	2	5	12
Pattern of Unbalanced Means and Buffer Sizes	(V) +	A	7.774	7.897	8.422	2.431	3.013	5.431	9.752	9.812	10.131	3 .234	3.304	5.437
		B	6.202	6.666	7.742	2.551	3.046	5.422	7.999	8.692	9.648	3.261	4.020	6.717
		C	6.599	7.081	7.614	1.973	3.211	5.314	9.483	9.583	9.967	2.780	2.851	6.280
		D1	**5.435**	**6.095**	**6.987**	**1.827**	**2.424**	**5 .240**	**6.275**	**6.316**	**7.714**	**2.341**	**2.808**	**5.427**
		D2	6.371	6.438	6.995	2.819	2.952	5.627	6.843	7.116	9.179	2.349	3.324	5.569
		D3	6.289	6.478	7.385	2.242	3.097	5.697	6.642	6.749	7.753	3.852	3.929	6.049
	(\) +	A	8.835	10.206	14.390	2.601	5.476	12.014	10.973	11.927	16.951	3.217	5.543	13.159
Balanced Line			4.985			2.066			5.935			2.174		

Table 6. MT&BC investigation: IT data for MT pattern (V) & BC patterns A- D3, MT pattern (\) & BC pattern A and the balanced line

Line Length			5		
Pattern of Unbalanced CVs			P4	P5	P8
Mean Buffer Capacity	Pattern of Buffer Capacity Imbalance				
	2	A	7.460	3.857	14.709
		B	7.389	7.660	10.879
		C	**3.417**	7.373	14.779
		D2	3.932	7.317	13.472
		D3	3.781	3.422	15.916
		Balanced Line	4.985		
	6	A	2.187	1.583	7.028
		B	3.723	3.888	4.659
		C	**1.544**	3.491	6.065
		D2	1.632	3.316	6.370
		D3	1.737	1.706	8.673
		Balanced Line	2.066		

Table 7. CV&BC investigation: IT data for CV patterns P4, P5, P8 & BC patterns A-C and D2-D3 and the balanced line (line length = 5)

Line Length			8		
Pattern of Unbalanced CVs			P4	P5	P8
Mean Buffer Capacity	Pattern of Buffer Capacity Imbalance				
	2	A	13.642	9.155	18.168
		B	13.539	12.779	12.839
		C	**6.487**	12.964	18.124
		D2	7.250	12.267	14.648
		D3	7.866	10.805	18.552
		Balanced Line	5.935		
	6	A	6.023	3.574	9.147
		B	6.725	6.635	6.048
		C	**2.299**	5.438	8.958
		D2	2.537	4.458	6.609
		D3	2.472	5.481	9.259
		Balanced Line	2.174		

Table 8. CV&BC investigation**:** IT data for CV patterns P4, P5, P8 & BC patterns A-C, D2-D3 and the balanced line (line length = 8)

11

4.1.2 Ranking of Policies and Patterns

From Tables 4-8, it is interesting to note the following:

MT & CV investigation:
- No single CV policy can be classified as the most or least favourable in terms of all of its patterns, but some specific individual configurations may be viewed as the best or the worst.
- The best unbalanced MT pattern is an inverted bowl arrangement (^), combined with CV pattern P5 (bowl shaped), i.e. the MT bottleneck station is located in the middle of the line and the two CV bottleneck stations are positioned at the beginning and end of the line.
- There seems to be no clear second best configuration. The lines having MT pattern (\) coupled with CV pattern P5 and MT pattern (^) along with CV pattern P4 have relatively low IT levels.
- The worst line configuration is any MT pattern together with CV pattern P8.

MT & BC investigation:
- While individual patterns may be regarded as either best or worst, the results are not conclusive enough for any particular policy to be classified as such with regard to all of its patterns.
- The best unbalanced pattern is (V) & D1, i.e. an MT bowl configuration, coupled with a distribution of buffer capacity as evenly as possible.
- Other good patterns are (V) & A, B, C, D2, D3, representing an MT bowl arrangement in conjunction with any buffer policy (except for the best one; D1).
- Amongst the worst patterns are (\) & A, (/) & C, and (^) & C.

   CV&BC investigation:
- Broadly speaking, no overall best or worst policy can be identified.
- The most superior pattern can be considered as P4 & C, i.e., the combination of CV pattern P4 (bowl-shaped) and buffer arrangement C (a descending order).
- CV&BC configurations P5 & D3 (line length = 5 stations) and P4 & D2 (line length = 8 stations) can be viewed as the second best patterns
- P4 & D3, P3 & B, and P5 & A are also good arrangements.
- The worst pattern is P8 & D3.
- Combined with any of the BC patterns, both CV bowl patterns P4 and P5 consistently produce good IT results.

4.1.3 Effect of Independent Factors on IT

The simulation data exhibit the following relationships between the independent variables and idle time:

MT & CV investigation:
- Increasing N leads to an increase in IT, particularly for smaller BC values in the case of the best pattern.
- As BC goes up IT decreases, with the rate of decline for the best configuration slowing down as BC and N increase.
- When DI becomes higher, IT also increases.

12

MT&BC investigation:

- IT goes up with an increase in N, particularly for lower MB values.
- IT decreases as MB is increased. For the best pattern, the marginal decrease in IT increases as MB and N continue to increase.
- As DI is increased, so does IT. The marginal increase in IT as DI rises tends to become more substantial as MB increases.

CV&BC investigation:

- IT increases with line length, particularly for low MB.
- As MB is increased, IT will decline.

4.1.4 ANOVA

ANOVA results for the three investigations indicate that all of the main effects are highly significant at the 0.01 level and that all of the first order interactions are significant at the 0.05 level or above. The batch effect is insignificant, consolidating the belief that the steady state condition is attained. The variables affecting IT can be ranked as follows:

MT & CV investigation: the most influential factor on IT is BC, followed by DI. CV pattern and MT pattern are respectively, 3rd and 4th in order of importance.

MT&BC investigation: the main independent variable affecting IT is MB, followed respectively by DI, MT pattern and the buffer pattern.

CV&BC investigation: the most significant factor on IT is MB, followed by CV pattern. BC pattern is on the other hand the least influential factor.

4.2 Average Buffer Level Results

4.2.1 ABL Data

Tables 9-15 exhibit ABL data for various unbalanced and balanced line configurations for the three two imbalance source investigations:

Buffer Size		1			2			6		
% Degree of Means Imbalance)		2	5	12	2	5	12	2	5	12
Pattern of Unbalanced CVs	P4	**0.289**	**0.204**	**0.133**	**0.421**	**0.333**	**0.224**	**0.933**	**0.553**	**0.332**
	P5	0.626	0.441	0.181	0.985	0.476	0.240	1.971	0.559	0.349
Balanced Line		0.526			1.033			3.321		

Table 9. MT&CV investigation: ABL data for MT pattern (\) & CV patterns P4, P5 and the balanced line (line length = 5)

13

Buffer Size		1			2			6	
% Degree of Means Imbalance)		2	5	12	2	5	12	2	5	12
Pattern of Unbalanced CVs	P4	**0.257**	**0.196**	**0.135**	**0.389**	**0.301**	**0.216**	**0.921**	**0.698**	**0.392**
	P5	0.724	0.574	0.239	1.348	0.599	0.242	2.115	0.858	0.400
Balanced Line		0.559			0.970			2.601		

Table 10. MT&CV investigation: ABL data for MT pattern (\) & CV patterns P4, P5 and the balanced line (line length = 8)

Buffer Size		1			2			6	
% Degree of Means Imbalance)		2	5	12	2	5	12	2	5	12
Pattern of Unbalanced CVs	P5	0.754	0.828	0.879	1.551	1.699	1.795	5.252	5.432	5.607
Balanced Line		0.526			1.033			3.321		

Table 11. MT&CV investigation: ABL data for MT pattern (/) & CV pattern P5 and the balanced line (line length = 5)

Buffer Size		1			2			6	
% Degree of Means Imbalance)		2	5	12	2	5	12	2	5	12
Pattern of Unbalanced CVs	P5	0.813	0.846	0.887	1.626	1.711	1.775	5.202	5.199	5.609
Balanced Line		0.559			0.970			2.601		

Table 12. MT&CV investigation: ABL data for MT pattern (/) & CV pattern P5 and the balanced line (line length = 8)

14

Line Length			5						8					
Mean Buffer Capacity			2			6			2			6		
% Degree of Means Imbalance)			2	5	12	2	5	12	2	5	12	2	5	12
Pattern of Unbalanced Means and Buffer Sizes	(/) +	B	1.380	1.540	1.804	4.460	5.097	5.681	1.419	1 .684	1.803	5.112	5.298	5.707
		C	1.550	1.688	1.824	4.549	5.105	5.610	1.622	1.702	1.796	4.163	5.002	5.560
		D1	1.432	1.599	1.819	4.298	5.127	5.595	1.251	1.560	1.782	3.984	4.778	5.649
		D2	1.482	1.627	1.812	3.197	4.930	5.613	1.344	1.508	1.776	4.194	5.027	5.447
	(\) +	A	**0.467**	**0.353**	**0.216**	**1.251**	**0.753**	**0.346**	**0.466**	**0.385**	**0.246**	**1.629**	**0.942**	**0.401**
		B	0.690	0.438	0.222	1.458	0.777	0.376	0.689	0.438	0.248	2.044	1.154	0.520
		D1	1.109	0.611	0.304	1.991	0.756	0.439	0.808	0.583	0.350	1.637	1.154	0.541
		D2	0.988	0.741	0.382	1.257	0.820	0.435	1.063	0.904	0.555	1.926	0.951	0.640
Balanced Line			1.033			3.321			0.970			2.601		

Table 13. MT&BC investigation**:** ABL data for MT pattern (/) & BC patterns B-D2, MT pattern (\) & BC patterns A-B, D1-D2, and the balanced line

Line Length			5	
Pattern of Unbalanced CVs			P4	P5
Mean Buffer Capacity	2	Pattern of Buffer Capacity Imbalance	A	**0.317**	0.759
			B	0.417	1.649
			C	0.899	1.627
			Balanced	1.033	
	6		A	**0.733**	2.126
			B	0.741	5.184
			C	3.472	5.228
			Balanced	3.321	

(Table 14 — columns: P4, P5; values: A 0.317/0.759, B 0.417/1.649, C 0.899/1.627, Balanced 1.033; A 0.733/2.126, B 0.741/5.184, C 3.472/5.228, Balanced 3.321)

Table 14. CV&BC investigation: ABL data for CV patterns P4, P5 & BC patterns A-C and the balanced line (line length =5)

Line Length			8	
Pattern of Unbalanced CVs			P4	P6
Mean Buffer Capacity	2	Pattern of Buffer Capacity Imbalance	A	**0.259**	0.291
			B	0.303	0.345
			C	1.333	1.146
			Balanced	0.970	
	6		A	**0.631**	0.822
			B	0.888	0.896
			C	2.010	1.634
			Balanced	2.601	

Table 15. CV&BC investigation: ABL data for CV patterns P4, P6 & BC patterns A-C and the balanced line (line length =8)

4.2.2 Ranking of Policies and Patterns

From Tables 9-15, it is interesting to note the following:

MT&CV investigation:
- None of the CV policies considered can be described as being the best or worst, as each consists of a number of patterns with different performance. Specific patterns however, may be ranked as the most or least performing.
- The best pattern is MT (\) coupled with CV pattern P4, i.e. the MT bottleneck station is positioned at the beginning of the line, while positioning the CV bottleneck stations at both ends.
- The second best configuration is MT (\) along with CV pattern P6.
- Other patterns providing reasonably good performance are MT (\) combined with CV

15

patterns P2 and P5.
- Lines having MT patterns (\) are better than all other configurations, while those with MT patterns (/) are consistently the worst.

MT&BC investigation:
- Each BC policy is comprised of a number of different patterns with varying performance; therefore none of the policies can be ranked as the best or worst. However, a specific individual pattern within each policy stands out as the most superior or inferior.
- The best pattern is (\) & A, i.e. a monotone decreasing MT order, together with an ascending BC allocation.
- The second best configuration is generally (\) & B (an MT descending arrangement, along with an inverted bowl buffer size pattern).
- Other good MT&BC patterns are (\) & D1 and (\) & D2.
- The worst patterns are generally the MT (/) patterns combined with any buffer capacity allocation.
- In general, the (\) patterns are better than any other MT patterns considered.

CV&BC investigation:
- Again, no single policy can be said to be the best or worst, but some individual patterns may be viewed as being so.
- The best CV&BC pattern is P4 & A, i.e., the combination of CV pattern P4 (bowl-shaped) and BC arrangement A (concentrating the buffers towards the end of the line).
- The second best configuration is P4 & B (N = 5) and P6 & A (N = 8).
- The least favourable patterns are P5 & C and P5 & B for N = 5. No clear-cut worst configuration was discerned for N = 8.

4.2.3 Effect of Design Factors on ABL

With regards to the influence of the independent variables on average buffer level, the simulation results point out to the following relationships:

MT&CV investigation:
- ABL goes up with an increase in BC. For the best configuration the rate of increase in ABL as BC increases, slows down as N is reduced.
- As DI is increased ABL falls. In the case of the best pattern, this drop in ABL becomes less pronounced as DI continues to go up and as BC declines.

MT&BC investigation:
- ABL tends to rise as N and/or MB increase.
- ABL decreases when DI rises. The decline in ABL in the case of the best pattern becomes less dramatic as DI continues to go up, particularly at lower MB levels.

CV&BC investigation:
- ABL becomes higher when MB is increased. This increase in ABL continues at a diminishing rate as MB continues to increase.

16

4.2.4 ANOVA

The analysis of variance of the ABL data for the MT&CV, MT&BC and CV & BC investigations resulted in the same conclusions as those presented in section 4.1.4 regarding IT results. The relative rankings of the design factors affecting ABL are as follows:

MT&CV investigation: the most prominent factor affecting ABL is BC. The second, third, and fourth factors are respectively, DI, MT pattern and CV pattern.

MT&BC investigation: the most significant independent variable in terms of its impact on ABL is MT pattern, followed respectively by MB, DI and buffer pattern.

CV&BC investigation: in terms of their relative importance, the factors affecting ABL are respectively, MB, CV pattern, and BC pattern.

5. BEST UNBALANCED PATTERNS' SAVINGS OVER THE BALANCED LINE

In the MT& CV investigation, the most favourable unbalanced pattern in terms of IT was an arrangement whereby the MT bottleneck station is located in the middle of the line and the two CV bottleneck stations are positioned at the beginning and end of the line. Table 16 exhibits the percentage differences in this pattern's IT vis-a-vis the balanced line (the control):

Line Length = 5			Line Length = 8		
Buffer Size	% Degree of Means Imbalance)	% Difference	Buffer Size	% Degree of Means Imbalance)	% Difference
1	2	**-20.73**	1	2	-5.31
1	5	-20.58	1	5	-4.56
1	12	-4.48	1	12	-2.25
2	2	-16.12	2	2	20.37
2	5	-0.18	2	5	21.92
2	12	26.48	2	12	30.68
6	2	-3.58	6	2	16.05
6	5	20.43	6	5	27.65
6	12	147.68	6	12	135.19

(-) indicates saving

Table 16. MT&CV investigation: % difference in the best pattern's IT over the balanced line

From Table 16, the following can be observed:

- The biggest obtained saving in IT for the most favourable configuration over the balanced line is -20.73%.
- As DI increases, any improvement in IT either decreases or is immediately wiped out, especially for higher BC values.
- When BC is increased the savings disappear either immediately or gradually.

17

- As N rises, such savings become smaller.

On the other hand, under the best MT&CV pattern with regard to ABL, the MT bottleneck station is positioned at the beginning of the line, while placing the CV bottleneck stations at both ends.

Table 17 exhibits the percentage savings in the best configuration's ABL over the balanced line.

The following can be discerned from Table 17:

- The highest saving in ABL for the best pattern over the balanced line is 90.06%.
- The best pattern outperformed the balanced line for all line length, buffer size and imbalance degree levels.
- When DI goes up, the advantage in ABL of the unbalanced line over the control increases.
- Increasing BC raises the best pattern's advantage.
- For all the factor levels considered, the best pattern outperformed the balanced line in terms of ABL.

Line Length = 5			Line Length = 8		
Buffer Size	% Degree of Means Imbalance)	% Saving	Buffer Size	% Degree of Means Imbalance)	% Saving
1	2	-45.06	1	2	-54.12
1	5	-61.22	1	5	-64.94
1	12	-74.72	1	12	-75.85
2	2	-59.25	2	2	-59.95
2	5	-67.76	2	5	-68.97
2	12	-78.36	2	12	-77.78
6	2	-71.91	6	2	-64.36
6	5	-83.35	6	5	-73.16
6	12	**-90.06**	6	12	-84.95

(-) indicates saving

Table 17. MT&CV investigation: % saving in the best pattern's ABL over the balanced line

For the MT&BC investigation, the best pattern for reducing idle times came from an MT bowl configuration, coupled with a distribution of buffer capacity as evenly as possible. On the other hand, the most advantageous pattern for lowering average buffer levels turned out to be a monotone decreasing MT order, together with an ascending BC allocation. A summary of the % differences in IT and ABL are shown for the best patterns in Tables 18 and 19.

18

Line Length = 5			Line Length = 8		
Mean Buffer Size	% Degree of Means Imbalance	% Difference	Mean Buffer Size	% Degree of Means Imbalance	% Difference
2	2	9.03	2	2	5.73
2	5	22.27	2	5	6.42
2	12	40.16	2	12	30.43
6	2	**-11.57**	6	2	7.68
6	5	17.33	6	5	29.16
6	5	153.63	6	12	149.63

(-) indicates saving

Table 18. MT&BC investigation: % difference in the best pattern's IT over the balanced line

Line Length = 5			Line Length = 8		
Mean Buffer Size	% Degree of Means Imbalance	% Saving	Mean Buffer Size	% Degree of Means Imbalance	% Saving
2	2	-54.79	2	2	-51.96
2	5	-65.83	2	5	-60.31
2	12	-79.14	2	12	-74.64
6	2	-51.22	6	2	-37.37
6	5	-77.34	6	5	-63.80
6	12	**-89.58**	6	12	-84.58

(-) indicates saving

Table 19. MT&BC investigation: % savings in the best pattern's ABL over the balanced line

From Tables 18 and 19, the following can be seen:

- Whilst the highest saving in the best unbalanced pattern's IT over the balanced line is 11.57%, the biggest saving in ABL (89.58%) is over 7.5 times larger.
- As DI increase, the ABL advantage increases, whereas any saving in IT disappears.
- When N increases, the % saving in ABL falls.
- The best ABL pattern consistently shows improvements over the balanced line for all the N, MB and DI values considered.

As regards CV&BC investigation, it was found that the most superior pattern with respect to IT was a combination of a bowl-shaped CV pattern and a descending order buffer arrangement. It was also observed that the best configuration in terms of ABL was a combination of a bowl-shaped CV pattern and a BC arrangement stipulating the concentration of the buffers towards the end of the line. Tables 20 and 21 summarise the % differences in IT and ABL for the best unbalanced patterns in comparison with those of the balanced line.

19

Line Length	5		8	
Mean Buffer Size	2	6	2	6
% Difference	**-31.46**	-25.27	9.30	5.75

(-) indicates saving

Table 20. CV&BC investigation: % difference in the best pattern's IT over the balanced line

Line Length	5		8	
Mean Buffer Size	2	6	2	6
% Saving	-69.31	**-77.93**	-73.35	-75.76

(-) indicates saving

Table 21. CV&BC investigation: % savings in the best pattern's ABL over the control

As is exhibited in Tables 20 and 21, the following can be observed:

- The highest IT and ABL savings for the best patterns over the balanced line are respectively around 31% and 78%.
- As N increases, the saving in IT disappears.
- When MB level becomes higher, the advantage in ABL increases.
- The best pattern consistently shows substantially lower ABL levels over the balanced line for all factor levels considered.

6. SUMMARY

A number of combined MT, CV and BC unbalancing policies and configurations were analysed in this study. One of the main conclusions is that there is no overall best policy that suits all types of imbalance but there were particular patterns within each policy that showed improvements of performance either in idle time or in average buffer level when compared to the balanced line counterpart.

Table 22 below summarizes the best performing configurations in terms of IT and ABL for the three investigations:

Best Performance Configurations	Idle Time	Average Buffer Level
MT + CV	MT Inverted bowl + CV bowl	MT Descending order + CV bowl pattern
MT + BC	MT Bowl + BC close to balance	MT Descending order + BC ascending order)
CV + BC	CV Bowl pattern + BC descending order	CV Bowl + BC ascending order

Table 22. The influence of double source imbalance patterns on idle time and average buffer levels.

It was observed that as BC increases, IT declines but ABL goes up. In contrast, as DI rises, IT increases and ABL falls. This demonstrates that both BC and DI seem to influence IT and ABL in opposing directions. Also, when either DI or BC is increased, the superiority of the unbalanced line over the control decreases for IT, but becomes more pronounced for ABL. In

20

addition, the best pattern with respect to ABL consistently showed improvements over the balanced line for all the N, BC and DI levels considered.

As regards the influence of the various factors on IT and ABL, Table 23 below summarizes ANOVA findings:

Investigation	Performance Measure	Factor 1	Factor 2	Factor 3	Factor 4
MT&CV	IT	BC	DI	CV pattern	MT pattern
MT&BC	IT	MB	DI	MT pattern	BC pattern
CV&BC	IT	MB	CV pattern	BC pattern	---
MT&CV	ABL	BC	DI	MT pattern	CV pattern
MT&BC	ABL	MT pattern	MB	DI	BC pattern
CV&BC	ABL	MB	CV pattern	BC pattern	---

Table 23. Ranking of the independent variables in their effects on IT and ABL

Furthermore, the best patterns found in the three investigations have all resulted in various degrees of savings over the balanced line, as is shown in Table 24 below:

Investigation	% Saving in IT	% Saving in ABL
MT&CV	20.73	90.06
MT&BC	11.57	89.58
CV&BC	31.46	77.93

(-) indicates saving

Table 24. Highest obtained % savings in IT and ABL

7. DISCUSION AND CONCLUSIONS

This research has set out to learn what impact two sources of imbalance might have on the performance of flow lines. It has shown that none of the patterns considered simultaneously achieved both low IT and ABL levels. Whether a line manager considers a low IT as more beneficial than a reduced ABL will partly depend on the relative costs of inventory (buffer space, work in process and stock holding costs) and lost production.

A line manager is going to have to opt for a line where the greatest advantages can be had. It may be to reduce idle time, should it be costly, for instance in an industry where demand is high and operators are working full out, such as on the assembly or production lines in consumer goods (toys, shoes, office supplies), or where manpower is expensive. In these cases, where any idle time leads to great expense, the best or other favourable jointly unbalanced designs may be selected to get the largest possible idle time reduction.

It may be, however, that the principal aim is lean buffering, as in the automotive industry, where just-in-time management requires it. Here, the best or some other advantageous unbalanced patterns which bring average buffer levels down would be the most appropriate.

It has been demonstrated that significant improvement in the performance of a non-automated flow line can be achieved by unbalancing it in an appropriate manner. Savings as high as 31% in IT (a substantial amount), and 90.06% in ABL (a very significant percentage) over the balanced line were obtained (see Table 23 above).

21

Savings of such magnitudes would appear to justify unbalancing production lines in many cases where operators differ in their mean processing times, variability, or when space and other restrictions require uneven buffer capacity allocation, especially given that these savings can be achieved while allocating workers to the same stations and without any need for extra investment in capital, labour, space, or other resources. The scale of the potential reductions in idle time and ABL when calculated over the lifecycle of a production line means that purposely unbalancing the buffer sizes and operators with different speeds or variability could lead to real benefits for the manufacturer and so might be a strategy to take into account when designing the production line.

It should be remembered, however, that the patterns considered are specific patterns among numerous possibilities, and that imbalance directed in the wrong way could lead to the opposite effect, i.e. increases in average buffer levels and/or idle times.

It is hoped that this research has contributed to the body of knowledge of unbalanced lines in furnishing additional insights into how to fine tune imbalanced flow lines with the objective of enhancing performance.

Ample opportunities for further research into this field of production lines are still available. For example, investigating the effects of having double sources of imbalance on merging (assembly) lines, or unreliable lines will expand knowledge in this field and furnish managers with a more enhanced line design and operation principles.

REFERENCES

[1] Rao, N.P. *On the Mean Production Rate of a Two-Stage Production System of the Tandem Type.* International Journal of Production Research, 1975a, Vol.13, No.2, pp. 207-217.

[2] Rao, N.P. *Two-Stage Production Systems with Intermediate Storage.* International Journal of Production Research, 1975b, Vol.13, No.4, pp. 414-421.

[3] Rao, N.P. *A Generalisation of the Bowl Phenomenon in Series Production Systems.* International Journal of Production Research, 1976, Vol.14, No. 4, pp. 437-443.

[4] Mishra, R., Sahu, K.C. & Rao, N.P. *Variability Imbalance in Series Production Systems.* International Conference in Production Engineering, 1977, New Delhi, India.

[5] Mishra, A., Acharya, D., Rao, N. P. & Sastry, G. P. *Composite Stage Effects in Unbalancing of Series Production Systems.* International Journal of Production Research, 1985, Vol. 23, No. 1, pp. 1–20.

[6] Muth, E.J. & Alkaff, A. *The Bowl Phenomenon Revisited.* International Journal of Production Research, 1987, Vol. 25, No. 2, pp. 161-173.

[7] Chow, W.M. *Buffer Capacity Analysis for Sequential Production Lines with Variable Process Times.* International Journal of Production Research, 1987, Vol. 25, No. 8, pp. 1183-1196.

[8] Liu, C.M. & Lin, C.L. *Performance Evaluation of Unbalanced Serial Production Lines.* International Journal of Production Research, 1994, Vol. 32, No. 12, pp. 2897-2914.

22

[9] Powell, S.G. *Buffer Allocation in Unbalanced Three - Station Serial Lines.* International Journal of Production Research, 1994, Vol. 32, No. 9, pp. 2201-2217.

[10] Harris, J.H. & Powell, S.G. *An Algorithm for Optimal Buffer Placement in Reliable Serial Lines.* IIE Transactions, 1999, Vol. 31, No.4, pp. 287-302.

[11] Chiadamrong, N. & Limpasontipong, P. *Using Storage Buffer to Improve Unbalanced Asynchronous Production Flow Line's Performance.* International Journal of Manufacturing Technology and Management, 2003, Vol. 5, Nos. 1/2, pp. 149-161.

[12] Sabuncuoglu, I., Erel, E. & Gocgun, Y. *Analysis of Serial Production Lines: Characterisation Study and a New Heuristic Procedure for Optimal Buffer Allocation.* International Journal of Production Research, 2006, Vol.44, No. 13, pp. 2499-2523.

[13] Conway, R., Maxwell, W., McClain, J. & Thomas, L. J. *The Role of Work in Progress Inventory in Serial Production Lines.* 1988, Operations Research, Vol. 36, No. 2, pp. 229 - 241.

[14] Lambrecht, M. & Segaert, A. *Buffer Stock Allocation in Serial and Assembly Type of Production Lines.* International Journal of Operations and Production Management, 1990, Vol.10, No. 2, pp. 47-61.

[15] Hillier, F.S. & So, K.C. *On the Optimal Design of Tandem Queuing Systems with Finite Buffers.* Queuing Systems, 1995, Vol. 21, pp. 245-266.

[16] Powell, S. G. & Pyke, D. F. *Allocation of Buffers to Serial Production Lines with Bottlenecks.* IIE Transactions, 1996, Vol. 28, No. 1, January, pp. 18-29.

[17] So, K. C. *Optimal Buffer Allocation Strategy for Minimizing Work-in-Progress Inventory in Un-paced Production Lines.* IIE Transactions, 1997, Vol. 29, pp. 8 −88.

[18] Spinellis, D. D. & Papadopoulos, C. T. *A Simulated Annealing Approach for Buffer Allocation in Reliable Production Lines.* Annals of Operations Research, 2000, Vol. 92, pp. 373-384.

[19] Hillier, M.S. *Characterizing the Optimal Allocation of Storage Space in Production Line Systems with Variable Processing Times.* IIE Transactions, 2000, Vol. 32, No.1, pp. 1–8.

[20] Hillier, M. S. & Hillier, F. S. *Simultaneous Optimization of Work and Buffer Space in Unpaced Production Lines with Variable Processing Times.* IIE Transactions, 2006, Vol. 38, pp. 39-51.

[21] Smith, L.D. & Brumbaugh, P. *Allocating Inter-Station Inventory Capacity in Unpaced Production Lines with Heteroscedastic Processing Times.* International Journal of Production Research, 1977, Vol. 15, No. 2, pp. 163–172.

[22] De La Wyche, P. & Wild, R. *The Design of Imbalanced Series Queue Flow Lines.* Operational Research Quarterly, 1997, Vol.28, No.3, ii, pp. 695-702.

[23] Hillier, F.S. & So, K.C. *The Effect of the Coefficient of Variation of Operation Times on the Allocation of Storage Space in Production Line Systems.* IIE Transactions, 1991, Vol. 23, No. 2, pp. 198–206.

[24] Dallery, Y. & Gershwin, S.B. *Manufacturing Flow Line Systems: A Review of Models and Analytical Results.* Queuing Systems, 1992, Vol. 12, pp. 3–94.

[25] Yamashita, H. & Altiok, T. *Buffer Capacity Allocation for a Desired Throughput in Production Lines*. IIE Transactions, 1998, Vol. 30, pp. 883–891.

[26] Slack, N. *Work Time Distributions in Production System Modelling.* Research Paper, Oxford Centre for Management Studies, 1982.

24

The Importance of Selecting the Capacity Disruption Level Used in Network Disruption Modeling

Network-disruption is a methodological approach that has been used to model various types of disruptive events on transportation networks and can used to quantify the importance of the disrupted link relative to all other links in the network. When modeling network disruptions it is critical to consider the underlying assumptions related to the disruptive event itself –for example the severity and duration of the disruption – and the traffic flow regimen. The traffic flows throughout the network can vary substantially depending on whether a disruption is modeled as a partial reduction in the capacity where some capacity on the disrupted link remains open or as a complete reduction in the capacity where the disrupted link is completely disabled or impassible. These assumptions impact how the traffic re-routes as a result of the disruption and the traffic flows on alternative routes. The *capacity-disruption level* is the reduction in the capacity on a given link due to some type of disruption expressed as a percentage. For example, a complete capacity reduction is synonymous with 100% capacity-disruption level that reduces the capacity on the link to zero.

We argue that it is important to carefully consider the nature of the disruptive event being modeled, the assumptions related to the traffic flow regime and duration of the disruption, and the network-disruption methodology that is used, as different approaches and assumptions can produce different performance results. We measure performance as the increase in travel time resulting from a disruption on a network link compared to the base case scenario of a fully or "normally" functioning network with no disruption. We show that the use of different capacity-disruption levels results in substantially different network-wide travel time outcomes and that the relationships between the capacity disruption level, network connectivity, and performance are non-linear and non- intuitive. These results are highly relevant with respect to considering how specific network disruptions are modeled and the capacity disruption value that is used. The common use of complete link removal is rather arbitrary and is not necessarily a defensible methodology, nor is it necessarily realistic with respect to modeling the impacts of many types of non-catastrophic disruptions that occur most frequently on transportation networks.

ON THE ANALYSIS OF COST-EFFECTIVE MILITARY FORCE STRUCTURES

Jean-Pierre Amor, School of Business Administration, University of San Diego, San Diego, CA 92110, jpamor@sandiego.edu, 1-619-260-2377

ABSTRACT

The problem of identifying cost-effective force structures is very important to military leadership. Ideally, one would like to develop optimal, time-phased mixes of weapon systems, over a given horizon, using widely accepted cost and effectiveness measures. In this paper, we briefly review several approaches that have been used in the past and outline some of their limitations. We then propose a quasi-dynamic optimization methodology for generating cost-effective military force structures. This conceptual work is based on the author's past experience with military force structures analysis and assessment.

Keywords: Military force structures, Combat simulation, Optimization

INTRODUCTION

An important problem facing the leadership of the military is the early identification of cost-effective force structures -- strategic and tactical for the Army, Navy, and Air Force. For the long run, say ten to fifteen years ahead, the generic force structure problem focuses on determining optimal mixes of weapon systems for each year of a given planning horizon. In this paper, a time-phased mix of weapon systems (i.e., a phase-in and phase-out schedule for various systems) will be considered optimal if it maximizes the effectiveness of the force assigned to a given theater/scenario of combat operations, while honoring various cost and possibly other types of resource constraints, such as production rates, total force size, etc.

A high-level mathematical statement of the problem has the form:

$$\text{"Max" } E(y) \quad \text{subject to } C(y) \leq k \tag{1}$$

Where:

- y is a vector of decision variables representing a potential force structure,
- E is a vector function representing the effectiveness of the force assigned to a given theater/scenario of combat operations.
- C is a vector function providing the cost or resource consumption of the force, and
- k is a vector of parameters representing budgetary or other resource constraints.

The quotation marks in (1) reflect the fact that it is generally impossible to optimize several, usually conflicting, objectives simultaneously. Note that this macro-formulation of the problem does not introduce symbols to deal explicitly with the time factor, or the various effectiveness

measures and cost/resource categories. For example, a component of the vector cost function could be the operating cost of a particular system which will be ten years old in 2015. Such detailed specifications are assumed to be understood for now so that we may quickly classify several approaches that have been tried in the past and outline their limitations. The purpose of this paper is to propose a more desirable methodology for generating cost-effective military force structures.

A BRIEF CLASSIFICATION OF SOME PREVIOUS APPROACHES

Approach 1

In the assessment of force modernization programs, Approach 1 selects reasonable ranges for the components of k and, for certain values in those ranges, it heuristically develops feasible force structures, which reduce the average age of the force while honoring various constraints on its size. The age of the force is viewed as a surrogate for its effectiveness -- the younger the force, the more effective it is assumed to be. No attempt is made at finding an optimal solution.

Approach 2

Approach 2 compares the currently programmed force structure to a few (politically?) proposed alternatives to this force structure for several values of k. Typically, the alternatives change the time-phasing and/or the mixes of the various types of weapon systems, but they occasionally violate some of the components of k. The age and size of the overall force are viewed as indices of its effectiveness, and no attempt is made at finding an optimal solution.

Approach 3

Approach 3 explores feasible, alternative structures for several values of k. Additional constraints on the minimum force size and system production rates are enforced. Sensitivity analyses are conducted with respect to the components of the vector function C. Once again, the size of the force is viewed as a substitute for its effectiveness, and no attempt is made at finding an optimal solution.

Approach 4

Approach 4 develops alternative mixes of systems which can be procured and operated over several years for a given cost, and then compares their effectiveness using combat simulation models. In these simulations, the effectiveness comparisons are not responsive to the time evolution of the size of the force and of the mix of systems, which occur during the multi-year period for which the costs were developed. This limitation essentially removes the time dimension in the definitions of y, C, and k in (1). Also, the force structure problem is artificially decomposed into independent sub-problems and no attempt is made at finding an optimal solution. A brief classification of the four approaches outlined above is presented below.

	Approach 1	Approach 2	Approach 3	Approach 4
Time-Phased Mixes of Weapon Systems	YES	YES	YES	NO
Effectiveness Measures or Surrogates (E)	Age	Age Size	Size	Combat Exchange Ratios
Compatible Time Frames for Relating Cost (C) and Effectiveness (E)	YES	YES	YES	NO
Optimization Methodology	NO	NO	NO	NO

Table 1: Classification of Some Previous Approaches.

A PROPOSED FORCE STRUCTURE OPTIMIZATION METHODOLOGY

There exist optimization models which consider (implicitly) a multitude of alternatives in their search for an optimal solution to (1), i.e., identifying the most effective phase-in and phase-out schedules for weapons systems, subject to yearly cost constraints. However, some are quite limited in their representation of combat effectiveness (e.g., they assume that force effectiveness is separable in the various combat tasks, or that task achievement is separable in the various weapon systems), or they ignore the learning aspects of production processes, or they leave out certain costs of combat operations. An optimization methodology which addresses these limitations is proposed below.

Force Structure Analysis Model

1. Indices
   - j indexes systems; $J \equiv$ Number of systems considered.
   - l, m & b index time periods; $L \equiv$ Number of periods considered.
   - $(l\text{-}b)$ denotes the age of a system.
   - n indexes system assignments; $N \equiv$ Number of assignments considered; n=1 represents assignment to the combat theater/scenario of interest.

2. Decision Variables
   - $x_j \equiv$ Number of systems of type j purchased over the entire horizon.
   - $x_{jl} \equiv$ Number of systems of type j purchased at the beginning of period l.
   - $x_{jlm} \equiv$ Number of systems of type j purchased at the beginning of period l and retired at the end of period m.
   - $w_{jlm} \equiv$ Number of systems of type j inherited from period l and retired at the end of period m.

- z_{jln} ≡ Number of systems of type j available in period l for assignment n.

3. Functions
   - Ψ ≡ Unit step function (=0 if x_j =0; =1 if x_j >0). It is used for calculating R&D costs.
   - Θ_{jl} ≡ Procurement cost function for system j in period l.
   - γ_{jln} ≡ Operating cost function for system j in period l while in assignment n.
   - ε_l ≡ Effectiveness function of the systems available for assignment 1 during period l. This function is developed from the results of a combat simulation.

4. Parameters

   a. Technological
      - R_{jl} ≡ Cost of developing system j in period l.
      - A_{jln} ≡ Fraction of systems j in assignment n during period l.
      - V_{jbn} ≡ Fraction of systems j in assignment n still usable after b periods.

   b. Objective Function
      - Parameters characterizing the effectiveness of individual systems, the scenario, etc., are input to the combat simulation from which ε_l is derived.

   c. Constraints
      - R_l ≡ Limit on R&D expenditures in period l.
      - P_l ≡ Limit on procurement expenditures in period l.
      - E_l ≡ Limit on operating expenditures in period l.

5. The Model

$$\text{Max Min } \{ \varepsilon_l (z_{1l1}, z_{2l1}, \ldots, z_{Jl1}) \} \qquad (2)$$
$$1 \le l \le L$$

   Subject to:

   a. Cost related constraints

$$\text{R\&D: } \sum_{j=1}^{J} R_{jl} \Psi(x_j) \le R_l \qquad l = 1,\ldots, L \qquad (3)$$

$$\text{Procurement: } \sum_{j=1}^{J} \Theta_{jl} (x_{jl}) \le P_l \qquad l = 1,\ldots, L \qquad (4)$$

$$\text{Operating: } \sum_{j=1}^{J} \sum_{n=1}^{N} \gamma_{jln} (z_{jln}) \le E_l \qquad l = 1,\ldots, L \qquad (5)$$

b. Force distribution constraints

$$z_{jln} = A_{jln} \left\{ \sum_{b=L_j}^{l} \sum_{m=l}^{L} V_{j(l-b)n}(x_{jbm} + w_{jbm}) \right\} \qquad (6)$$

$$j = 1,\ldots, J \qquad l = 1,\ldots, L \qquad n = 1,\ldots, N$$

Where: $L_j \equiv \{=1$, if system j is new (x_{jlm}); an integer <0, if system j is inherited $(w_{jlm})\}$

c. Structural constraints

$$x_j = \sum_{l=1}^{L} x_{jl} \qquad\qquad\qquad j = 1,\ldots, J \qquad (7)$$

$$x_{jl} = \sum_{m=l}^{L} x_{jlm} \qquad\qquad\qquad \begin{matrix} j = 1,\ldots, J \\ l = 1,\ldots, L \end{matrix} \qquad (8)$$

Discussion

This model formulation may be viewed as a specialization of (1) in which the objective function (2) represents the worst possible effectiveness/combat outcome which might occur, if hostilities were to break out in any one of the time periods under consideration. Each effectiveness outcome (ε_l) is derived from a force-on-force combat simulation, using the systems available for assignment 1 during period l. Assignment 1 represents the combat theater/scenario of interest.

The derivation of effectiveness from a simulation avoids the limitations of other optimization models, which ignore key aspects of force-on-force interactions, e.g., system interactions when performing a task (possibly with optimized tactics) and task interactions when determining effectiveness (possibly with optimized strategies). Consequently, this model requires, for each time period, an appropriate "black box" which accepts the numbers and types of systems available for combat and produces a vector of effectiveness outcomes. The same effectiveness simulation may be used for each time period when appropriate changes in the technological parameters, the scenario, and the threat are made. By using the output of a simulation, force effectiveness need not be separable in the various tasks and task achievement need not be separable in the various systems. By maximizing the minimum effectiveness which could occur over the horizon, the proposed approach (a) removes the requirements of some models for additivity over the time periods, (b) retains the compatibility of time frames for relating effectiveness to costs, and (c) provides one way of hedging against uncertain events (e.g., we do not always know when a war may start).

The constraint functions (C) in (1) include R&D, procurement, and operating cost components for each year of the planning horizon. The R&D cost functions (3) are step-functions which are discontinuous at the origin. The procurement functions (4) are concave to reflect the learning aspects of production. The operating cost functions (5) account for different loss rates and different expenses due to different system assignments – systems are assigned to the theater of interest as well as to other regions. They can be of any nonlinear variety. With adequate data, these cost functions can be closely fitted with piecewise linear approximations.

This formulation indicates that while we compute the costs of all the systems in the force -- regardless of their assignment -- only those systems involved in the theater of interest (assignment 1) are involved in the effectiveness simulations. Assumptions are made regarding the distribution of each type of system among each type of assignments during each time period (6). Assignments may be defined as desired, but no transfer of systems between assignments are allowed during a given time period, i.e., during the exercise of a combat simulation. Bookkeeping constraints (7) and (8) are included to insure that the model is internally consistent.

One approach for solving this model would be to fit the output of the effectiveness simulations using the techniques of nonlinear regression. This would provide analytic approximations to the effectiveness functions, which can then be integrated into the objective as the ε_l functions. If these functions turn out to be concave, then a solution method could be developed using a combination of Branch and Bound and Sequential Unconstrained Optimization techniques. If not, such a method would on yield local solutions.

SUMMARY

The problem of developing a powerful methodology for identifying cost-effective military force structures has many facets. Ideally, one would like an algorithm which generates optimal, time-phased mixes of systems, using widely accepted cost and effectiveness measures. None of the approaches examined above satisfy all these criteria.

The Force Structure Analysis Model, proposed in this paper, addresses force structure development problems with a coordinated, quasi-dynamic, optimization methodology. Among the major tasks that remain to be done, one is to specify equations (3) through (6), another is to develop a specific algorithm to solve this model, and a third is to select an appropriate combat/effectiveness simulation. These tasks will be addressed in future research.

REFERENCES

This is a conceptual paper based on the author's past experience with military force structures analysis. No references are available.

SEQUENTIAL AND INTERACTIVE SOLUTION APPROACH FOR AIRCRAFT AND CREW RECOVERY PROBLEMS

Nazan ZEYBEKCAN
Dokuz Eylül University, Dept. of Industrial Engineering
DEU. Endustri Muhendisligi Bolumu,Tinaztepe Kampusu, Buca, Izmir, TURKIYE.
Phone: 905325422706, e-mail: nazan.zeybekcan@deu.edu.tr
Irem OZKARAHAN
Troy University- Montgomery, Dept. of Computer Science, USA
Troy University Montgomery, Bartlett Hall 325, P.O. Box 4419, Montgomery, AL 36104, USA.
Phone: 334 832 7293, e-mail: iozkarahan@troy.edu

ABSTRACT

This paper presents an algorithm which solves aircraft recovery and crew recovery problems sequentially and in interactive manner. When any disruption occurs, airlines have to find a minimal cost aircraft reassignments and crew reschedules taking into account the available resources and satisfying all the operational and safety rules. In this paper, two different multi-commodity network flow problems have been modeled; one for aircraft recovery and one for crew recovery problems. Since the aircraft and crew recovery problems solutions affect each other, a solution algorithm that integrates the effects of two problems have been developed. The proposed solution algorithm has been applied to real data obtained from one of the Turkish major airlines' domestic flights.

Key words: *Aircraft recovery, crew recovery, network flow problems, airline irregular operations*

1. INTRODUCTION

In a typical day, airlines face various problems such as crew unavailability, unscheduled maintenance problems, gate delays, bad weather conditions, station congestion and airport facility restrictions that cause disruptions and the prepared airline schedules can not be operated as planned. Because of the disruptions several flights may be delayed or canceled, and aircraft and crews may miss the rest of their assigned flights.

When any of these disruptions occurs, operations personnel in the airlines must find real-time solutions in order to return the airline to its original schedule as soon as possible, considering available aircrafts, pilots, flight attendants, passengers and cargo. Some of the recovery options mostly used is as follows:

- Flights may be cancelled or delayed.
- Aircrafts may be diverted or ferried to a destination without passengers.
- Swapping aircraft among scheduled flights
- Flight attendants may be rerouted, and reserve flight attendants may be called.
- Passengers may be rescheduled, and may fly on other airlines.

While generating the recovery plan, many resources of airline such as aircraft, crew, passengers, cargo, etc. have to be re-planned. Since it is a complex task and the resource re-planning problems are usually solved sequentially. First aircraft re-assignments are made, and then crew re-scheduling problem is solved. It is more applicable to consider the each problem separately, since the integrated problem makes the solution more difficult and complex due to large number of variables to be considered and increased number of constraints to be satisfied.

In disruption management literature, aircraft recovery and crew recovery problems have been usually studied separately. Most of the recovery literature has focused on aircraft recovery problems, since the number of aircraft is much smaller than the number of crew and the rules for crews are much more complex. Teodorovic and Guberinic [26] were among the first to study the aircraft recovery problem. They discuss the problem of minimizing total passenger delays on an airline network for the schedule perturbation, by reassigning and retiming flights. The model is based on a type of connection network, which consists of two types of nodes. The first type represents the flights to be flown whereas the other represents operational aircraft. They attempt to determine the least expensive set of aircraft routings and schedule plan using a Branch and Bound procedure. Teodorovic and Stojkovic [27] consider aircraft shortage and discuss a greedy heuristic algorithm for solving a lexicographic optimization problem which considers aircraft scheduling and routing in a new daily schedule. Teodorovic and Stojkovic [28] further extend their model to include also crew and maintenance considerations. Jarrah et al. [13] present an overview of a decision support framework for airline flight cancellations and delays at United Airlines. Their underlying solution methodology is based on network flow theory. They develop two network flow models which provide solutions in the form of a set of flight delays or a set of flight cancellations. Mathaisel [17] reports on the development of a decision support system for AOCC (Airline Operations Control Centers) which integrates computer science and operations research techniques. The application integrates real-time flight following, aircraft routing, maintenance, crew management, gate assignment and flight planning with dynamic aircraft rescheduling and fleet rerouting algorithms for irregular operations. Talluri [23] deals with the problem of changing the aircraft type for a single flight while still satisfying all the constraints. They describe different algorithms for the swapping procedure in the airline schedule development process. They tested the algorithms on a connection network of two equipment types, 700 arcs and 200 nodes, 10 swapping solutions was found. Yan and Tu [30] develop a framework to assist carriers in fleet routing and flight scheduling for schedule perturbations in the operations of multifleet and multistop flights. The framework is based on a basic multifleet schedule perturbation model constructed as multiple commodity network flow problems. They use a time-line network, in which flights are represented by edges from origin to destination. Lagrangian relaxation and subgradient methods have been used for solving the problems. Yan and Yang [31] develop a decision support framework for handling schedule perturbations which incorporates concepts published by United Airlines. The framework is based on a basic schedule perturbation model constructed as a dynamic network (time-space network) from which several perturbed network models are established for scheduling following irregularities. They formulate both pure network flow problems which are solved using a network simplex algorithm, and network flow problem with side constraints, which are solved using Lagrangian relaxation with subgradient methods. Argüello, Bard and Yu [4] present a method based on the metaheuristic GRASP (Greedy Randomized Adaptive Search Procedure) to reschedule the aircraft routing during an aircraft shortage. Lou and Yu [16] address the airline schedule perturbation problem caused by the Ground Delay Program of the Federal Aviation Authorities. The goal is to improve airline dependability statistics defined by Department of Transportation as percentage of flights delayed more than 15 minutes. They design the polynomial algorithm for minimizing maximum delay among out flights. The problem is modeled as an integer program. To solve the model, they derive valid inequalities for strengthening LP relaxation bound. Cao and Kanafani [8, 9] discuss a real-time decision support tool for the integration of airline flight cancellations and delays. This research is an extension of the work of Jarrah [13], using many of the modeling concepts presented and discussed in Jarrah's paper. Thengvall, Bard and Yu

[24] present a network model with side constraints in which delays and cancellations are used to deal with aircraft shortages while ensuring a significant portion of the original aircraft routings remain intact. Bard, Yu and Argüello [6] present the time-band optimization model for reconstructing aircraft routings in response to groundings and delays experienced in daily operations, where the objective is to minimize the costs of flight delays and cancellations. This model is constructed by transforming the aircraft routing problem into a time-based network in which the time horizon is discretized. Rosenberger, Johnson and Nemhauser [20] propose a model which addresses each aircraft type as a single problem. The model principally follows an approach traditionally used in planning problems, namely a Set Partitioning master problem and a route generating procedure. Andersson and Varbrand [5] solve the complex problem of reconstructing aircraft schedules. A mixed integer multi-commodity flow model with side constraints, that each aircraft is a commodity, is developed. Side constraints are also used to model possible delays. The model is then reformulated into a set packing model using the Dantzig–Wolfe decomposition. Cancellations, delays and aircraft swaps are used to resolve the perturbation and the model ensures that the schedule returns to normal within a certain time. Two column generation schemes for heuristically solving the model are tested on real problem data obtained from a Swedish domestic airline.

Table 1. Overview of Aircraft Recovery Problem Literature

Authors	Year	Network	Cancellation	Delay	Swapping/Ferrying	Crew Considerations	Multiple Fleet	Passenger Recovery	Objective function		Solution Approach
Teodorovic and Guberinic	1984	CN		✓	✓				Min:	Total passenger delays	Branch-and Bound
Teodorovic and Stojkovic	1990	CN	✓	✓					Min:	Number of cancellations and delay minutes	Lexicographic Dynamic Programming Goal Programming Greedy Heuristics
Jarrah et al.	1993	TLN	✓	✓	✓				Min:	Delay, swap and cancellation costs	Busacker-Gowen's Dual Algorithm
Teodorovic and Stojkovic	1995	CN	✓	✓	✓	✓			Min:	Total number of cancelled flights Total passenger delays	Lexicographic Dynamic Programming Goal Programming Greedy Heuristics
Mathaisel	1996	TLN	✓	✓	✓				Min:	Revenue loss	Out-of-Kilter algorithm
Talluri	1996	CN			✓		✓		Min:	Swapping costs	Heuristic algorithm for swapping
Yan and Tu	1996	TLN	✓	✓	✓		✓		Max:	Total system profit	Lagrangian relaxation with subgradient methods
Yan and Yang	1996	TLN	✓	✓					Max:	Revenue - costs	Lagrangian relaxation with subgradient methods
Cao and Kanafani	1997	TLN	✓	✓	✓				Max:	Revenue - costs	Algorithm for 0-1 quadratic programming
Argüello et al.	1997	TBN	✓	✓	✓		✓		Min:	rerouting and cancellation costs	GRASP (Greedy Randomized Adaptive Search Procedure)
Lou and Yu	1997	Integer Programming		✓					Min:	percentage of flights delayed more than 15 minutes	LP Relaxation
Thengvall et al.	2000	TLN	✓	✓	✓				Max:	Revenue - costs	LP Relaxation Rounding Heuristic
Bard et al.	2000	TBN	✓	✓					Min:	Delay and cancellation costs	LP Relaxation Branch and Bound
Rosenberger et al.	2003	CN	✓	✓	✓	✓			Min:	Rerouting, delay and cancellation costs	Aircraft Selection Heuristic
Andersson and Varbrand	2004	CN	✓	✓	✓		✓		Max:	Revenue - costs	Column Generation
Eggenberg et al.	2009	CSN	✓	✓	✓		✓	✓	Min:	Operating, cancellation, delay costs Passenger inconvenience costs	Column Generation Dynamic Programming
Jafari and Zegordi	2010	Mixed Integer Programming	✓	✓	✓		✓	✓	Min:	Operating, cancellation, delay and passenger inconvenience costs	

* CN: connection network, TLN: time line netwok, TBN: time band network, CSN: constraint spesific network

Eggenberg et al. [11] has developed the constraint-specific recovery network model which can be seen as an extension of the time-band model by Bard et al. [6]. With this network model structural and unit- specific constraints are separated and checked independently. They have applied the model to the aircraft recovery problem with maintenance planning and passenger recovery problem. The most recent study on aircraft and passenger recovery has been performed by Jafari and Zegordi [12]. They developed a model to recover flight, aircraft and passenger simultaneously. Their recovery scope is using aircraft rotations and passengers' itineraries instead of flights which helps limiting the disruption scope and is useful to return original schedule as soon as possible.

Since the problem complexity of the crew recovery problems, these problems have not been studied as much as aircraft recovery problems. Most of the research for these problems is made in the recent years. Wei, Yu and Song [29] develop an integer programming model and an algorithm for managing crew in case of disruption. The model repairs broken pairings and assigns crew to flights that are not covered. Stojkovic et al. [21] describe the operational airline crew scheduling problem. The problem consists of modifying personalized planned monthly assignments of airline crew members during day-to-day operation. The problem requires that all flights are covered at a minimum cost while minimizing the disturbances of crew members. They formulate the crew recovery problem as an integer non-linear multi-commodity flow problem. Lettovsky et al. [15] present a method based on an integer programming formulation. They develop a new solution framework. It provides, in almost real time, a recovery plan for reassigning crews to restore a disrupted crew schedule. Stojkovic et al. [22] present a model that involves determining appropriate real-time changes to planned airline schedules when perturbations occur to minimize customer inconvenience and costs to the airline. They propose a model that determines new flight schedules based on planned crew transfers, rest periods, passenger connections and maintenance. Medard and Sawhney [18] consider the crew recovery problem and integrate both crew pairing and crew rostering to solve time critical crew recovery problems arising on the day of operations. Abdelhany et al. [2] present a decision support tool that automates crew recovery during irregular operations for large scale commercial airlines. Nissen and Haase [19] present a new duty-period-based formulation for airline crew rescheduling problem where the aim is to determine new crew assignments minimizing the impact on the original crew schedule, after a disturbance in the schedule.

Table 2. Overview of Crew Recovery Problem Literature

Authors	Year	Models	Uncovered flights	Flight delays	Deadheading	Stand-by / Reserve crew	Modifications to schedule	Objective function	Solution Approach
Wei et al.	1997	Integer Multicommodity Network Flow problem	✓		✓	✓		Min: the number of uncovered flights	Depth-First Branch-and-Bound
Stojkovic et al.	1998	Set Partitioning Problem	✓		✓		✓	Min: pairing, deadheading, undercovering costs the disturbances of crew members	Column Generation Branch-and-Bound
Lettovsky et al.	2000	Set Covering Problem	✓				✓	Min: crew reassignment and cancellation costs	Column Generation Branch-and-Bound
Stojkovic et al.	2002	PERT/CPM (time constrained models) (dual: Network Problem)	✓	✓			✓	Min: modifications, uncovered flights, delay costs	Dual Problem
Medard and Sawhney	2003	Set Covering Problem	✓				✓	Min: illegal crew, uncovered flights, affected crew costs	Simple Tree Search Column Generation
Abdelhany et al.	2004	Mixed Integer Program		✓	✓	✓	✓	Min: deadheading, standby, swap, flight delay costs	Optimization Tool
Nissen and Haase	2006	Duty-period-based Network Model					✓	Min: the costs of changing each crew's original schedule	Branch-and Price

Because of the complex nature of the integrated aircraft and crew recovery problem, only a few researches have been made on integrated problems. Lettovsky's Ph.D. thesis [14] is the first to consider truly integrated approach in the literature. His thesis presents a linear mixed-integer mathematical problem that maximizes total profit of the airline while capturing availability of aircrafts, crews and passengers. Lettowsky suggests solving problem using decomposition algorithms, but his approach has not been completely tested. Furthermore, Bratu and Barnhart [7] presents two models that considers aircraft and crew recovery and through the objective function focuses on passenger recovery. While reserve crews are included into the models they do not consider how to recover disrupted crews. They present two models: passenger delay metric and disrupted passenger metric. Both have same objective function which incorporates operation costs and passenger recovery costs. They test both models and conclude that only the disrupted passenger metric model is fast enough to be used in a real-time environment. The most recent studies on integrated recovery problems have been performed by Abdelghany et al. [3]. They have presented a decision support tool which integrates schedule simulation model and a resource assignment optimization model. The simulation model predicts the list of disrupted flights in the system. The optimization model is formulated as mixed integer programming and combines different recovery actions to minimize projected flight delays and cancellations. Although the developed tool is capable to generate efficient recovery plan, several extensions are still possible.

The general assumption in the aircraft recovery problem literature is that the crew members are available at any time. Most of the studies do not include the crew availability constraints in the aircraft recovery problems. As seen in *Table 1*, only a few studies involve the crew related considerations in the problems. However, these problems which involve crew considerations do not attempt to recover disrupted crew. Similarly crew recovery problems do not consider the aircraft availability. They only focus on recovering disrupted crew without considering other airline resources. On the other hand, integrated problems consider both aircraft and crew availabilities. However, these problems become very complex due to large number of the constraints and variables coming from the integrated problem. The integrated problem is much harder to be solved. They require special attention and complex techniques in order to be solved to optimality in real time.

In this paper, both aircraft recovery and crew recovery problems will be considered. Furthermore, a new solution algorithm will be presented which makes possible to solve the two problems together including the correlation between them and without integrating the whole problem. Thus, with this algorithm both aircraft and crew availabilities will be considered and both aircraft and crew disruptions will be recovered without having the complexity of the integrated problem. First, the aircraft recovery and crew recovery problems will be introduced individually. Then, taking into account the correlation between the two problems, sequential and interactive solution approach will be used to find real time solutions to both aircraft and crew disruptions.

The paper is organized as follows: In section 2, aircraft recovery problem related assumptions, the mathematical model and notations will be presented. In section 3, crew recovery problem related assumptions, the mathematical model and notations will be presented. In section 4, the sequential and interactive solution approach for aircraft recovery and crew recovery problems and the advantages of this approach will be presented. In section 5, computational experiments will be given. Finally, the section 6 will be on the conclusions.

2. AIRCRAFT RECOVERY PROBLEM : THE MULTI-COMMODITY NETWORK FLOW MODEL

The aircraft recovery problem presented in this paper is formulated as multi-commodity network flow problem where each commodity represents an aircraft. The underlying network used in the model is connection network, where nodes represent scheduled flight legs with corresponding departure/arrival times and origin/destination stations. The arcs in the network represent connections between two consecutive flight legs where the destination of the first flight leg is the same as the origin of the next flight leg. Node i representing the flight leg l_i is connected to node j representing the flight leg l_j by a directed arc, if it is feasible to flight leg l_j after flight leg l_i by the same aircraft with respect to turn over times and other operational rules. In addition there is a set of origin and destination nodes representing the possible positions of aircraft at the beginning and at the end of the planning horizon.

The problem is formulated considering the multiple aircrafts shortage and crew availability constraint. Most of the recovery problems in the literature have not taken into account the crew availability constraints. However, the crew availability directly affects the flight legs' operability. There should always be sufficient number of crews in order to fly all scheduled flight legs. Although there is an available aircraft to be assigned a scheduled flight leg, the flight cannot be operated until the available crew is found to fly it. Thus, the crew availability related constraints are also included into the problem.

The Aircraft Recovery problem presented in this paper considers 3 recovery strategies:

- Cancelling flights
- Swapping aircrafts among scheduled flights.
- Delaying flights

The notation used in this problem is as follows:

K	: set of aircrafts
R	: set of available aircrafts in the recovery period
F	: set of all flight legs
F^k	: set of flight legs that can be flown by aircraft k
N^k	: rerouting connection network; set of flight connections for aircraft k during recovery period. (Feasible connection between two consecutive flight legs)
N_{30}^k	: 30 minutes of delayed connection network; set of flight connections for aircraft k during recovery period.
N_{60}^k	: 60 minutes of delayed connection network; set of flight connections for aircraft k during recovery period.
L^k	: set of flight connections for aircraft k before the recovery period begins
M^k	: set of flight connections for aircraft k after the recovery period finishes
k	: index representing the aircrafts
i	: index representing the first flight in the connection arc
j	: index representing the next flight in the connection arc
arr_i	: arrival time of flight i
$dept_j$	: departure time of flight j
org_j	: origin station of flight j
$dest_i$	: destination station of flight i
$turnover$	: minimum required turnover time between two consecutive flight legs
$lower$	: lower bound on the number of flight legs that an aircraft can fly
$upper$	: upper bound on the number of flight legs that an aircraft can fly
sch_{ij}^k	: the original aircraft – flight assignment. It is equal to 1, if the aircraft k is assigned to two consecutive flights i and j in the original schedule, 0 otherwise.

c_{ij}^k : the cost of rerouting aircraft k between flights i and j

d_{ij}^k : the cost of delaying aircrafts.

X_{ij}^k : the binary decision variable which takes value 1 if the aircraft k fly two consecutive flights i and j and 0 otherwise.

Y_j : the binary decision variable which takes value 1 if the flight j is cancelled and 0 otherwise.

The mathematical model of the aircraft recovery problem is as follows:

$$Minimize \sum_{k \in K} \sum_{(i,j) \in N^k \cup M^k \cup L^k} c_{ij}^k X_{ij}^k + \sum_{k \in K} \sum_{(i,j) \in N_{30}^k \cup N_{60}^k} d_{ij}^k X_{ij}^k + \sum_{i \in F} b_i Y_i \tag{1.1}$$

Subject to:

$$\sum_{k \in K} \sum_{j:(i,j) \in N^k \cup M^k \cup L^k} X_{ij}^k + \sum_{k \in K} \sum_{j:(i,j) \in N_{30}^k \cup N_{60}^k} X_{ij}^k + Y_i = 1 \qquad \forall i \in F^k \tag{1.2}$$

$$X_{ij}^k = 0 \qquad \forall k \in K - \{R\} \quad \forall(i,j) \in N^k \cup N_{30}^k \cup N_{60}^k \tag{1.3}$$

$$X_{ij}^k = sch_{ij}^k \qquad \forall k \in K \ \forall(i,j) \in L^k \tag{1.4}$$

$$X_{ij}^k = sch_{ij}^k \qquad \forall k \in K \ \forall(i,j) \in M^k \tag{1.5}$$

$$\sum_{j:(i,j) \in N^k \cup N_{30}^k \cup N_{60}^k} X_{ij}^k - \sum_{j:(j,i) \in N^k \cup N_{30}^k \cup N_{60}^k} X_{ji}^k = 0 \qquad \forall k \in R \ \forall i \in F^k \tag{1.6}$$

$$X_{ij}^k \left(arr_i + turnover - dept_j \right) = 0 \qquad \forall k \in R \ \forall(i,j) \in N^k \cup N_{30}^k \cup N_{60}^k \tag{1.7}$$

$$X_{ij}^k \left(dest_i - org_j \right) = 0 \qquad \forall k \in R \ \forall(i,j) \in N^k \cup N_{30}^k \cup N_{60}^k \tag{1.8}$$

$$\sum_{(i,j) \in N^k \cup M^k \cup L^k \cup N_{30}^k \cup N_{60}^k} X_{ij}^k \leq lower \qquad \forall k \in K \tag{1.9}$$

$$\sum_{(i,j) \in N^k \cup M^k \cup L^k \cup N_{30}^k \cup N_{60}^k} X_{ij}^k \geq upper \qquad \forall k \in K \tag{1.10}$$

$$\sum_{k \in K - \{R\}} \sum_{j:(i,j) \in N^k \cup M^k \cup L^k \cup N_{30}^k \cup N_{60}^k} X_{ij}^k \leq Xac_i^k \qquad \forall i \in F^k \tag{1.11}$$

$$X_{ij}^k \in \{0,1\} \qquad \forall k \in K \quad \forall(i,j) \in N^k \cup M^k \cup L^k \cup N_{30}^k \cup N_{60}^k \tag{1.12}$$

$$Y_i \in \{0,1\} \qquad \forall i \in F \tag{1.13}$$

where the objective function minimizes the total cost of rerouting aircrafts, delaying flights and cancelling flights. The constraint (1.2) is the flight cover constraint and ensures that all flight legs will be either covered by an aircraft (delayed or on time) or cancelled. This constraint also ensures that flight legs can be either a member of rerouting connection network or delay connection network. The constraint (1.3) ensures that during recovery period no flights will be assigned to the shortage aircrafts. The constraints (1.4) and (1.5) ensure that before and after the recovery period the aircraft assignment will be same as the original schedule. In other words, it will be returned to the original flight aircraft assignment after the shortage aircrafts is recovered. The constraint (1.6) is the flow conservation constraint which ensures that the total arc flow which enters the node i must be equal to the total arc flow which leaves the node i for the available aircrafts during recovery period. The constraint (1.7) guarantees the time requirements between consecutive flight legs: there should be at least the minimum required turnover time between the arrival of the first flight and the departure of the second flight for the aircrafts available during recovery period. The constraint (1.8) ensures that the destination station of the first flight and the origin of the second flight are the same for the aircrafts available during recovery period. Constraints (1.9) and (1.10) guarantees the usage balance for the aircrafts ensuring the upper and lower bounds on the number of flight legs that an aircraft can fly. The constraint (1.11) ensures that if there is not sufficient number of crews for flying any flight legs, this flight leg cannot be flown. Thus, the decision variables for the aircraft- flight leg assignment will take value 0. The right hand side of this constraint directly takes value from the solution of the crew recovery problem with the automatic calculation of number of available crews which will be explained in the next section. This constraint may be considered as connection constraint between two recovery problems. Constraints (1.12) and (1.13) ensure that all the decision variables are binary.

3. CREW RECOVERY PROBLEM : THE MULTI-COMMODITY NETWORK FLOW MODEL

The crew recovery problem is modeled similar to the aircraft recovery problem. Thus, multi-commodity network flow model is also developed for crew recovery problem. In this case, each commodity represents each crew. The connection network is used as underlying network as in the aircraft recovery problem. Different form aircraft recovery problem, in the crew recovery problem crew related regulations are included into the model as new constraints.

In this problem following crew recovery strategies are taken into account:

- Reserve/stand by crews can be used to fly uncovered flight leg.
- Good (undistrupted) crews can be swapped to cover broken pairs

The notation used in the problem formulation is as follows:

K : set of crews
A : set of available crews in the recovery period
F : set of all flight legs
F^k : set of flight legs that can be flown by crew k
N^k : rescheduling network; set of flight connections for crew k during recovery period. (Feasible connection between two consecutive flight leg)
N_{30}^k : 30 minutes of delayed connection network; set of flight connections for aircraft k during recovery period.
N_{60}^k : 60 minutes of delayed connection network; set of flight connections for aircraft k during recovery period.

L^k	: set of flight connections for crew k before the recovery period begins
M^k	: set of flight connections for crew k after the recovery period finishes
k	: index representing the crews
i	: index representing the first flight in the connection arc
j	: index representing the next flight in the connection arc
h	: index for flight legs
NC_h	:total number of available crews that can fly flight leg h
arr_i	: arrival time of flight i
$dept_j$	: departure time of flight j
org_j	: origin station of flight j
$dest_i$	: destination station of flight i
$resttime$	: minimum required rest time time between two consecutive flight legs
$flytime$	: maximum allowed flying time for a crew
$dutytime$	: maximum allowed duty time for a crew
$lower$	: minimum number of flight legs assigned to crews
$upper$	: maximum number of flight legs assigned to crews
$pairing_{ij}^k$	: the original crew – flight assignment. It is equal to 1, if the crew k is assigned to two consecutive flights i and j in the original schedule, 0 otherwise.
c_{ij}^k	: the cost of reassigning crew k between flights i and j
r_j	: the cost of using reserve/standby crew for flight j.
b_j	: the cost of cancelling flight j.
X_{ij}^k	: the binary decision variable which takes value 1 if the crew k fly two consecutive flights i and j and 0 otherwise.
Y_j	: the binary decision variable which takes value 1 if the flight j is cancelled and 0 otherwise.
Z_j	: the binary decision variable represents the usage of reserve/standby crew. It takes value 1 if the flight j is assigned to reserve crew and 0 otherwise.

The mathematical formulation of the problem is as follows:

$$Minimize \sum_{k \in K} \sum_{(i,j) \in N^k \cup M^k \cup L^k} c_{ij}^k X_{ij}^k + \sum_{k \in K} \sum_{(i,j) \in N_{30}^k \cup N_{60}^k} d_{ij}^k X_{ij}^k + \sum_{i \epsilon F} b_i Y_i + \sum_{i \epsilon F} r_i Z_i \qquad (2.1)$$

Subject to:

$$\sum_{k \epsilon K} \sum_{j:(i,j) \in N^k \cup M^k \cup L^k} X_{ij}^k + \sum_{k \in K} \sum_{(i,j) \in N_{30}^k \cup N_{60}^k} X_{ij}^k + Y_i + Z_i = 1 \qquad \forall i \in F^k \qquad (2.2)$$

$$X_{ij}^k = 0 \qquad \forall k \in K - \{R\} \; \forall(i,j) \in N^k \cup N_{30}^k \cup N_{60}^k \qquad (2.3)$$

$$X_{ij}^k = sch_{ij}^k \qquad \forall k \in K \; \forall(i,j) \in L^k \qquad (2.4)$$

$$X_{ij}^k = sch_{ij}^k \qquad \forall k \in K \; \forall(i,j) \in M^k \qquad (2.5)$$

$$\sum_{j:(i,j) \epsilon N^k \cup N_{30}^k \cup N_{60}^k} X_{ij}^k - \sum_{j:(j,i) \epsilon N^k \cup N_{30}^k \cup N_{60}^k} X_{ji}^k = 0 \qquad \forall k \in R \; \forall i \in F^k \qquad (2.6)$$

$$X_{ij}^k \left(arr_i + resttime - dept_j \right) = 0 \qquad \forall k \in R \; \forall(i,j) \in N^k \cup N_{30}^k \cup N_{60}^k \qquad (2.7)$$

$$X_{ij}^k \left(dest_i - org_j \right) = 0 \qquad\qquad \forall k \in R \quad \forall (i,j) \in N^k \cup N_{30}^k \cup N_{60}^k \qquad (2.8)$$

$$\sum_{(i,j) \in N^k \cup M^k \cup L^k \cup N_{30}^k \cup N_{60}^k} X_{ij}^k \leq lower \qquad\qquad \forall k \in K \qquad (2.9)$$

$$\sum_{(i,j) \in N^k \cup M^k \cup L^k \cup N_{30}^k \cup N_{60}^k} X_{ij}^k \geq upper \qquad\qquad \forall k \in K \qquad (2.10)$$

$$\sum_{i:(i,j) \in N^k \cup M^k \cup L^k \cup N_{30}^k \cup N_{60}^k} X_{ij}^k \left(arr_i - dept_i \right) \leq flytime \qquad\qquad \forall k \in K \qquad (2.11)$$

$$\sum_{i:(i,j) \in N^k \cup M^k \cup L^k \cup N_{30}^k \cup N_{60}^k} X_{ij}^k \left(dept_i - arr_j \right) \leq dutytime \qquad\qquad \forall k \in K \qquad (2.12)$$

$$\sum_{i \in F} Z_i \left(dept_j - arr_i \right) \leq flytime \qquad\qquad\qquad (2.13)$$

$$\sum_{i \in F} Z_i \left(dept_i - arr_j \right) \leq dutytime \qquad\qquad\qquad (2.14)$$

$$NC_h = 1 - Y_h + \sum_{\substack{k \in K \\ }} \sum_{\substack{(i,j) \in N^k \cup N_{30}^k \cup N_{60}^k \\ arr[i] < dept[h] \\ dept[j] > arr[h]}} X_{ij}^k \qquad\qquad \forall h \in F \qquad (2.15)$$

$$X_{ij}^k \epsilon \{0,1\} \qquad\qquad \forall k \in K \; \forall (i,j) \in N^k \cup M^k \cup L^k \cup N_{30}^k \cup N_{60}^k \qquad (2.16)$$

$$Y_i \epsilon \{0,1\} \qquad\qquad\qquad\qquad \forall i \in F \qquad (2.17)$$

$$Z_i \epsilon \{0,1\} \qquad\qquad\qquad\qquad \forall i \in F \qquad (2.18)$$

where the objective function minimizes the total cost of reassigning crews to the flight legs, using reserve crew, delaying flight legs and cancelling flights. The constraint (2.2) is the flight cover constraint and ensures that all flight legs will be either covered by a good crew (delayed or on time) or by a reserve crew or cancelled. The constraint (2.3) ensures that during recovery period no flights will be assigned to the shortage crew. The constraints (2.4) and (2.5) ensure that before and after the recovery period the crew assignment will be same as the original schedule. In other words, it will be returned to the original crew flight assignment after the shortage crew becomes available again. The constraint (2.6) ensures that the total arc flow which enters the node i must be equal to the total arc flow which leaves the node i for the available crews during recovery period. The constraint (2.7) guarantees the time requirements between consecutive flight legs: there should be at least the minimum required rest time between the arrival of the first flight and the departure of the second flight for the available crews during recovery period. The constraint (2.8) ensures that the destination station of the first flight and the origin of the second flight are the same for the available

crews during recovery period. Constraints (2.9) and (2.10) limit the number of flight legs to be flown by a crew. Constraint (2.11) limits the total flying time for a crew. A crew cannot fly more than predefined total flying time in one pairing. Constraint (2.12) limits the duty time for a crew. Total time spent on duty by a crew can not be more than predefined total duty time. Constraint (2.13) limits the total flying time for a reserve/standby crew. Constraint (2.14) limits the total duty time for a reserve/standby crew. Constraint (2.15) is used for computing the number of available crews at certain time at certain station. The result of this computation will be used by the aircraft recovery problem (see section 2, constraint (1.11).

The total number of available crews is computed as follows:

- The crew originally assigned to flight h is taken into consideration.
- If the flight h is cancelled the sum will be affected accordingly
- For flight h, the total number of crews who are stationed on station $org[h]$ waiting for their next flight leg between the departure and arrival time of the flight h is computed.

Finally, constraints (2.16) - (2.18) ensure that all the decision variables are binary.

4. SEQUENTIAL AND INTERACTIVE SOLUTION APPROACH

While solving airline resource scheduling problems, the crew and aircraft related decisions affect each other. In order to operate any flight leg there must be at least 1 available crew and 1 available aircraft to be assigned to the flight leg. The same consideration exists also for recovery problems. In case of schedule disruption even if there is an available aircraft for the flight leg, subject flight leg can not be operated until an available crew is found or wise versa.

When the aircraft and crew recovery problems are solved separately, some feasibility problems occur due to lack of congruency of the solutions. For example, the solution obtained from the crew recovery problem indicates that one of the flight legs should be cancelled. On the other hand, the solution obtained from aircraft recovery problem indicates that the subject flight leg is covered by an available aircraft. In this case, there is a flight leg that is covered by an aircraft but not covered by a crew. That causes incongruence between the solutions of two problems; aircraft recovery problem should take into account the crew availabilities and the crew recovery problem should consider the aircraft availabilities. If one of the resources is not available for a flight leg, the subject flight leg should be cancelled or delayed until available aircraft and/or crew is found. By integrating the two problems subject incongruence can be dealt. However integrated problem has other challenges.

The integrated crew and aircraft recovery problems are complex in terms of modeling. Moreover, the integrated problem includes both crew and aircraft related constraints and variables. Thus, the number of constraints and variables become very huge compared to the individual recovery problems. Because of the increased complexity and number of variables, the integrated problem is NP-hard.

When aircraft and recovery problems are solved individually and independently, above mentioned incongruence may occur any time. By taking into account the dependency and correlation between the two problems a new sequential and interactive solution algorithm is developed in order to solve aircraft and crew recovery problems in real time without

integrating the problems. Thus, the complexity resulted from the integration will be reduced and both aircraft and crew disruptions will be recovered at one time. The algorithm is as follow:

1. Solve the initial airline crew recovery problem.
2. Obtain a new crew schedule for recovery period.
3. According to the obtained crew re-schedule compute the number of available crews.
4. Send the computed number of available crews to the aircraft recovery problem.
5. Use the number of available crews as a resource constraint in the aircraft recovery problem.
6. Solve the aircraft recovery problem by using number of available crews.
7. Obtain new aircraft-flight assignments for recovery period.
8. Compare the congruency of the solutions of the aircraft and crew recovery problems.
    - Are all the flight legs covered by a aircraft in both problems?
    - Are all the flight legs covered by crew in both problems?
    - If any flight leg is cancelled in one problem, is it also cancelled in the other problem?
    - If any flight leg is delayed in one problem, it is also delayed in the other problem?
9. If they are congruent Stop.
10. If not send the solution of aircraft recovery problem to the crew recovery problem.
11. Obtain restricted airline crew recovery problem.
12. Go to Step 2

Below *Algorithm 1* represents the interactive aircraft and crew recovery algorithm which will later be converted in ILOG script language.

Begin
Models : Crew (Crew recovery model), Aircraft (Aircraft recovery model).
Data : DF1 = 1, DF2 = 1.

Repeat

        If (DF1$\neq$ 0 & DF2$\neq$ 0)

      Solve model Crew.
          For all flight legs:
          Aircraft. NOAC[i] = Crew. NOAC[i] (number of available crew)
          End {**for**}
      Solve model Aircraft.
          For all flight legs:
          DF1= Σ (Crew.cancel[i] - Aircraft.cancel[i])
          DF2= Σ (Crew.X[i,j].dept[i] - Aircraft. X[i,j].dept[i])
          End {**for**}
      End {**if**}
End {**repeat**}

End {algorithm}
Algorithm 1. The interactive aircraft and crew recovery algorithm
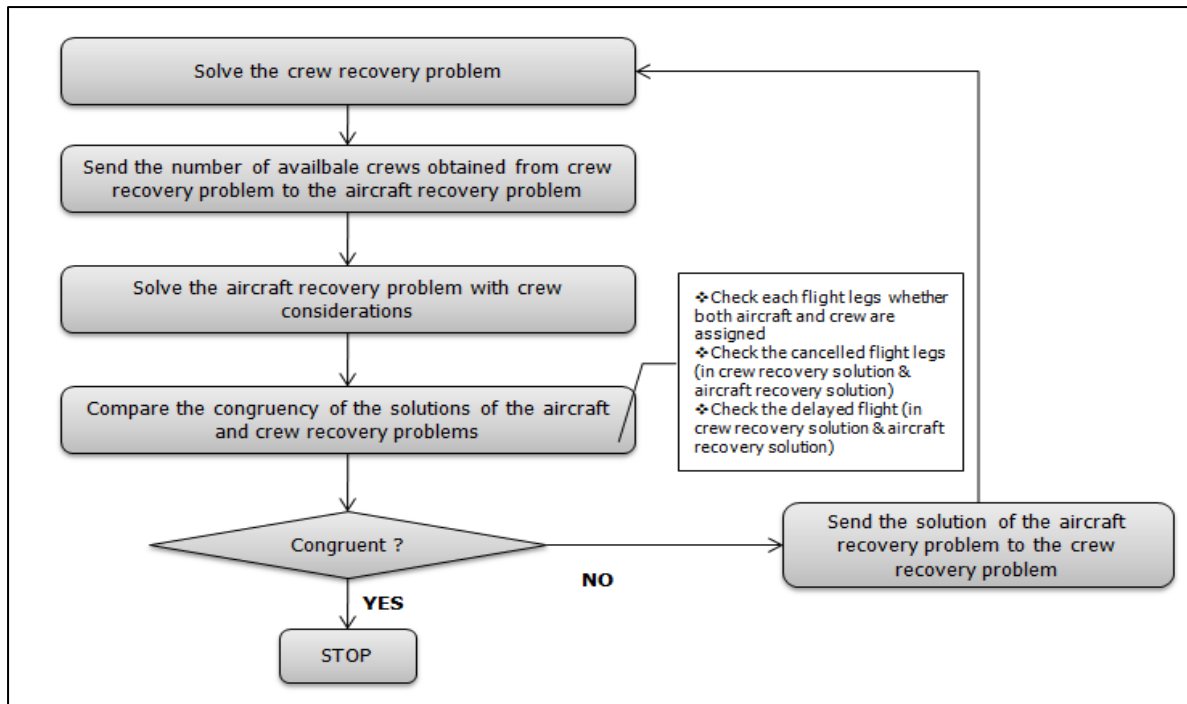
Figure 1. The sequential and interactive algorithm for aircraft and crew recovery problem

The advantages of the sequential and interactive algorithm can be summarized as follows:

- Both aircraft and crew availabilities are considered in the algorithm. Without integration, both aircraft and crew disruptions are recovered at one time.
- Since the solution obtained for the aircraft and crew recovery problems without integration, the problem complexity and huge number of constraints and variables are prevented. Thus, the overall problem is less complex than the integrated problem in terms of modeling and the real time solutions can be obtained in easier and more practical way.
- Several combinations of aircraft and crew availabilities can be considered. The algorithm can find solutions for multiple dependent or independent crew and aircraft unavailabilities: In some cases only aircraft disruptions can be recovered. In some cases aircraft and crew disruptions for same flight leg can be considered or aircraft unavailability for flight *a* and crew unavailability for flight *b* can be solved.

5.   COMPUTATIONAL EXPERIMENTS

Computational experiments have been applied to one of the major Turkish airlines' daily domestic flight schedule. There are total of 38 flights, all the flights are flown within the same day over a network of 12 stations. There are 2 different fleet types and total of 9 aircrafts: Boeing MD-88 and Boeing MD-83. There are 5 aircrafts of Boeing MD-88 type and 4 aircrafts of Boeing MD-83 type. Both types of aircrafts have the same technical specifications and are capable of flying all the domestic flights of the airline. So connection networks for each aircraft type are same and as seen in *Figure 2*. The *Table 3* presents the original flight schedule, flight- aircraft assignment and flight-crew assignment.

Table 3. The original flight schedule and aircraft/ crew assignments

Flight No	Origin	Destination	Departure	Arrival	Aircraft	Crew No
OHY010	IST	ADA	06:45	08:15	MD83c	Crew10
OHY011	ADA	IST	09:10	10:40	MD83c	Crew10
OHY016	IST	ADA	15:00	16:30	MD83c	Crew4
OHY017	ADA	IST	17:30	19:00	MD83c	Crew4
OHY018	IST	ADA	19:00	20:30	MD88a	Crew2
OHY019	ADA	IST	21:30	23:00	MD88a	Crew2
OHY022	IST	AYT	08:15	09:15	MD88c	Crew2
OHY023	AYT	IST	10:45	11:45	MD88c	Crew2
OHY024	IST	AYT	16:45	17:45	MD83b	Crew9
OHY025	AYT	IST	18:45	19:45	MD83b	Crew9
OHY026	IST	AYT	20:30	21:30	MD83d	Crew1
OHY027	AYT	IST	07:30	08:30	MD83b	Crew1
OHY028	IST	BJV	11:55	12:55	MD83d	Crew10
OHY029	BJV	IST	13:45	14:45	MD83d	Crew10
OHY034	IST	DIY	06:45	08:30	MD88a	Crew11
OHY035	DIY	IST	09:30	11:15	MD88a	Crew11
OHY036	IST	DIY	18:45	20:30	MD88b	Crew6
OHY037	DIY	IST	21:30	23:15	MD88b	Crew6
OHY038	IST	ERZ	12:30	14:20	MD88c	Crew8
OHY039	ERZ	IST	15:20	17:10	MD88c	Crew8
OHY042	IST	GZT	09:30	11:15	MD83a	Crew5
OHY043	GZT	IST	12:15	14:00	MD83a	Crew5
OHY050	IST	ADB	07:45	08:45	MD88d	Crew4
OHY051	ADB	IST	10:50	11:45	MD88d	Crew4
OHY054	IST	ADB	16:30	17:30	MD88d	Crew3
OHY055	ADB	IST	18:50	19:45	MD88d	Crew3
OHY058	IST	ADB	20:45	21:45	MD88e	Crew11
OHY059	ADB	IST	07:30	08:25	MD83d	Crew6
OHY062	IST	MLX	12:15	13:45	MD88e	Crew1
OHY063	MLX	IST	14:30	16:00	MD88e	Crew1
OHY066	IST	KSY	07:15	08:30	MD88e	Crew3
OHY067	KSY	IST	09:15	10:30	MD88e	Crew3
OHY074	IST	SZF	09:30	10:45	MD83b	Crew9
OHY075	SZF	IST	11:45	13:00	MD83b	Crew9
OHY080	IST	TZX	06:50	08:25	MD88b	Crew7
OHY081	TZX	IST	09:20	10:55	MD88b	Crew7
OHY086	IST	TZX	18:45	20:20	MD83a	Crew12
OHY087	TZX	IST	21:15	22:50	MD83a	Crew12

The connection network of aircrafts and crews have been designed by taking into account the aviation and crew feasibility rules such as limit on the total flying time in duty, limit on the total length of a duty, minimum required rest time for crews between two flight legs, turnover time for aircrafts between two flight legs and etc.
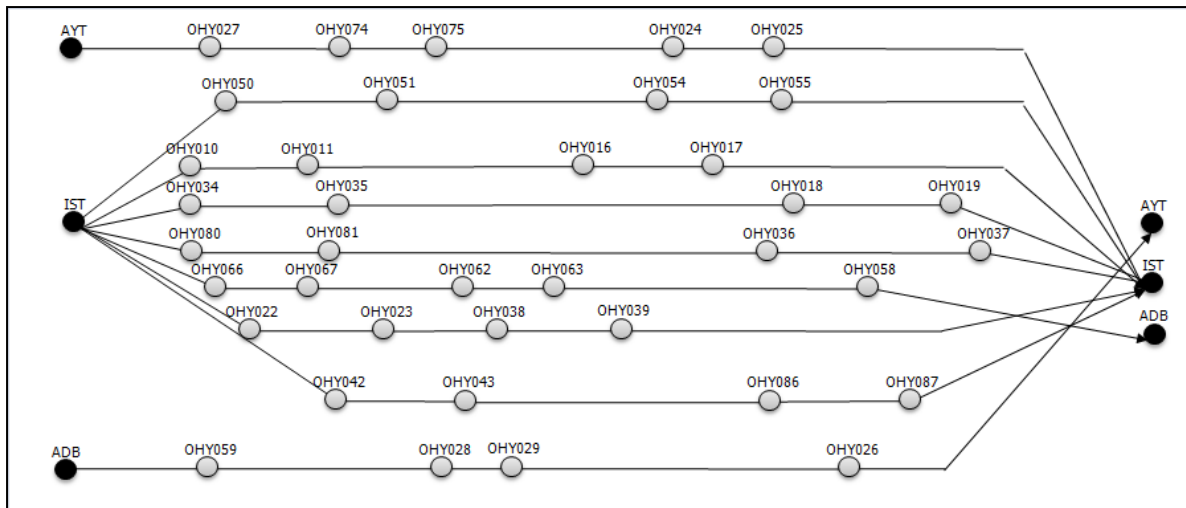
Figure 2. Connection network representation of aircraft assignments

The input data of the overall problem involves:

- The flight schedule of the domestic flight legs
- The initial aircraft-flight leg assignment
- The initial crew assignment
- The connection network of the aircrafts
- The connection network of the crews
- The reassignment cost, delay cost and cancellation costs

Before applying the sequential and interactive solution approach, the individual aircraft recovery and crew recovery problems have been solved. In order to solve integer programming models of the problems which have been explained in sections 2 and 3, ILOG OPL Studio 3.7. has been used as optimization software. Moreover, the sequential and interactive solution algorithm has also been coded with the Script in ILOG. The ILOG script has been written based on the algorithm represented in *Algorithm 1*.

Several cases have been tested in order to evaluate the efficiency of the approach: Multiple aircraft disruptions, both aircraft and crew disruption of the same flight leg, and also aircraft and crew disruptions for different flight legs. Only an example case's results will be presented in this paper.

As an example case, there occurs a shortage on crew "Crew 3" and on aircraft "MD88e" between 06:00 and 12:00. During the recovery period crew "Crew3" and aircraft "MD88e" which is originally assigned to flight legs "OHY066" and "OHY067" cannot perform their corresponding flight legs. The problem is to find a feasible aircraft re-assignment and crew re-schedule during recovery period with limited available resources by minimizing the total operating and disruption costs.

Table 4. The size and the solution time of the example problem

Problem	# of Stations	# of Flight Legs	# of Resources	# of Constraints	# of Variables	Solution Time Individual Prob. (in sec.)	Solution Time Interactive Prob. (in sec.)
Crew Recovery Problem	12	38	12 crews	7394	25040	1.29	
Aircraft Recovery Problem	12	38	8 aircrafts	6190	18719	0.97	2.573

The size of the example problem including number of constraints and variables to be considered can be seen in *Table 4*. *Table 4* also presents the solution times of the problems in following situations:

- when the aircraft recovery problem is solved individually,
- when the crew recovery problem is solved individually,
- when the sequential and interactive solution algorithm applied and aircraft and crew recovery problems solved together.

As it can be seen in *Table 4*, the sequential and interactive solution approach can solve both problem in real time, in about 2.5 seconds.

The crew reschedule and aircraft reassignment solution can be seen in *Table 5*. As it can be seen from the *Table 5*, no flight leg has to be cancelled or delayed because of crew or aircraft unavailability. Since the original flight schedule is not too tight, the remaining/available aircrafts and crews can cover all the flight legs with various aircraft swapping operations. However, 2 flight legs remain uncovered by a crew, and reserve crew is used for covering these flight legs.

Table 5. The recovered aircraft and crew assignments of example problem

Flight No	Origin	Destination	Departure	Arrival	Aircraft Assignment			Crew Assignment		
					Original	Recovered	Action	Original	Recovered	Action
OHY010	IST	ADA	06:45	08:15	MD83c	MD83d	S	Crew10	Crew10	-
OHY011	ADA	IST	09:10	10:40	MD83c	MD83d	S	Crew10	Crew10	-
OHY016	IST	ADA	15:00	16:30	MD83c	MD83c	-	Crew4	Crew4	-
OHY017	ADA	IST	17:30	19:00	MD83c	MD83c	-	Crew4	Crew4	-
OHY018	IST	ADA	19:00	20:30	MD88a	MD88a	-	Crew2	Crew2	-
OHY019	ADA	IST	21:30	23:00	MD88a	MD88a	-	Crew2	Crew2	-
OHY022	IST	AYT	08:15	09:15	MD88c	MD88d	S	Crew2	Crew2	-
OHY023	AYT	IST	10:45	11:45	MD88c	MD88d	S	Crew2	Crew2	-
OHY024	IST	AYT	16:45	17:45	MD83b	MD83b	-	Crew9	Crew9	-
OHY025	AYT	IST	18:45	19:45	MD83b	MD83b	-	Crew9	Crew9	-
OHY026	IST	AYT	20:30	21:30	MD83d	MD83d	-	Crew1	Crew1	-
OHY027	AYT	IST	07:30	08:30	MD83b	MD83a	S	Crew1	Crew1	-
OHY028	IST	BJV	11:55	12:55	MD83d	MD83d	-	Crew10	Crew10	-
OHY029	BJV	IST	13:45	14:45	MD83d	MD83d	-	Crew10	Crew10	-
OHY034	IST	DIY	06:45	08:30	MD88a	MD88c	S	Crew11	Reserve	R
OHY035	DIY	IST	09:30	11:15	MD88a	MD88c	S	Crew11	Reserve	R
OHY036	IST	DIY	18:45	20:30	MD88b	MD88b	-	Crew6	Crew6	-
OHY037	DIY	IST	21:30	23:15	MD88b	MD88b	-	Crew6	Crew6	-
OHY038	IST	ERZ	12:30	14:20	MD88c	MD88c	-	Crew8	Crew8	-
OHY039	ERZ	IST	15:20	17:10	MD88c	MD88c	-	Crew8	Crew8	-
OHY042	IST	GZT	09:30	11:15	MD83a	MD83a	-	Crew5	Crew5	-
OHY043	GZT	IST	12:15	14:00	MD83a	MD83a	-	Crew5	Crew5	-
OHY050	IST	ADB	07:45	08:45	MD88d	MD83b	S	Crew4	Crew4	-
OHY051	ADB	IST	10:50	11:45	MD88d	MD83b	S	Crew4	Crew4	-
OHY054	IST	ADB	16:30	17:30	MD88d	MD88d	-	Crew3	Crew3	-
OHY055	ADB	IST	18:50	19:45	MD88d	MD88d	-	Crew3	Crew3	-
OHY058	IST	ADB	20:45	21:45	MD88e	MD88e	-	Crew11	Crew11	-
OHY059	ADB	IST	07:30	08:25	MD83d	MD83c	S	Crew6	Crew11	S
OHY062	IST	MLX	12:15	13:45	MD88e	MD88e	-	Crew1	Crew1	-
OHY063	MLX	IST	14:30	16:00	MD88e	MD88e	-	Crew1	Crew1	-
OHY066	IST	KSY	07:15	08:30	MD88e	MD88a	S	Crew3	Crew9	S
OHY067	KSY	IST	09:15	10:30	MD88e	MD88a	S	Crew3	Crew9	S
OHY074	IST	SZF	09:30	10:45	MD83b	MD83c	S	Crew9	Crew11	S
OHY075	SZF	IST	11:45	13:00	MD83b	MD83c	S	Crew9	Crew11	S
OHY080	IST	TZX	06:50	08:25	MD88b	MD88b	-	Crew7	Crew7	-
OHY081	TZX	IST	09:20	10:55	MD88b	MD88b	-	Crew7	Crew7	-
OHY086	IST	TZX	18:45	20:20	MD83a	MD83a	-	Crew12	Crew12	-
OHY087	TZX	IST	21:15	22:50	MD83a	MD83a	-	Crew12	Crew12	-

*R : reserve crew, S : swapping, D : delay, C : cancellation

6. CONCLUSION

In this paper both aircraft recovery and crew recovery problems have been considered. Both the aircraft recovery and crew recovery problems have been modeled as multi-commodity network flow problems where the underlying network is connection network. For each aircraft and crew a different connection network have been developed.

This paper has also considered the correlation and dependency of two problems. A new algorithm has been presented which solves aircraft recovery and crew recovery problems sequentially and in interactive manner. The subject algorithm takes into account the dependency of two problems, represents the correlation between them without integrating the two recovery problems.

With this solution algorithm, both aircraft and crew disruptions are recovered at one time. The overall algorithm does not include any integration of aircraft and crew recovery problems. Thus, the problem complexity and huge number of constraints and variables are prevented; the real time solutions are obtained in easier and more practical way. Furthermore, it is possible to consider several combinations of dependent or independent aircraft and crew unavailabilities through the algorithm.

The field of aircraft and crew recovery problems has increasing interest over the last decade and still very promising for further research opportunities. In this paper, the interaction and dependency between aircraft and crew recovery problems have been considered. However, in airline industry there are not only aircraft and crews but also other resources which are affected from the inconvenience resulting from the disruption such as passengers, gate assignments etc. The current algorithm may be extended taking into account the recovering passengers with the objective of minimizing the passengers' inconvenience.

In the literature, the solution methodologies for recovery problems are based on network flow theory and traditional optimization techniques such as set covering set partitioning and mixed integer programming. In this paper, the individual aircraft and crew recovery problems have been modeled based on multi-commodity flow network as well. As a future research, new modeling and solution techniques such as Constraint Programming may be applied rather than traditional optimization models. Although Constraint Programming approach is widely applied to scheduling problems and NP-hard problems as an alternative to linear programming based approaches (e.g. Integer programming, Mix integer programming, etc.), there is no attempt for solving airline recovery problems with Constraint Programming approach in the literature. However, constraint programming may be useful tool for solving our interactive aircraft and crew recovery problem, by enabling to capture many complex constraints in a fast and effective manner.

7. REFERENCES

[1] Abdelghany K.F., Shah S.S., Raina S. and Abdelghany A.F. *A Model for Projecting Flight Delays During Irregular Operation Conditions.* Journal of Air Transport Management, 2004, 10, 385-394.

[2] Abdelghany A.F., Ekollu G., Narasimhan R. and Abdelghany K.F. *A Proactive Crew Recovery Decision Support Tool for Commercial Airlines During Irregular Operations.* Annals of Operations Research, 2004, 127, 309-331.

[3] Abdelghany K.F., Abdelghany A.F. and Ekollu G. *An integrated decision support tool for airlines schedule recovery during irregular operations.* European Journal of Operational Research, 2008, 185, 825 - 848.

[4] Argüello M.F., Bard J.F., Yu G. *A GRASP for Aircraft Routing in Response to Grounding and Delays.* Journal of Combinatorial Optmization, 1997, 5, 211-228.

[5] Andersson T. and Varbrand P. *The Flight Perturbation Problem.* Transportation Planning and Technology, 2004, 27(2), 91-118.

[6] Bard, J.F., Yu, G. and Argüello, M.F. M.F. *Optimizing Aircraft Routings in Response to Groundings and Delays.* IEE Transactions. 2000, 33, 931-947.

[7] Bratu S. and Barnhart C. *Flight operations recovery: New approaches considering passenger recovery.* Journal of Scheduling, 2006, 9 (3), 279 – 298.

[8] Cao J.M., Kanafi A. *Real-time decision support for integration of airline flight cancellations and delays. part i: Mathematical formulation.* Transportation Planning and Technology, 1997, 20:183 - 199.

[9] Cao J.M., Kanafi A. *Real-time decision support for integration of airline flight cancellations and delays. part ii: Algorithm and computational experiments.* Transportation Planning and Technology, 1997, 20:201 - 217.

[10] Clarke, L., Hane C., Johnson E., Nemhauser G. *Maintenance and crew considerations in fleet assignment.* Transportation Science, 1996, 30, 249-260.

[11] Eggenberg N., Salani M. and Bierlaire M. *Constraint-specific recovery network for solving airline recovery problems.* Computers and Operations Research, 2010, 37, 1014-1026.

[12] Jafari N. and Zegordi S.H. (in press). *Simultaneous recovery model for aircraft and passengers.* Journal of the Franklin Institute.

[13] Jarrah A.I.Z., Yu G., Krishnamurthy N., Rakshit A. *A Decision Support Framework for Airline Flight Cancellations and Delays.* Transportation Science, 1993, 27(3), 266-280.

[14] Lettovsky L. *Airline Operations Recovery: An Optimization Approach.* PhD thesis, Georgia Institute of Technology, 1997.

[15] Lettovsky L.,Johnson E.L. and Nemhauser G.L. *Airline Crew Recovery.* Transportation Science, 2000, 34(4), 337-348.

[16] Lou S. and Yu G. *On the Airline Schedule Perturbation Problem Caused by the Ground Delay Program*. Transportation Science, 1997, 31(4), 298-311.

[17] Mathaisel D.F.X. *Decision support for airline system operations control and irregular operations*. Computers and Operations Research, 1996, 23(11), 1083 - 1098.

[18] Medard C.P and Sawhney N. *Airline Crew Scheduling From Planning to Operations*. Carmen Systems, Research and Technology Reports, 2003.

[19] Nissen R. and Haase K. *Duty-period-based Network model For Crew Rescheduling in European Airlines*, Journal of Scheduling, 2006, 9, 255-278.

[20] Rosenberger, J.M., Johnson, E.L. and Nemhauser, G.L. *Rerouting Aircraft for Airline Recovery*. Transportation Science, 2003, 37(4), 408-421.

[21] Stojkovic M., Soumis F. and Desrosiers J. *The Operational Crew Scheduling Problem*. Transportation Science, 1998, 32(3), 232- 245.

[22] Stojkovic G., Soumis F. and Desrosiers J., Solomon M.M. *An optimization model for a real-time flight scheduling problem*. Transportation Research Part A. 2002, 36, 779–788.

[23] Talluri K.T. *Swapping Applications in a Daily Airline Fleet Assignment*. Transportation Science, 1996, 30(3), 237-248.

[24] Thengvall, B.G., Bard J.F. and Yu, G. *Balancing User Preferences for Aircraft Schedule Recovery During Irregular Operations*. IIE Transactions. 2000, 32, 181-193.

[25] Teodorovic D. *A model for designing the meteorologically most reliable airline schedule*. European Journal of Operations Research, 1985, 21, 156-164.

[26] Teodorovic D. and Guberinic S. *Optimal Dispatching Strategy on an Airline Network After a Schedule Perturbation*. European Journal of Operational Research, 1984, 15, 178-181.

[27] Teodorovic D. and Stojkovic G. *Model for Operational Daily Airline Scheduling*. Transportation Planning and Techology, 1990, 14, 273- 285.

[28] Teodorovic D. and Stojkovic G. *Model to Reduce Airline Schedule Disturbances*. Journal of Transportation Engineering, 1995, 121(4), 324-331.

[29] Wei G., Yu G. and Song M. *Optimization Model and Algorithm for Crew Management During Airline Irregular Operations*. Journal of Combinatorial Optimization, 1997, 1, 305-321.

[30] Yan S. and Tu Y.P. *Multifleet Routing and Multistop Flight Scheduling for Schedule Perturbation*. European Journal of Operational Research, 1996, 103, 155-169.

[31] Yan S., Yang D-H. *A decision support framework for handling schedule perturbation*. Transportation Research B, 1996, 30(6), 405 - 419.

A STUDY OF EFFICIENCY TRENDS IN THE U.S. TL CARRIERS: 2000-2009.

Dinesh R. Pai

Penn State University, 2809 Saucon Valley Road, Center Valley, PA 18034

1-610-285-5029, drp18@psu.edu

Kenneth D. Lawrence

School of Management, New Jersey Institute of Technology, Newark, NJ 07102

1-973-596-6425, carpetfour@ yahoo.com

ABSTRACT

In this study the efficiency trends in the U.S. truckload (TL) carriers is investigated using nonparametric techniques. DEA window analysis is conducted on publicly available financial data with a representative sample of 10 publicly listed TL carriers for the period 2000-2009. The results reveal that the average overall efficiencies of the TL carriers are relatively low. They do not show a definite trend, but tend to follow the general economic conditions prevalent during the study period. Furthermore, a majority of the carriers are scale inefficient, which demonstrates that the TL carriers are not operating at their optimal size.

Key Words: DEA window analysis, Operational efficiency, Technical efficiency, Scale efficiency, Truckload.

1. INTRODUCTION

For-hire truckload (TL) carriers form an important segment of the U.S. motor carrier industry. Every year they transport almost everything from building materials, bulk commodities, household goods, heavy machinery, and motor vehicles to specialized and refrigerated commodities, catering to the needs of both individuals as well as businesses. Due to its flexibility, almost all other modes of freight transportation depend on motor carriers to provide access to air cargo, railroad, and seaport terminals. Numerous factors such as deregulation, technological innovations, economic development, and integrated supply chain management among others have catapulted the trucking business to its current status. In 2009, the for-hire TL carriers handled about 45 per cent (or $246 billion) by value of the U.S. commercial freight [40]. For these reasons alone, TL segment deserves a considerable degree of research attention.

The motor carrier industry in general, and TL carriers in particular, face several challenges, which have or threaten to erode their productivity and compromise service levels currently provided to shippers. First, increased customer expectations such as on-time windows, value-added services, time-definite services, etc., and competitive pressures have necessitated higher productivity and service quality [20] [47]. Second, higher operating costs due to escalating cost of fuel, shortage of drivers, rising insurance costs, and government regulations have affected industry profit margins. Corsi (2005) report that the average operating ratio for truckload (TL)

1

carriers increased from 97.7 to 98.4 per cent between 1987 and 2001, citing lower freight, a weak economy, and restructuring costs as reasons for higher operating ratios. Finally, the onslaught of Internet sales and the resultant disintermediation of retailers have had significant impact on the TL operations.

The carriers have adopted a variety of ways to overcome these challenges, and improve their operational efficiency, such as consolidation, enhanced communication and information systems, and value added logistical services.

While there have been numerous studies on motor carriers in general, and less-than-truckload (LTL) in particular [4] [6] [16] [21] [33], there are a few studies which specifically delve on TL carriers. For example, Stephenson and Stank (1994) analyze the profitability strategies adopted by U.S. TL carriers during 1991-1993; Taylor and Meinert (2000) discuss the quality of operations in the TL trucking industry; whereas, Snyman (2006) investigate the relationship of strategic decision processes of motor carriers in the truckload segment of the motor carrier industry and their organizational performance.   Till date, few studies have analyzed the efficiency trends and productivity changes in TL carriers.

This study presents the results of a longitudinal performance analysis of the 10 publicly owned TL carriers in the U.S. over the 10 year period 2000-2009. The objective of this research is to study the efficiency trends in a sample of publicly-owned TL carriers, and compare their performances.   Data envelopment analysis (DEA) window analysis [12], a non-parametric technique, is used to study the efficiency trends of the TL carriers or decision making units (DMUs)[2] for the period 2000-2009; DMUs are basically a collection of private firms, social organizations, departments, and groups with similar goals, and functions. The DEA window analysis results indicate that the overall long run average efficiencies of most TL carriers are substantially low and there exists a high scope for efficiency improvements.

Although, DEA technique has been utilized to analyze the motor carrier industry ([28]; Min and Joo 2006, and Zhou et al. 2008), there are no studies thus far, which have applied DEA window analysis to the U.S. motor carrier industry. This could be considered as the main contribution of this paper.

The rest of the paper is organized as follows. Section 2 gives a brief review of the U.S. motor carrier industry and Section 3 discusses the relevant literature. Section 4 introduces the DEA window analysis methodology. Section 5 delves on the data and variable selection for the study. Analysis of the results and related issues are discussed in Section 6. Conclusions and limitations of this study are discussed in section 7.

2. BACKGROUND AND LITERATURE REVIEW

This study focuses on publicly traded for-hire TL carriers. TL involves the movement of shipments of 10,000 pounds or more, and involves the transportation of a shipment from origin to destination, without making intermediate stops. Hence, the TL carriers have very little requirement for their own terminals and special equipment. As a result, the entry barriers are

2

low, thereby making this segment extremely competitive. Most of the costs incurred by TL carriers are variable costs, which includes the fuel costs and the driver services cost and accounts for about 60 per cent of the operating costs [40]. TL carriers can acquire these services either by hiring employee drivers or by engaging the services of independent contractors (owner operators), who own/lease their own trucks and/or trailers, while performing driving services on behalf of the for-hire carriers. Owner operators are self-employed truck drivers [36]. TL carriers, which require very little coordination, tend to engage owner operators. However, the high driver turnover rates among TL carriers coupled with safety and security restrictions such as driver age limitations, background checks, health standards and training have had a negative impact on their cost structure.

DEA has been extensively used as a tool for evaluating the operating efficiency of a variety of business and social organizations for the past two decades. For example, DEA has been successfully explored in measuring the operational efficiency of banks [2] [35], container ports [18] [37] [44], public transit [31] [32], and others. Only a handful of studies have applied DEA to trucking industry. For example, McMullen (1997) applied DEA to the U.S. motor carriers, whereas Min and Joo (2006), and Zhou et al. (2008) applied DEA to the third part logistics (3PL) companies. McMullen (1997) uses DEA to examine the U.S. motor carriers for a fifteen year time series (1976-90) to measure efficiency, and productivity changes in industry over time. The author concludes that the DEA efficiency scores and the Malmquist indices show little change in the efficiency and overall productivity, respectively, following deregulation. Min and Joo (2006) benchmark the operational efficiency of six U.S. based 3PL providers for the years 1999-2002 using DEA.

Unlike DEA, very few studies have utilized window analysis approach to DEA (see [3] [12] [13] [17] [19] [23] [25] [42] [46] [48]), a majority of which have been in banking, with a few others in areas such as ports, and defense. For example: Charnes et al. (1985) utilized the technique to measure the efficiency of maintenance units in the US Air Force; Itoh (2002) utilized window analysis to study the efficiency changes at major ports in Japan, whereas, Cullinane et al. (2004) applied window analysis to a sample of world's major container ports to estimate their relative production efficiency. In a way, this is the first study to apply DEA window analysis to the U.S. motor carrier industry.

Motor carrier industry has seen considerable changes in past two decades. While the economic expansion of the late 1990s served the motor carrier industry well, the early twenty-first century began with a recession primarily attributed to the "dot-com bust". The terrorist attacks of September 11, 2001 exposed the vulnerability of the U.S. transportation infrastructure causing businesses as well as government to search for means of securing supply chains and making them more resilient in case of future disruptions [45]. Recent events such as the rise in fuel prices, changes in hours of service rule (HOS), deteriorating driver shortage problems, and stricter emission standards have all added to the complexities faced by the motor carrier industry. The setting of this study differs considerably from those that existed during 1976-1990, which is the period of study covered by McMullen (1997). Furthermore, McMullen (1997) adopts an output oriented DEA model and works with operational output measures such as: ton-miles and number of shipments. Boyer and Burks (2007) state that a standard measure of physical output i.e., ton-miles is beneficial as it can be aggregated across many different types of trucking

3

operations, however, aggregated ton-miles is highly sensitive to traffic composition. According to McMullen (2005), motor carrier operational and financial data during the regulated years, 1935-1980 was comprehensive and reliable; however, the quality of data deteriorated post deregulation. The subsequent changes in the reporting instrument, Form M, and reporting requirements lead to data losing its granularity, making it difficult to evaluate and compare industry performance [9]. This study adopts an input oriented model for reasons explained in section 4, and uses financial input and output measures, which have been successfully used in the recent studies by Min and Joo (2006), and Zhou et al. (2008), among others. Homogeneity of DMUs is one of the basic requirements of DEA method i.e., DMUs transform the same type of resources, or *inputs*, into the same type of products, or *outputs*.

3. METHODOLOGY

3.1 Data Envelopment Analysis

Over the last three decades DEA has been used by researchers across disciplines as a tool to estimate performance efficiencies and productivity of decision making units (DMUs), both in public and private sector activities. DEA offers several characteristics that are unique and useful in comparison to traditional financial analysis methods like ratio analysis or regression analysis. One of the most important features of DEA is the ability to compare many financial parameters simultaneously and come up with a scalar measure of relative efficiency for a given DMU, among a set of DMUs. Another unique feature of DEA is that the type of units used for all the inputs(outputs) does not have to be the same, as long as same set of inputs (outputs) are used for all DMUs, and the measure of efficiency becomes 'units invariant' [10]. This gives a tremendous flexibility in choosing the inputs and outputs, and a convenient way to compare relative efficiencies of DMUs.

The first classical DEA model was proposed by Charnes, Cooper, and Rhodes (1978) and is called the CCR model. The CCR model assumes constant returns to scale (CRS) to assess relative productive efficiencies of decision making units (DMUs) with multiple inputs and outputs. This means that the size of a DMU does not affect the efficiency of a firm and provides global *technical efficiency* measure. Since, this assumption may not always hold good in practice, the second one, called the BCC model, was developed by Banker et al. (1984), under the assumption of variable returns to scale (VRS). The BCC model calculates local "*pure technical efficiency*". When the DMUs exhibit VRS, it means they might be increasing returns to scale (IRS), decreasing returns to scale (DRS), or both. IRS means the output increases by more than the proportional change in the input (i.e. economies of scale); whereas the DRS mean the output decreases by less than the proportional change in the input (diseconomies of scale).

Motor carriers are suffering from ever decreasing margins, due to increased customer expectations and highly competitive environment. Achieving cost efficiency through efficient use of inputs is vital for a motor carrier to staying competitive. Hence, input oriented models are more appropriate to determine the efficiency of the motor carriers. In this case, the resultant DEA score will be equal to one, if the firm is efficient and less than one, if the firm is not efficient. For details of both input oriented DEA models used in the paper, and for output

4

oriented CCR and BCC models, avid readers are referred to Charnes et al. (1978), and Banker et al. (1984).

3.2   Window Analysis

The window analysis technique was first introduced by Charles et al. (1985) and works on the principle of moving averages, which is useful to detect performance trends of a unit over time [48]. The technique treats each DMU as a different company so that data on truckload carriers in different time periods are incorporated into the model by treating them as different carriers. This treatment helps to contrast the performance of a unit in a particular period with its own performance in other periods in addition to the performance of other units. This increases the number of data points in the analysis, which can be useful when dealing with small sample sizes. If a truckload carrier is found to be DEA efficient in one year despite the window in which it is placed, it is likely to be considered strongly efficient compared to its contemporaries.

For the present study, due to homogeneity constraints, the number of truckload carriers for which data was originally available was 10, so let $n = 10$. The data was collected between the time period 2000-2009, which is 10 yearly periods, so let $P = 10$. To increase the observations to a reasonably large amount but not to over extend the period beyond the realms of plausibility a three year window was chosen, let $w = 3$. There are no straightforward methods to select a window size. Previous researchers have used window sizes of three, six, or 10 years [12] [17] [25]. The window size of three years adopted in this paper agrees with Charnes et al. (1985) and is sufficient enough to capture operational changes in truckload carriers [7]. Each DMU is the placed in the window as if it were a different DMU for each of the three years, so in the first window, Y00, Y01, and Y02. This increases the number of DMUs to 30 (i.e., $n.w = 10.3 = 30$) and the analysis is performed on these 30 truckload carriers instead of just 10 as was the case with DEA VRS model. The window is then shifted one yearly period and the analysis is performed on the next three year set, dropping the original year and adding a new year, so the window incorporates Y01, Y02, and Y03. This is run for the entire time period until window eight analysis Y07, Y08, and Y09.

DEA window analysis has several advantages when investigating the efficiency of the entire industry or individual companies [23] [42]. First, the approach lends itself to a study of trends over a specified time period. Second, it can be used to examine the stability and other properties of the efficiency scores across, as well as within windows, by adopting row views as well as column views, respectively. Third, according to Webb (2003), the approach is most meaningful when analyzing closely homogenous firms in closely homogenous markets. This means, comparing a truckload carrier in 2000 with one in 2009 could render 'relative' results meaningless due to changes in the technology employed and other structural changes in the market. However, by analyzing carriers in three-year time frames this problem is reduced.

4.   DATA VARIABLES

The sample consists of 10 publicly listed (NASDAQ) truckload carriers, whose data are available throughout the study period of 2000-2009. The data for all the 10 carriers were collected using

5

multiple sources: MergentOnline, COMPUSTAT, SEC filings, and the company websites. The sample carriers chosen for this study have their annual sales revenues in excess of $200 million for each year of the study period. The main reason for choosing this sample is the continuous availability of data for a common sample, which enables analysis of continuous evolution of relative efficiencies of those motor carriers that have survived at least 10 years or more. In order to contain the widths of the tables, the NASDAQ codes of carriers are used instead of their full names. The names of the sample carriers along with their NASDAQ codes and services they provide are shown in Table 1. A notable exclusion from the sample is TL carrier Landstar (LSTR). Though trade journals refer to LSTR as a TL carrier, the company's annual reports suggest that they also provide less-than-truckload services, which was reconfirmed by the author's telecom with a company representative.

Table 1: List of sample TL carriers along with their NASDAQ code and service offerings

Carrier	NASDAQ Code	Service(s)	Trucking	Express Delivery	Freight Services	Consulting	Warehousing	Others
Celadon Group	CLDN	TL	X		X			X
Covenant Transportation Group	CVTI	TL	X	X	X			
Heartland Express	HTLD	TL	X	X	X			
J.B. Hunt	JBHT	TL	X		X	X		
Knight Transportation	KNX	TL	X			X		X
Marten Transport	MRTN	TL						
P.A.M. Transportation Services	PTSI	TL	X		X			X
Quality Distribution	QLTY	TL	X	X	X		X	X
USA Truck	USAK	TL	X		X			X
Werner Enterprises	WERN	TL	X		X			X

The assessment of operational efficiency using DEA begins with the selection of appropriate input and output measures that can be aggregated into a composite index of overall performance standards. With an objective to look at the internal efficiencies of the 10 firms from both carriers and shippers point of view, three different inputs that represent physical resources and financial values have been selected: (i) Property, Plant, and Equipment (PPE), (ii) Salaries and Wages of employees, including fringe benefits (SW), and (iii) Operating Expenses (OPEX) (excluding the salaries and wages). PPE is considered vital to any business operations including motor carrier operations and hence considered as an input. PPE typically includes tractors, trailers, containers, warehouses, terminals, material handling equipment, IT infrastructure, etc. PPE add value by their ability to move, consolidate, and track freight, manage critical inventory, and offer other value-added services [30]. Industry experts' estimate that labor, including wages and benefits, represent about 34 percent of operating costs on average, making it the single largest component of operating costs [40]. Motor carriers, for their labor intensive operations, also hire part-time or full-time workers over and above their regular employees. For instance, many motor carriers hire owner operators, to augment their capacity. Hence, salaries and wages (SW) would be a more appropriate input instead of the number of employees. Major operating expenses (OPEX)

6

for a motor carrier (excluding salaries and wages) are fuel, transportation (purchased), vehicle maintenance, taxes, insurance premiums, and depreciation. Hence, OPEX is included as an input.

With regards to output(s), EBIT or operating income represents the most commonly used output measure in the literature. Literature on motor carriers also cites strong relationship between productivity gains and firm financial performance measures: operating ratio, net profit margin, and earnings before interest and taxes divided by total assets (EBIT/TA). Of the three measures, EBIT/TA is considered to be a better measure of firm performance because it provides some measure of return on assets. According to Golany and Roll (1989), ratio forms or percentages interfere with the notion of technical efficiency; hence EBIT is used as an output measure instead of other measures such as operating ratio and EBIT/TA. Zhou et al. (2008) observed that both revenue and profit margin were not good measures for evaluating operational efficiency. Revenues of a carrier could be inflated due to its scale and high prices, whereas favorable change in fuel prices, insurance premiums, and tax rates can increase profitability, but not necessarily the operational efficiency (e.g. equipment utilization or labor productivity). Furthermore, Sherman (1984) observed that profit measure was not a good indicator of operational efficiency of a logistical service firm. Table 2 shows the inputs and output used in this paper.

Table 2: DEA inputs and output

Inputs	
PPE	Property, Plant, and Equipment
SW	Salaries and wages, including fringe benefits
OPEX	Operating expenses, excluding salaries and wage
Output	
EBIT	Earning before Interests and Taxes

The inputs: PPE, SW, and OPEX by their very nature have positive values; however, EBIT need not be positive for all the firm years. A few firms in the sample used in this paper have negative values for EBIT. This implies that EBIT does not fulfill the condition of positiveness that DEA models require. Currently, there are no software packages, which can handle negative data. In order to avoid this situation, the property of translation invariance is applied as demonstrated by Ali and Seiford (1990), Lovell and Pastor (1995), and Pastor (1996) for additive models and the BCC model. Translation invariance is critical when the data contain zero or negative values, and must be translated prior to analysis with available software packages. Translation invariance makes it possible to change negative variables into positive ones by adding a fixed amount for all units, without giving rise to variations in the results of the analysis.

5. EMPIRICAL RESULTS AND DISCUSSION

7

Table 3 and 4 report the results for the overall efficiency (DEA-CCR), and pure technical efficiency (DEA-BCC) windows analysis, respectively. Table 6 presents the scale efficiencies, which are derived from the overall efficiencies and pure technical efficiencies (PTE). Each TL carrier is represented as if it were a different DMU at each of the three successive dates noted at the top of each column. As noted earlier, there are no straightforward methods to determine the appropriate window size. This paper has selected a window size of three years, which agrees with Charnes et al. (1985). The eight separate windows are represented as separate rows in Tables 3, 4 and 6 for each of the DMUs, i.e. truckload carriers.

5.1   Overall Efficiency

Table 3 indicates that the average overall efficiencies of the truckload carriers have been fluctuating throughout the study period which more or less coincides with the general economic conditions prevalent in the 2000s following the dot-com bubble burst and terrorist attack (2000-2002), and recent economic volatility and global mortgage crisis (2007-2008) [24] [45]. In most cases, the average overall efficiencies of truckload carriers increased in windows 4 and 5 before following a downward trend, which coincides with the economically stables years of 2005 and 2006.

The results further indicate that the truckload carriers have exhibited mean overall efficiency score of 0.491 (i.e. 49.10 per cent) during the period 2000-2009, which shows that the U.S. motor carriers have not performed well in its basic function – moving freight while incurring input waste of about 50.9 per cent. Hence, there is a high scope for improving truckload carrier's operational efficiency. It is clear from Table 3 that, HTLD was the best performer during the study period with mean overall efficiency of score of 0.87 (or 87 per cent), which is accompanied by relatively lower standard deviation of 0.19. Charnes et al. (1985) suggested that DMUs with high efficiency levels tend to demonstrate lower standard deviations compared to its peers with lower efficiency levels. While HTLD was the best truckload carrier in terms of minimizing inputs to produce the same level of outputs, on the other hand results suggest that USAK was the worst performer with 23 per cent mean overall efficiency level.

One of the advantages of the windows analysis, as exhibited by Tables 3, 4, and 6, is that it lends itself to a study of 'trends' and the examination of the 'stability' of efficiency scores across, as well as within, windows by adoption of 'row views' and 'column views', respectively. Taking HTLD for example, in Table 3, the overall efficiency of HTLD in first window is 1.00, 1.00, and 1.00. These figures correspond to the estimated relative efficiency of HTLD for 2000, 2001, and 2002, respectively. In the second window, relative efficiency estimates of HTLD of 1.00, 0.95, and 1.00 correspond respectively to 2001, 2002, and 2003, respectively. The same interpretive process can be applied to Tables 4 and 5. For each DMU, the average of the 24 DEA efficiency scores and associated standard deviations are presented in the summary measures columns denoted "Mean" and "Std. Dev.". Again, taking the HTLD example, the carrier's efficiency varies from 1.00 in 2004 to 0.84 in 2006 (i.e. window 5) by adopting a 'row view' perspective. At the same time, the efficiency of a carrier within the different windows can also vary substantially (adopting a 'column view' perspective). This variation reflects simultaneously both

8

Table 3: Overall efficiency (or CCR Efficiency) window analysis for the truckload carriers

DMU	Window	Efficiency Scores										Mean / Window	Summary Measures	
		2000	2001	2002	2003	2004	2005	2006	2007	2008	2009		Mean	Std. Dev.
CLDN	1	1.00	0.10	0.22								0.44	0.50	0.29
	2		0.10	0.22	0.32							0.21		
	3			0.22	0.33	0.54						0.36		
	4				0.32	0.53	0.85					0.57		
	5					0.53	0.85	0.73				0.71		
	6						0.85	0.73	0.43			0.67		
	7							0.86	0.53	0.29		0.56		
	8								0.30	0.17	1.00	0.49		
CVTI	1	0.32	0.17	0.32								0.27	0.29	0.31
	2		0.14	0.26	0.25							0.22		
	3			0.26	0.25	0.35						0.29		
	4				0.25	0.34	1.00					0.53		
	5					0.34	1.00	0.05				0.47		
	6						1.00	0.05	0.01			0.35		
	7							0.06	0.01	0.02		0.03		
	8								0.01	0.01	0.54	0.19		
HTLD	1	1.00	1.00	1.00								1.00	0.87	0.19
	2		1.00	0.93	1.00							0.98		
	3			0.93	1.00	1.00						0.98		
	4				0.96	0.99	0.50					0.82		
	5					0.99	0.50	0.84				0.78		
	6						0.50	0.84	0.85			0.73		
	7							1.00	1.00	0.92		0.97		
	8								0.60	0.55	1.00	0.71		
JBHT	1	0.15	0.19	0.25								0.20	0.46	0.19
	2		0.17	0.24	0.41							0.27		
	3			0.24	0.41	0.62						0.42		
	4				0.41	0.60	0.56					0.52		
	5					0.60	0.56	0.58				0.58		
	6						0.56	0.58	0.53			0.56		
	7							0.76	0.74	0.74		0.75		
	8								0.38	0.37	0.39	0.38		
KNX	1	0.79	0.85	0.90								0.85	0.74	0.18
	2		0.80	0.84	0.94							0.86		
	3			0.84	0.94	0.98						0.92		
	4				0.58	0.57	0.75					0.63		
	5					0.57	0.75	0.53				0.62		
	6						0.75	0.53	0.45			0.57		
	7							1.00	0.86	0.80		0.89		
	8								0.43	0.40	0.86	0.56		
MRTN	1	0.40	0.27	0.22								0.29	0.39	0.28
	2		0.27	0.22	0.32							0.27		
	3			0.22	0.32	0.45						0.33		
	4				0.23	0.27	1.00					0.50		
	5					0.27	1.00	0.21				0.49		
	6						1.00	0.21	0.17			0.46		
	7							0.41	0.31	0.42		0.38		
	8								0.15	0.21	0.91	0.42		
PTSI	1	0.71	0.73	0.86								0.77	0.43	0.28
	2		0.56	0.66	0.35							0.53		
	3			0.66	0.35	0.28						0.43		
	4				0.24	0.24	0.73					0.40		
	5					0.24	0.73	0.30				0.42		
	6						0.73	0.30	0.08			0.37		
	7							0.43	0.11	0.02		0.19		
	8								0.06	0.01	0.86	0.31		
QLTY	1	0.89	0.81	0.74								0.82	0.59	0.27
	2		1.00	0.92	0.53							0.82		
	3			0.93	0.53	0.81						0.76		
	4				0.28	0.48	0.28					0.35		
	5					0.48	0.28	0.73				0.50		
	6						0.28	0.73	0.45			0.49		
	7							1.00	0.57	0.69		0.75		
	8								0.32	0.39	0.01	0.24		
USAK	1	0.15	0.16	0.20								0.17	0.23	0.17
	2		0.12	0.15	0.16							0.14		
	3			0.15	0.16	0.22						0.18		
	4				0.12	0.15	0.53					0.26		
	5					0.15	0.53	0.19				0.29		
	6						0.53	0.19	0.12			0.28		
	7							0.30	0.17	0.21		0.23		
	8								0.08	0.10	0.74	0.31		
WERN	1	0.38	0.35	0.43								0.39	0.40	0.20
	2		0.29	0.34	0.38							0.34		
	3			0.34	0.38	0.43						0.38		
	4				0.30	0.33	0.90					0.51		
	5					0.33	0.90	0.31				0.51		
	6						0.90	0.31	0.23			0.48		
	7							0.51	0.34	0.33		0.39		
	8								0.17	0.17	0.35	0.23		

9

the absolute performance of a carrier over time and the relative performance of that carrier in comparison to the others in the sample.

As discussed in section 4, the overall efficiency score is a composite of both pure technical and scale efficiency scores. The comparison of scale efficiency and pure technical efficiency sheds light on the main sources of inefficiency of a DMU, may it be the technical problems associated with the quantity and combination of input and output factors or the whole operational scale. An insight into the decomposition of overall efficiency into its pure technical and scale efficiency components suggests that during the period of study, scale inefficiency (output related) dominated pure technical inefficiency (input related) in the truckload carriers.

5.2   Pure Technical Efficiency

Table 4 presents the results for the pure technical efficiency (PTE) of truckload carriers. As in the case of average overall efficiency, the PTE for the truckload carriers does not indicate a definite trend. The results indicate that the truckload carriers have exhibited mean pure technical efficiency score of approximately 0.76 (i.e. 76 per cent) during the period 2000-2009, which means that though the truckload carriers have allocated their inputs and outputs well, there is still scope for improving the efficiencies. Furthermore, the results imply that the inefficiencies of the truckload carriers are to a greater extent due to the factors concerning operational scale, and to a lesser extent due to misallocation of inputs and outputs.

The average slacks generated by the truckload carriers give an indication of misallocation of inputs and outputs. Slacks are the input excesses or the output shortfalls for a particular truckload carrier. Table 5 shows the average slacks for the individual truckload carriers as well as the average investment in PPE for the study period. The average slack for inputs PPE, SW and OPEX indicate that there is a scope for reducing the inputs for generating the same amount of output to increase the PTE of the truckload carriers. An interesting observation from Table 5 is that the two truckload carriers, WERN and JBHT, which have relatively high average investments in PPE, have higher average input slacks. Figure 1, for instance shows the number of tractors owned by the TL carriers for the study period 2000-2009.

From the figure, it can be conjectured that during years of relative economic stability as was the case during 2005-2006, the number of tractors owned by the truckload carrier shows an upward trend indicating increased freight movements, whereas; when the economy is turbulent, the number of tractors follow a downward trend. This especially is the case with TL carriers with relatively large number of tractors such as JBHT and WERN, who respond to the slowing freight markets by reducing their fleet size. For instance, WERN reduced its tractor count by 19 per cent from 9,000 (including owner-operations) at the end of 2006 to 7,250 units in 2009. Similarly, JBHT reduced its tractor count by 22 per cent from 12,127 in 2006 to 9,502 in 2009 (Standard & Poor's 2010). KNX has been a slight exception to this observation, in that, its tractor count increased consistently from 1,694 tractors in 2000 to 3,758 tractors in 2007, before stabilizing around 3,700 during 2008-2009. During economic turbulence, a truckload carrier with high investment in PPE has to contend with underutilized assets on one hand, and lower

10

Table 4: Pure technical efficiency (BCC efficiency) window analysis for the truckload carriers

DMU	Window	Efficiency Scores 2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	Mean / Window	Summary Measures Mean	Std. Dev.
CLDN	1	1.00	0.87	0.94								0.94	0.91	0.12
	2		0.96	0.97	1.00							0.98		
	3			1.00	1.00	1.00						1.00		
	4				0.75	0.93	1.00					0.89		
	5					0.93	1.00	0.81				0.91		
	6						1.00	0.81	0.48			0.76		
	7							1.00	0.91	0.76		0.89		
	8								0.87	0.79	1.00	0.89		
CVTI	1	0.43	0.46	0.45								0.45	0.56	0.20
	2		0.51	0.49	0.49							0.50		
	3			0.53	0.53	0.57						0.54		
	4				0.39	0.43	1.00					0.61		
	5					0.43	1.00	0.29				0.57		
	6						1.00	0.29	0.29			0.53		
	7							0.62	0.62	0.55		0.60		
	8								0.68	0.68	0.64	0.67		
HTLD	1	1.00	1.00	1.00								1.00	0.91	0.13
	2		1.00	0.95	1.00							0.98		
	3			1.00	1.00	1.00						1.00		
	4				1.00	1.00	0.65					0.88		
	5					1.00	0.65	0.84				0.83		
	6						0.65	0.85	0.85			0.78		
	7							1.00	1.00	0.93		0.98		
	8								0.75	0.73	1.00	0.83		
JBHT	1	0.21	0.40	1.00								0.54	0.70	0.28
	2		0.18	0.34	1.00							0.51		
	3			0.25	0.61	1.00						0.62		
	4				0.46	0.72	0.63					0.61		
	5					0.72	0.63	0.72				0.69		
	6						0.63	0.72	0.66			0.67		
	7							1.00	0.98	0.98		0.99		
	8								1.00	0.98	1.00	0.99		
KNX	1	1.00	0.99	0.98								0.99	0.80	0.20
	2		1.00	0.97	0.94							0.97		
	3			1.00	1.00	0.99						1.00		
	4				0.68	0.61	0.76					0.68		
	5					0.61	0.76	0.53				0.63		
	6						0.76	0.53	0.45			0.58		
	7							1.00	0.87	0.80		0.89		
	8								0.50	0.48	0.91	0.63		
MRTN	1	0.88	0.81	0.79								0.82	0.75	0.23
	2		0.93	0.90	0.81							0.88		
	3			1.00	0.91	0.82						0.91		
	4				0.52	0.48	1.00					0.67		
	5					0.48	1.00	0.34				0.61		
	6						1.00	0.34	0.30			0.55		
	7							0.84	0.77	0.77		0.79		
	8								0.66	0.65	0.96	0.76		
PTSI	1	1.00	0.97	1.00								0.99	0.85	0.20
	2		1.00	1.00	0.79							0.93		
	3			1.00	0.90	0.86						0.92		
	4				0.61	0.56	1.00					0.73		
	5					0.56	1.00	0.50				0.69		
	6						1.00	0.50	0.43			0.64		
	7							1.00	0.95	0.98		0.97		
	8								0.89	0.82	1.00	0.91		
QLTY	1	1.00	0.86	0.76								0.87	0.87	0.18
	2		1.00	1.00	1.00							1.00		
	3			1.00	0.96	1.00						0.99		
	4				0.42	0.55	1.00					0.66		
	5					0.55	1.00	0.78				0.78		
	6						1.00	0.78	0.52			0.77		
	7							1.00	0.94	0.79		0.91		
	8								1.00	0.87	1.00	0.96		
USAK	1	0.90	0.81	0.78								0.83	0.73	0.19
	2		0.90	0.87	0.81							0.86		
	3			0.98	0.92	0.76						0.89		
	4				0.56	0.44	0.81					0.60		
	5					0.44	0.81	0.38				0.54		
	6						0.81	0.38	0.35			0.51		
	7							0.87	0.80	0.77		0.81		
	8								0.79	0.65	0.84	0.76		
WERN	1	0.75	0.65	1.00								0.80	0.52	0.27
	2		0.29	0.54	0.78							0.54		
	3			0.38	0.51	0.58						0.49		
	4				0.30	0.34	1.00					0.55		
	5					0.34	1.00	0.34				0.56		
	6						1.00	0.34	0.24			0.53		
	7							0.59	0.36	0.34		0.43		
	8								0.20	0.20	0.45	0.28		

11

prices for their assets in the market for used equipment on the other. This overcapacity leads to a general decline in carrier rates and hence operating revenues. Shippers can have an upper hand in their negotiations with such TL carriers, which have high capacity, to gain better rates.

Table 5: Average slacks and PPE for the study period for individual truckload carriers

DMU	Excess PPE	Excess SW	Excess OPEX	Shortage EBIT	PPE
CLDN	0	24	24	12	117
CVTI	0	31	0	16	241
HTLD	1	24	1	0	190
JBHT	9	136	224	0	1095
KNX	26	17	5	0	315
MRTN	16	4	5	11	245
PTSI	2	9	0	12	176
QLTY	1	5	40	1	139
USAK	9	7	1	17	220
WERN	37	42	23	0	879
Average	10	30	32	7	

Figure 1: Number of tractors owned by TL carriers for the study period 2000-2009.

12
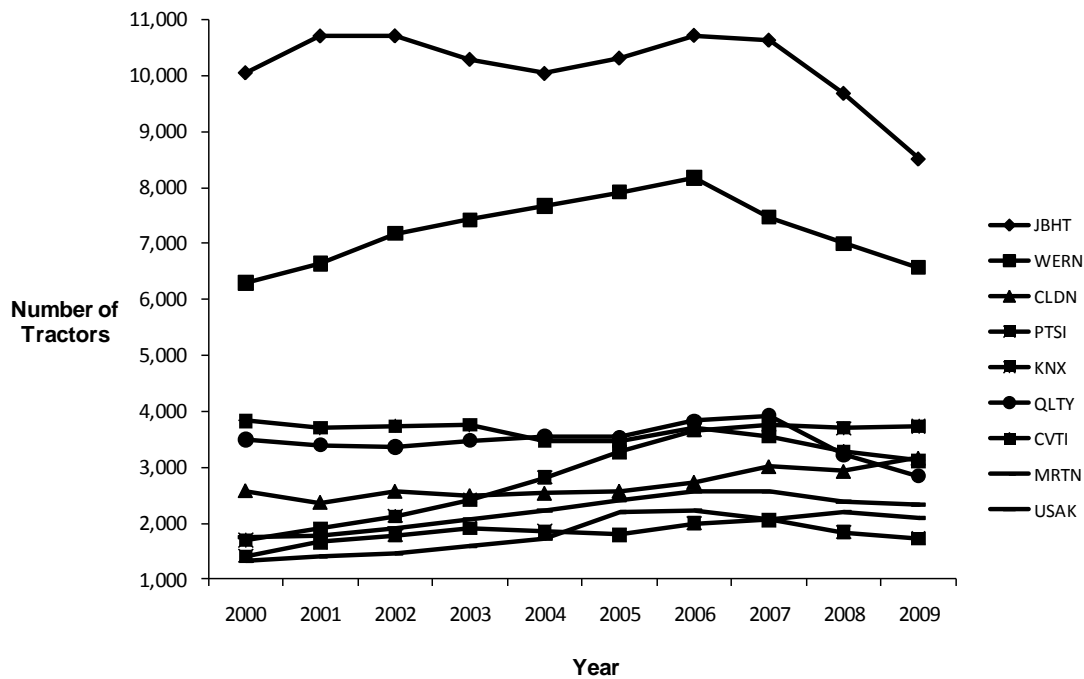
Figure 2: Average annual diesel (all types) prices in the United States*
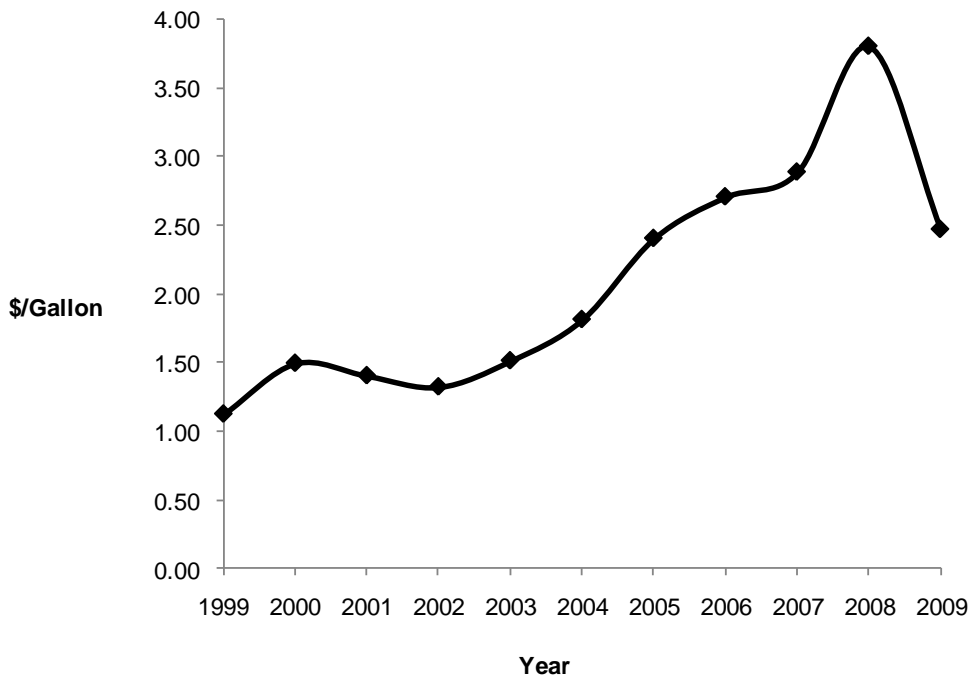
*Source: U.S. Energy Information Administration, Department of Energy.

Labor and fuel represent approximately 60 per cent of a truckload carrier's operating expenses [40]. From Figure 2, it is clear that the fuel prices have seen a general upward trend since 1999, reaching its peak in 2008, and declining sharply in 2009, which partially explains the underutilization of OPEX. A possible explanation for underutilization of SW among truckload carriers is the high annual driver turnover rates, which in many cases approach 100 per cent. The need to hire so many new drivers, and impart training to these relatively young and inexperienced drivers further drives up cost [15].

5.3   Scale Efficiency and Returns to Scale

As discussed in section 4, scale efficiency (SE) is defined as the ratio of overall efficiency score (CCR) and pure technical efficiency score (BCC), and indicates how close the production scale of a DMU (i.e. a truckload carrier) is to the most productive scale. Table 6 presents the estimated values of scale efficiency of each term and the window average for each of the sample truckload carriers. A DMU is fully scale efficient if the SE score is equal to 1.0. The results indicate that the truckload carriers have exhibited mean scale efficiency score of approximately 0.65 (i.e. 65 per cent) during the period 2000-2009 indicating divergence of the DMUs (truckload carriers) from the most productive scale size (scale inefficiency). Only two truckload carriers, HTLD and KNX, are relatively scale efficient, with SE scores of more than 0.9. While WERN, JBHT, and QLTY are not too scale inefficient with SE scores of 0.82, 0.70, and 0.70, respectively, the SE score of USAK is extremely low at 0.33.

13

Table 6: Scale efficiency window analysis for the truckload carriers

DMU	Window	Efficiency Scores										Mean / Window	Summary Measures	
		2000	2001	2002	2003	2004	2005	2006	2007	2008	2009		Mean	Std. Dev.
CLDN	1	1.00	0.12	0.23								0.45	0.55	0.30
	2		0.11	0.22	0.32							0.22		
	3			0.22	0.33	0.54						0.36		
	4				0.43	0.57	0.85					0.62		
	5					0.57	0.85	0.90				0.78		
	6						0.85	0.90	0.89			0.88		
	7							0.86	0.58	0.39		0.61		
	8								0.35	0.22	1.00	0.52		
CVTI	1	0.74	0.36	0.72								0.61	0.47	0.35
	2		0.27	0.54	0.52							0.44		
	3			0.48	0.47	0.61						0.52		
	4				0.64	0.81	1.00					0.81		
	5					0.81	1.00	0.19				0.66		
	6						1.00	0.19	0.02			0.40		
	7							0.10	0.01	0.03		0.05		
	8								0.01	0.01	0.85	0.29		
HTLD	1	1.00	1.00	1.00								1.00	0.95	0.09
	2		1.00	0.98	1.00							0.99		
	3			0.93	1.00	1.00						0.98		
	4				0.97	0.99	0.77					0.91		
	5					0.99	0.77	1.00				0.92		
	6						0.77	0.99	1.00			0.92		
	7							1.00	1.00	0.99		1.00		
	8								0.80	0.75	1.00	0.85		
JBHT	1	0.72	0.46	0.25								0.48	0.70	0.21
	2		0.98	0.69	0.41							0.69		
	3			0.93	0.68	0.62						0.75		
	4				0.88	0.83	0.88					0.86		
	5					0.83	0.88	0.81				0.84		
	6						0.88	0.81	0.81			0.84		
	7							0.76	0.76	0.76		0.76		
	8								0.38	0.38	0.39	0.39		
KNX	1	0.79	0.86	0.92								0.86	0.93	0.07
	2		0.80	0.87	1.00							0.89		
	3			0.84	0.94	1.00						0.93		
	4				0.85	0.95	0.98					0.92		
	5					0.95	0.98	1.00				0.98		
	6						0.98	1.00	0.98			0.99		
	7							1.00	0.99	0.99		1.00		
	8								0.85	0.84	0.95	0.88		
MRTN	1	0.45	0.33	0.28								0.35	0.52	0.25
	2		0.29	0.24	0.40							0.31		
	3			0.22	0.36	0.55						0.37		
	4				0.44	0.57	1.00					0.67		
	5					0.57	1.00	0.62				0.73		
	6						1.00	0.62	0.55			0.72		
	7							0.48	0.40	0.55		0.48		
	8								0.23	0.32	0.94	0.50		
PTSI	1	0.71	0.75	0.86								0.77	0.49	0.26
	2		0.56	0.66	0.44							0.56		
	3			0.66	0.39	0.32						0.46		
	4				0.40	0.42	0.73					0.52		
	5					0.42	0.73	0.60				0.58		
	6						0.73	0.60	0.19			0.51		
	7							0.43	0.12	0.02		0.19		
	8								0.07	0.01	0.86	0.31		
QLTY	1	0.89	0.95	0.98								0.94	0.70	0.29
	2		1.00	0.92	0.53							0.82		
	3			0.93	0.56	0.81						0.77		
	4				0.66	0.87	0.28					0.61		
	5					0.87	0.28	0.93				0.70		
	6						0.28	0.93	0.87			0.69		
	7							1.00	0.61	0.87		0.83		
	8								0.32	0.45	0.01	0.26		
USAK	1	0.17	0.20	0.25								0.21	0.33	0.21
	2		0.14	0.17	0.19							0.17		
	3			0.15	0.17	0.29						0.20		
	4				0.21	0.35	0.65					0.40		
	5					0.35	0.65	0.50				0.50		
	6						0.65	0.50	0.34			0.50		
	7							0.35	0.21	0.28		0.28		
	8								0.10	0.16	0.88	0.38		
WERN	1	0.51	0.54	0.43								0.49	0.82	0.18
	2		1.00	0.63	0.48							0.71		
	3			0.89	0.75	0.74						0.79		
	4				1.00	0.96	0.90					0.95		
	5					0.96	0.90	0.91				0.92		
	6						0.90	0.91	0.98			0.93		
	7							0.86	0.96	0.98		0.93		
	8								0.87	0.85	0.78	0.83		

14

6. CONCLUSIONS

This study uses a combination of nonparametric techniques to examine the efficiency trends in the U.S. TL carriers for the period 2000-2009. The DEA window analysis methodology discussed in this study has allowed us to distinguish between three types of efficiency: technical or overall, pure technical, and scale efficiencies. This method has the potential to provide crucial information about a TL carrier's operational conditions and management performance for the benefit of regulators, managers and investors. Moreover, the method requires only data on the quantity of inputs and outputs; no price data are necessary. This is especially appealing in the analysis of motor carriers' because of the difficulties inherent in defining and measuring the prices of carriers' inputs and outputs [28].

The DEA window analysis revealed that the overall long run efficiency trends of the TL carriers fluctuate over time. This leads to the conclusion that the fluctuation may have been due to the prevailing economic environment. Furthermore, there exists substantial inefficiency in most of the TL carriers for the period 2000-2009. The results were disaggregated to test for levels of pure technical efficiency and scale efficiency. In line with Ying (1990a), and Harmatuck (1992), the evidence contradicted the generally held notion of constant returns to scale. Most TL carriers' exhibit increasing returns to scale (IRS), indicating that production scale may be one of the main source of inefficiency. However, as noted earlier, the estimated scale efficiencies may reflect insufficient production due to inefficient use of inputs rather than a non-optimal size of the TL carriers. The average pure technical efficiency score of 0.76 indicate that though the TL carriers have been reasonable in allocating their inputs and outputs, there is still scope for improving the efficiencies. Another interesting observation from this study is the distribution of slacks. TL carriers' like JBHT and WERN, which have relatively higher investment in PPE, showed higher slacks or underutilization of resources, especially PPE and OPEX. In general, most TL carriers faced underutilization of SW, due to higher driver turnover, shortage of skilled drivers, and increased cost of recruiting and training new drivers.

This study has several practical implications for shippers. First, overcapacity among TL carriers generally leads to decline in rates, which a shipper should leverage to extract lower rates from carriers. Second, a shipper should contract TL carriers that have better technical efficiency, which is an indicator of improved technological capability in addition to higher productivity given a technology. Finally, a shipper should ensure that the TL carrier they contract have lower OPEX slack, especially, the TL carriers should have lower driver turnover rate to ensure safe and reliable services.

This study has several limitations, a few of which could form the basis for future research. First, this study has evaluated the performances of only the publicly listed TL carriers. However, there are many privately held carriers such as Wal-Mart, Sysco, and Coca-Cola Co., etc., which have not been considered in this study. Including these carriers may significantly impact the results of the study. Moreover, the sample size of 10 TL carriers used in this study is relatively small. Second, the results of the study could vary depending on the choice of the inputs and outputs. This paper considers only financial variables such as operating expenses (excluding salaries and wages), property, plant and machinery (PPE), and EBIT. However, non-financial inputs such as average number of company operated tractors per year, average number of trailers per year, total

15

loads, etc., and output such as average revenues per total mile, could be considered. Third, this study is confined to the analysis of TL carriers for a decade; whereas, many previous studies have considered a time frame of more than a decade [3] [12] [42] [46]. An advantage of considering a longer period is that it aids better analysis of efficiency trends and productivity changes. However, as noted before, one of the reasons for selecting a period of a decade is the availability of continuous data for the sample of TL carriers. Finally, the study utilizes a DEA approach, which is inherently deterministic, with no allowance for measurement or specification error.

REFERENCES

[1] Ali, A. and L. M. Seiford (1990). "Translation invariance in data envelopment analysis," *Operations Research Letters*, 9 (6), pp. 403-405.

[2] Angelidis, D. and K. Lyroudi (2006). "Efficiency in the Italian Banking Industry: Data Envelopment Analysis and Neural Networks," *International Research Journal of Finance and Economics*, 5, pp. 155-165.

[3] Asmild, M., J. Paradi, V. Aggarwall and C. Schaffnit (2004). Combining DEA window analysis with the Malmquist index approach in a study of the Canadian banking industry. *Journal of Productivity Analysis*, 21, pp. 67–89.

[4] Baker, J. (1991). "Emergent pricing structures in LTL transportation," *Journal of Business Logistics*, 12, pp. 191–202.

[5] Banker, R. D., A. Charnes, and W. W. Cooper, (1984). "Some models for estimating technical and scale inefficiencies in data envelopment analysis," *Management Science,* 30 (9), pp. 1078-1092.

[6] Barnhart, C., and D. Kim (1995). "Routing models and solution procedures for regional Less-than-Truckload operations," *Annals of Operations Research*, 61, pp. 67–90.

[7] Boussofiane, A., R. Dyson and E. Thanassoulis (1991). "Applied data envelopment analysis," *European Journal of Operational Research*, 52(1), pp. 1-15.

[8] Boyer, K. and S. Burks (2007). "Stuck in the Slow Lane: Traffic Composition and the Measurement of Labor Productivity in the U.S. Trucking Industry," *IZA Discussion Papers 2576*, Institute for the Study of Labor (IZA), Bonn, Germany.

[9] Burks, S., F. Guy and B. Maxwell (2004). "Shifting Gears In The Corner Office: Deregulation And The Earnings Of Trucking Executives," *Research in Transportation Economics*, Elsevier, 10(1), pp. 137-164.

[10] Charnes, A., W. W. Cooper, A.Y. Lewin and L.M. Seiford (eds.) (1994). Dam Envelopment

16

Analysis: Theory, Methodology, and Applications, Boston/Dordrecht/London: Kluwer Academic Publishers.

[11] Charnes, A., W. W. Cooper and E. Rhodes (1978). "Measuring the Efficiency of Decision-Making Units," European Journal of Operational Research, 2, pp. 429-444.

[12] Charnes, A., C. Clark, W. Cooper and B. Golany (1985). "A Developmental Study of Data Envelopment Analysis in Measuring the Efficiency of Maintenance Units in the U.S. Air Forces," *Annals of Operation Research*, 2 (1), pp. 95-112.

[13] Chung, S., A. Lee, H. Kang and C. Lai (2008). "A DEA window analysis on the product family mix selection for a semiconductor fabricator," Expert Systems with Applications, 35, pp. 379–388.

[14] Cooper, W. W., L. Seiford and K. Tone (2000). "Data Envelopment Analysis," Dordrecht: Kluwer.

[15] Corsi, T M. (2005). "The truckload carrier industry segment. In Dale Belman and Chelsea White III (Eds.) Trucking in the age of information, Ashgate, Hampshire, England. pp. 35-36.

[16] Corsi, T., C. Grimm, K. Smith and R. Smith (1992). "The effects of LTL motor carrier size on strategy and performance," Logistics and Transportation Review, 28, pp. 129-146.

[17] Cullinane, K., D. Song, P. Ji and T. Wang (2004). "An Application of DEA Windows Analysis to Container Port Production," *Review of Network Economics*, 3 (2), pp. 184–206.

[18] Cullinane, K. and Wang, T. (2006). "The Efficiency of European Container Terminals: A Cross-sectional Data Envelopment Analysis," *International Journal of Logistics: Research and Applications*, 9(1), pp. 19-31.

[19] Day, D., A. Lewin, H. Li and R. Salazar (1994). "Strategic Leaders in the U.S. Brewing Industry: A Longitudinal Analysis of Outliers." In Data Envelopment Analysis: Theory, Methodology, and Applications. A. Charnes, W.W. Cooper, A. Lewin, and L.M. Sieford. eds. Boston: Kluwer Academic Publishers, pp. 211-234.

[20] Dobie, K. (2005). "The Core Shipper Concept: A Proactive Strategy for Motor Freight Carriers," *Transportation Journal*, 44 (2), pp. 37-53.

[21] Feither, J., T. Corsi and C. Grimm (1998). "Strategic and performance changes among LTL motor carriers: 1976-1993," *Transportation Journal*, 37, pp. 5-13.

[22] Golany, B. and Y. Roll (1989). "An application procedure for DEA," *Omega*, 17 (3), pp. 237-250.

[23] Hartman, T. and J. Storbeck (1996), "Input congestion in loan operations", *International Journal of Production Economics*, 46-47(3), pp. 413-21.

17

[24] Iammartino, R. (2009). The Impact of Changing Economic Conditions on United States Army Recruiting Trends. *QMSS Thesis: Columbia University*, Quantitative Methods of the Social Sciences, pp. 18.

[25] Itoh, H. (2002). "Efficiency changes at major container ports in Japan: A window application of DEA," *Rurds*, 14, pp. 133–152.

[26] Lovell, C. and J. Pastor (1995). "Units invariant and translation invariant DEA models," *Operations Research Letters*, 18, pp. 147–151.

[27] McMullen, B. S. (1987). "The Impact of Regulatory Reform on U.S. Motor Carrier Costs." *Journal of Transport Economics and Policy,* 26 (3), pp. 307-319.

[28] McMullen, B. S. (1997). "The Impact of Deregulation on Technical Efficiency: The U.S. Motor Carrier Industry," *Technical Report # TNW 97-10*, Oregon State University.

[29] McMullen, B.S. (2005), "The Evolution of the U.S. Motor Carrier Industry", in Belman, D. and White, C. III (Ed.), *Trucking in the age of information*, Ashgate Publishers, Hampshire, England, pp. 1-19.

[30] Min, H and S. Joo (2006). "Benchmarking the operational efficiency of third party logistics providers using data envelopment analysis," *Supply Chain Management: An International Journal*, 11 (3), pp. 259-265.

[31] Nolan, J. (1995). "Efficiency of public agencies: A study of urban bus transit measurement," of public agencies: A Study of Urban Department of Economics, U.C. Irvine. November.

[32] Obeng, K. (1994) "The Economic Cost of Subsidy-Induces Technical Inefficiency," *International Journal of Transport Economics*, 21 (1), pp. 3-20.

[33] Özkaya, E., P. Keskinocak, V. Joseph and R. Weight (2010). "Estimating and Benchmarking Less-Than-Truckload Market Rates," *Transportation Research Part E: Logistics and Transportation Review*, 46 (5), pp. 667-682.

[34] Pastor, J. T. (1996). "Translation invariance in data envelopment analysis: a generalization," *Annals of Operations Research*, 66 (2), pp. 91-102.

[35] Pastor J. M., F. Perez and J. Quesada (1997). "Efficiency Analysis in Banking Firms: An International Comparison," *European Journal of Operational Research*, 98, pp. 395-407.

[36] Peoples, J. and M. Peteraf (1999). "The Effects of Regulatory Reform on Company Drivers and Owner-Operators in the For-hire and Private Sectors," *Transportation Journal,* 33, pp. 5-17.

[37] Roll, Y. and Y. Hayuth (1993). "Port Performance Comparison Applying Data Envelopment Analysis (DEA)", *Maritime Policy and Management*, 20 (2), 153-161.

18

[38] Sherman, H.D. (1984). "Improving the productivity of service businesses," *Sloan Management Review,* 125, pp. 11-23.

[39] Snyman, J. (2006). "Strategic Decision Processes and Firm Performance Among Truckload Motor Carriers," *Journal of American Academy of Business*, 8 (1), pp. 265-270.

[40] Standard & Poor's (2010). "Industry Survey, Transportation: Commercial," *Standard & Poor's*, pp. 10, New York, NY.

[41] Stephenson, F. and T. Stank (1994). "Truckload motor carrier profitability strategy," *Transportation Journal*, 34 (2), pp. 5-17.

[42] Sufian, F. and Majid, M-Z A. (2007). "Singapore banking efficiency and its relation to stock returns: A DEA window analysis approach," *International Journal of Business Studies*, 15(1), pp. 83-106.

[43] Taylor, G. and Meinert, T. (2000). "Improving the quality of operations in truckload trucking," *IIE Transactions*, 32, pp. 551-562.

[44] Tongzon, J. (2001). "Efficiency Measurement of Selected Australian and other International Ports using Data Envelopment Analysis," *Transportation Research, Part A*, 113-128.

[45] Voss, M. D., T.J. Page Jr, S.B. Keller and J. Ozment (2006). "Determining Important Carrier Attributes: A Fresh Perspective Using the Theory of Reasoned Action," *Transportation Journal*, 45 (3), pp.7-19.

[46] Webb, R. (2003). "Levels of efficiency in UK retail banks: a DEA window analysis," *International Journal of the Economics of Business*, 10(3), pp. 305-22.

[47] Wisner, J.D. and I.A. Lewis, (1996). "Quality Improvement Programs in the Motor Carrier Industry", *Transportation Journal*, 36(2), pp. 26-34.

[48] Yue, P. (1992). "Data envelopment analysis and commercial bank performance: A primer with applications to Missouri banks," *Federal Reserve Bank of St. Louis Review*, 74(1), pp. 31-45.

[49] Zhou, G., H. Min, C. Xu and Z. Cao, (2008). "Evaluating the comparative efficiency of Chinese third-party logistics providers using data envelopment analysis," *International Journal of Physical Distribution & Logistics Management*, 38 (4), pp. 262-279.

19

MINIMIZING MAKESPAN AND WORKSTATION UTILIZATION IN A FLOWSHOP WITH OPERATIONS FLEXIBILITY

Alex J. Ruiz-Torres, Departamento de Gerencia, Facultad de Administración de Empresas, Universidad de Puerto Rico – Rio Piedras, San Juan, PR 00931. alex.ruiztorres@uprrp.edu (787) 764-0000

Johnny C. Ho, Department of Management and Marketing, Turner College of Business and Computer Science, Columbus State University, Columbus, GA 31907. ho_johnny@colstate.edu (706) 565-4125

Jose H. Ablanedo-Rosas, Department of Information and Decision Sciences, The University of Texas at El Paso, 500 West University Ave., El Paso TX 79968. jablanedorosas2@utep.edu (915) 747-6041

ABSTRACT

Operations flexibility refers to production environments where there are more tasks than workstations and any task can be assigned to any workstation. This article investigates the advantages of operations flexibility in a flowshop when the objectives are to minimize the makespan and workstation utilization. The proposed heuristic framework consists of two phases, allocation of operations to workstations and generation of job sequencing. Moreover, an improvement procedure based on simulated annealing is implemented for this problem. Experiments are conducted to: analyze the impact of operations flexibility; demonstrate that schedules that are makespan optimal may not be machine utilization optimal and vice versa; and analyze the performance of the proposed heuristics.

Keywords: scheduling; flowshop; makespan; workstation utilization; operations flexibility.

1. INTRODUCTION

This paper investigates the benefits of operations flexibility in a flowshop when the goals are to minimize the completion time of all the jobs and the utilization of the workstations. Operations flexibility in a flowshop environment refers to the ability of the shop to organize the production tasks along the production resources. As in a regular flowshop, there is a fixed sequence of operations for the products; however, with this flexibility the "location" of each operation within the flow is not fixed. In the flowshop problem, the production resources are the machines or workstations; thus, operation flexibility relates to the assignment of the production tasks to the workstations, with the assumption that there are more tasks than workstations and that the workstations can be assigned to perform any of the tasks. This type of flexibility is typically available during the organization of production systems that employ people (manual labor) and small tools as the primary components. The concept and applications of operation to workstation flexibility, also called task redistribution, were presented by Burdett and Kozan [1] for the case of a mixed model assembly line; while Ruiz-Torres et al. [21] considered operation and workstation flexibility in flowshop scheduling, proposing a lower bound procedure and solution approaches to minimize the number of tardy jobs. There are similarities between the flowshop problem with operations flexibility and the assembly line balancing problem in that both problems deal with allocating operations/tasks to workstations. However, in our problem the allocation is restricted by the order of operations, whereas it is restricted by the precedence graph in assembling line balancing. Moreover, job sequencing is the primary issue in flowshop scheduling. On the other hand, due to mass-production of a homogeneous product, job sequencing is not a major concern in line balancing problems.

The proposed reconfiguration of the traditional flow shop problem comes from observations in industries where modifications to the allocation of work to the production process are performed based on the jobs at hand. The observed ability to allocate operations across production stages is formally analyzed in this

paper when the criteria of relevance are the makespan and workstation utilization. We describe the problem next. There are n jobs to be scheduled, $N = \{1,\ldots, n\}$ and each job must undergo t operations, $T = \{1,\ldots, t\}$. Jobs are available at time 0. There are m available workstations, $M = \{1,\ldots, m\}$. Each operation must be allocated in sequence to one of m workstations. Jobs flow through m workstations in sequence, and each job must follow the identical sequence in all the workstations, i.e., permutation flowshop. Each job j has a processing time for operation g, p_{jg}. Once a job is started at a workstation, all operations assigned to that workstation must be completed without interruption. Let f_k be the number of operations assigned to workstation k, where each workstation must be assigned at least one operation ($f_k \geq 1$); thus, if $f_1 = 2, f_2 = 1$, and $f_3 = 3$ then operations 1 and 2 are assigned to workstation 1, operation 3 is assigned to workstation 2 and operations 4 to 6 are assigned to workstation 3. Let F represent the set of assignments; $F = \{f_1, f_2, \ldots, f_m\}$. As in Burdett and Kozan [1], the redistribution of tasks can be constrained by factors, such as the workstation space. Hence, we limit the number of operations that can be assigned to any workstation; let f_{max} be this number and $f_{max} \geq f_k \ \forall \ k \in M$.

Let q_{jk} represent the sum of processing time for the operations of job j in workstation k, where $q_{jk} = \sum_{g = \varsigma(k-1)+1 \ldots \varsigma(k)} p_{jg}$ with $\varsigma(0) = 0$ and $\varsigma(k) = \sum_{g = 1 \ldots k} f_g$. For the purpose of this analysis, we assume the number of operations is an integer multiple of the number of workstations, and let $r \ (= t/m)$ be this multiple. In the case that no operations flexibility is allowed, each workstation will be assigned r operations, $f_k = r \ \forall \ k \in M$. Let C_j be the completion time of job j in the last machine (i.e., machine m) and $C_{max} = \boldsymbol{max}_{j \in N} (C_j)$, the maximum completion time or makespan. Let MC_k be the maximum completion time in machine k (i.e., the completion time of the last job on machine k) and obviously $MC_m = C_{max}$. Based on the definition used by Fondrevelle et al. [5], machine utilization is the sum of the machine completion times: $MU = \sum_{k \in M} MC_k$.

The data for a sample problem with five jobs, three workstations, and six operations are provided in Table 1. Figure 1 presents three schedules for the problem, noting that the labels include both the job code and operations number. In the first schedule (S_1), the operation assignment is set to the baseline (no flexibility or $f_k = 2 \ \forall \ k \in M$) and the job sequence results in a makespan of 34 time units and a machine utilization of 81 time units. We note that this sequence provides the optimal makespan of 34 time units assuming no operations flexibility (found by full enumeration). The second schedule (S_2) assumes operation flexibility and has an operation to workstation assignment of $f_1 = 1, f_2 = 3$ and $f_3 = 2$. The presented schedule has a makespan of 31 time units and a MU of 71 time units; thus, S_2 dominates S_1 in both criteria demonstrating the benefit of considering operations flexibility. It is also noted that S_2 is the optimal makespan schedule (again found by full enumeration). The third schedule (S_3) has an operation to workstation assignment of $f_1 = 1, f_2 = 2$ and $f_3 = 3$ and yields a makespan and machine utilization of 38 and 67 time units, respectively. Schedule S_3 is an efficient schedule and the optimal solution when only MU is considered. These schedules demonstrate a tradeoff between the two criteria and the effect of changes in work position assignment and job sequences.

TABLE 1: SAMPLE PROBLEM PROCESSING TIMES

Job	p_{j1}	p_{j2}	p_{j3}	p_{j4}	p_{j5}	p_{j6}
A	2	1	2	2	1	2
B	4	3	1	4	4	4
C	3	2	1	1	3	3
D	2	2	2	3	1	3
E	1	1	1	1	1	2

It should be noted that the traditional Permutation Flowshop Scheduling Problem (PFSP) with makespan minimization criterion has been the subject of a great deal of research since the seminal work of Johnson

[10], and that interest in the problem continues as evidenced by recent work (e.g., Laha and Sarin [13], Li et al. [14], Rad et al. [19], Naderi and Ruiz [17], and Vallada and Ruiz [24]). While Johnson [10] developed an algorithm that solved the two-machine flowshop problem optimally with respect to minimizing makespan, the general *m*-machine flowshop scheduling problem has been shown to be NP-hard (Coffman [2]; Garey et al. [7]). As a result, researchers have focused on the design of efficient heuristics and the identification of special flowshops which are solvable in polynomial time. The NEH heuristic (Nawaz et al. [18]) has been regarded as the most effective (Ruiz and Maroto [20]; Kalczynski and Kamburowski [11]) and multiple studies have analyzed the performance of diverse approaches to the problem (e.g., Rad et al. [19]). Reviews on flowshop scheduling include Ruiz and Maroto [20] and Gupta and Stafford [8]. The sum of machine completion times has also received attention in the literature, including Fondrevelle et al. [5], Framinan et al. [6], Sridhar and Rajendran [22], and Ho and Chang [9].
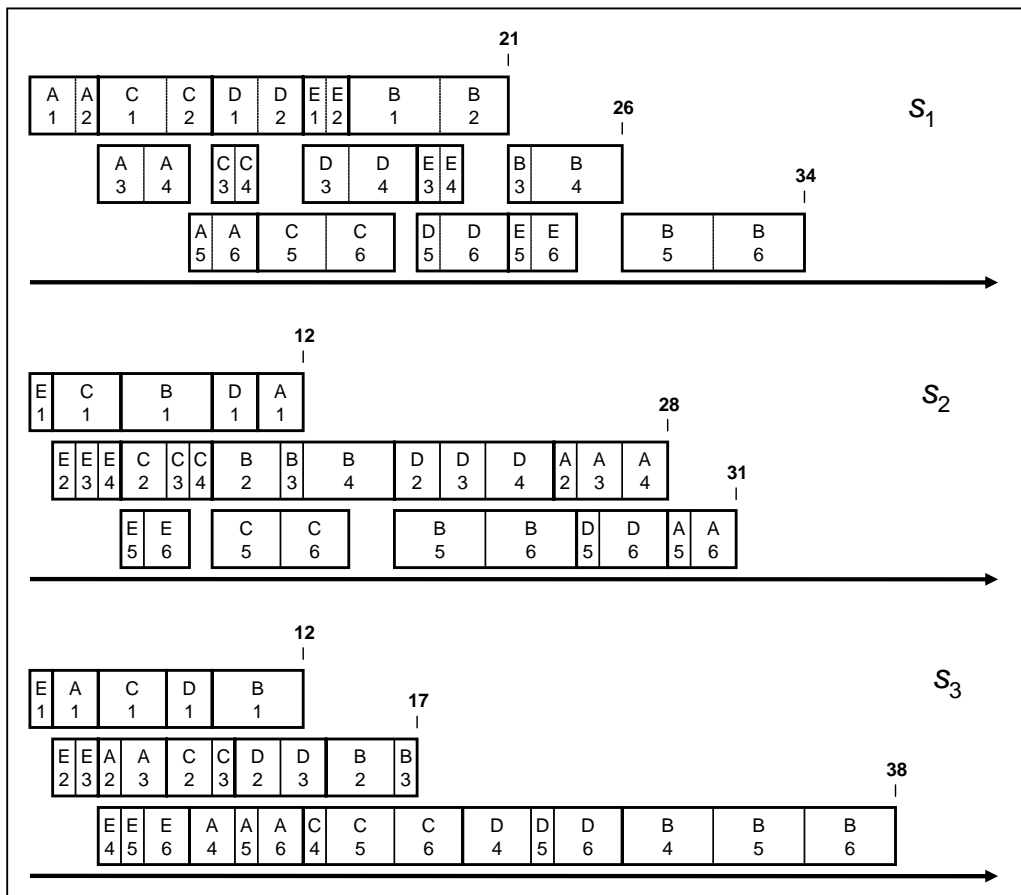
FIGURE 1: THREE SAMPLE SCHEDULES

The remaining of this paper is organized as follows. Section 2 describes the algorithms used to generate solutions for the described problem. Section 3 presents the experimental framework and discusses the results from the experiments. Finally, Section 4 provides a brief summary of this paper.

2. ALGORITHMS

This section describes the approaches used to generate solutions for the PSFP with operations flexibility. The solution approach combines two sets of heuristics: the first group of heuristics is aimed to determine the allocation of *operation to workstation* assignment; and, the second collection of heuristics is used to

generate the job sequencing. Let *Y* be the criterion of interest for a particular problem instance, that is, $Y = C_{max}$ or *MU*.

2.1 Operation allocation to workstations

This section describes two procedures which will be used to assign the operations to the workcenters. The first procedure, called *Balanced Loading* (*B*) proposed by Ruiz-Torres et al. [21], is similar to the Multifit procedure used for makespan minimization in parallel-machine scheduling (Coffman et al. [3]). As in Multifit, the procedure iteratively attempts to fit workload (in this case the operations) to the workstations up to a target value. If for an iteration of the process all the operations are assigned to the set of workstations (a feasible assignment is found), then the target is reduced; while in the opposite case (one or more operations fail to be assigned to the workstations, resulting in an infeasible solution), the target is increased. When the target is modified, we find out that, based on our pilot experiments, a maximum of 25 non-improvement loops are sufficient to achieve desirable results in both performance criteria.

Assignment of Operations by Balanced Loading (*B*)

Step 0.   Let $P_g = \sum_{j \in N} p_{jg} \ \forall \ g \in T$, *Loops* = 0, *Maxloops* = 25, $L_{max} = \infty$, $\upsilon = \sum_{g \in T} P_g$, $\lambda = \textbf{\textit{max}}\{\textbf{\textit{max}}\{P_g, g \in T\}, \upsilon/m\}$, and $\tau = (\upsilon + \lambda)/2$.

Step 1.   Let $h = 0, g = 1, L_k = 0 \ \forall \ k \in M, f^*_k = 0 \ \forall \ k \in M$ and $P_{left} = \sum_{g \in T} P_g$.

Step 2.   Let $h = h + 1$.

Step 3.   If $h > m$ or $P_{left} > \tau \ (m - h + 1)$, then $\tau = (\upsilon + \tau)/2$, *Loops* = *Loops* + 1, and go to Step 8.

Step 4.   If $L_h + P_g > \tau$ or $f^*_h + 1 > f_{max}$, then return to Step 2; else $f^*_h = f^*_h + 1$, $P_{left} = P_{left} - P_g$, and $L_h = L_h + P_g$.

Step 5.   If $g < t$, then $g = g + 1$ and return to Step 4.

Step 6.   Let $\upsilon = \textbf{\textit{max}}\{L_h, h \in M\}$, *Loops* = *Loops* + 1, and $\tau = (\upsilon + \lambda)/2$. If $L_{max} > \upsilon$, then *Loops* = 0 and $L_{max} = \upsilon$.

Step 7.   If $f^*_k \geq 1 \ \forall \ k \in M$, then $f_k = f^*_k \ \forall \ k \in M$.

Step 8.   If *Loops* < *Maxloops*, then return to Step 1; else, Stop.

Steps 0 and 1 define and initialize control parameters and calculate the total workload of the problem (υ). L_k is the total workload of workstation *k* and the parameter τ represents the target loading for all the workstations. Step 2 increments the counter for the current workstation (*h*). Step 3 performs two tests to identify if the current assignment will be infeasible, that is, it needs more than *m* workstations or the remaining load is larger than the target loading multiplied by the number of currently unassigned workstations. If any test fails, then the parameter *Loops* (which tracks the number of non-improvement loops performed) is increased by 1, and the process goes to Step 8. Step 4 also tests two conditions (the new load is above the target and this assignment exceeds the maximum number of operations per workstation) to determine if the current workstation can receive the current operation. If any condition fails, then the process returns to Step 2, where the current workstation counter is updated. If both conditions are satisfied, then the current workstation is assigned an additional operation and the remaining load parameter is reduced. Step 5 increases the current operation counter if this is less than the total number of operations, and then returns to Step 4 (to check if the assignment works). Step 6 is reached if all operations have been assigned and the workstation counter has not exceeded the number of workstations. If the maximum load in the workstations has decreased, the number of loops is reset to 0. Step 7 verifies the current assignment has at least one operation per workstation; and if so, Step 7 updates the operations to workstations assignment. Step 8 checks for a maximum number of non-improvement loops and stops the process when 25 non-improvement loops have been completed.

The second procedure, called *Uneven Loading* (*U*), iteratively tests uneven assignments of the workstation loads. The motivation behind this strategy is to achieve higher machine utilizations. Let $P_g = $

$\sum_{j \in N} p_{jg}$ and suppose that there are three workstations and six operations with $P_1 = P_2 = \ldots = P_6$. Then, an example of unbalanced assignment is: operations 1-3 are assigned to workstation 1, operations 4-5 to workstation 2, and operation 6 to workstation 3, resulting in a high load in the first workstation. Let Γ be the set of workstations assignment sets (F) generated by this procedure.

Procedure U iteratively modifies the targeted unbalance parameter, ϕ, representing the ratio of the workload on the *high load* workstation to that on the *low load* workstation, where $\phi \geq 1$. For example, if $\phi = 4$, the objective is to assign operations to the *high load* workstation such that the combined processing time (workload) of those operations is four times the combined workload of the operations assigned to the *low load* workstation. Clearly as ϕ increases, the workload imbalance increases; whereas $\phi = 1$ indicates that the goal is balanced load levels across all workstations. A second input to U is the "direction" of the imbalance, referred to as φ. If $\varphi = 1$, then the first workstation will be the one receiving the largest load. On the other hand, if $\varphi = m$, workstation m will be the one receiving the highest load. Procedure U starts assigning operations to the high load workstation first until the target is reached or the maximum number of operations is assigned to a workstation. Based on pilot experiments, ϕ is set to 4 in our implementation. The variable L_h is used in the process to determine the total workload assigned to a workstation.

Assignment of Operations by Uneven Loading (U)
Step 0.   Let $\varphi = 1$ and $\phi = 4$. Let $\Gamma = \varnothing$.
Step 1.   Let $f_h = 1 \; \forall \; h \in M$, $\upsilon = (\phi - 1)/(m - 1)$, $\sigma = m + \sum_{x = 1 \ldots m - 1} x \cdot \upsilon$, $P^* = \sum_{g \in T} P_g$, $g' = t - m$, and $h' = m$.
Step 2.   If $\varphi = 1$, then $h = 1$ and $a = 1$; else, $h = m$ and $a = -1$.
Step 3.   Let $g' = g' - 1$, $f_h = f_h + 1$.
Step 4.   If $\varphi = 1$, then let $u = \sum_{x = 1 \ldots h - 1} f_x$, $w = \sum_{x = 1 \ldots h} f_x$; else, let $u = t - \sum_{x = h \ldots m} f_x$, $w = u + f_h$.
Step 5.   Let $L_h = \sum_{x = u + 1 \ldots w} P_x$.
Step 6.   If $f_h < f_{max}$, $L_h < P^*(1 + \upsilon(h' - 1)) / \sigma$, and $g' > 0$, then return to Step 3.
Step 7.   Let $h = h + a$ and $h' = h' - 1$.
Step 8.   If $h' > 0$ and $g' > 0$, then return to Step 3.
Step 9.   If $h' = 0$ and $g' > 0$, then let $P^* = \infty$, $h' = m$, and return to Step 2.
Step 10.   Let $\Gamma = \Gamma \cup F$.
Step 11.   If $\phi > 1$, then let $\phi = \phi - 1$ and return to Step 1.
Step 12.   If $\phi = 1$ and $\varphi = 1$, then $\phi = 4$, $\varphi = m$ and Step 1; else, Stop.

Step 0 establishes the direction of the unbalance and the target unbalance. It also initializes the set of possible assignment sets. Step 1 calculates the parameters used to determine the percentage of the flowshop load (P^*) that will be targeted for each of the workstations. For example, if $m = 3$, $\varphi = 1$, and $\phi = 4$, then $\upsilon = 1.5$, $\sigma = 7.5$, and the target levels for the three workstations are: $(4/7.5)P^*$ or 53.33% of P^* for workstation 1, $(2.5/7.5)P^*$ or 33.33% of P^* for workstation 2, and $(1/7.5)P^*$ or 13.33% of P^* for workstation 3. These two steps also set the loop variables related to the number of operations to be assigned (g') and the total workstations (h'). At Step 2 depending on the direction of the imbalance parameter (φ), the loop variable h has initial values of 1 or m, and the step variable (a) is either 1 or -1. Step 3 adds an operation to the current workstation (h). Step 4 calculates the first and last operation presently assigned to the current workstation (h) and Step 5 determines the total load assigned to this workstation (L_h). Step 6 checks if the number of operations is less than the maximum allowed, if the current workload is below the loading target, and if there are still operations to assign. If all three conditions are satisfied, then the process returns to Step 3 (as to assign one more operation to the current workstation). Step 7 updates the current workstation variable and loop control variable. Step 8 verifies that the current workstation is valid and that there are still operations to assign, and if this is the case, the process returns to Step 3 (to assign the next operation). Step 9 represents the case where the process has

looped through all the workstations but did not assign all the operations. In this case the total processing time variable (P^*) is set to infinite and the process returns to Step 2. By doing this, the process will revisit all workstations assigning up to the maximum allowed by order of the direction of the unbalance. Step 10 is reached when $g' = 0$ and adds the current assignment F to the set of assignments. Step 11 updates the unbalance target for the next iteration. Finally, Step 12 starts the next set of iterations with the inverse unbalance direction.

2.2 Sequencing of Jobs

A variety of sequencing methods have been proposed for the PFSP, in particular for the C_{max} criterion. For the generation of a job sequence we assume an operations to workstations assignment has been established, so that the processing time per workstation (the q_{jh}) are fixed. We employ two heuristics, NEH-D from Dong et al. [4] and NEHKK1 from Kalczynski and Kamburowski [12], to develop the job sequence. These heuristics are based on the NEH heuristic (Nawaz et al. [18]) and worked well in the experiments reported by the authors. We use the best solution from these two methods.

2.3 Procedure Combinations

For each of the two criteria, three operation assignment procedures will be tested representing the baseline case of no operation flexibility, B, and U combined with the two sequence generation heuristics. Table 2 presents a summary of the procedure combinations. For those combinations that include U, each of the operation assignments generated by U, set Γ, is combined with each of the two sequence generation heuristics, and the best result is kept, with ties being solved by the secondary criterion.

TABLE 2. PROCEDURE COMBINATIONS

Operation to Workstation	Job Sequencing
fixed: $f_g = r \ \forall \ g \in M$	NEH-D
	NEHKK1
B	NEH-D
	NEHKK1
U	NEH-D
	NEHKK1

2.4 Improvement Procedure

A search based on simulated annealing is implemented to demonstrate that improvements for both criteria can be found by the implementation of meta-heuristics. Simulated annealing was selected from several available options (e.g., genetic algorithms, tabu search, or ant-colony algorithms) given it has been shown in previous studies (e.g., Ruiz and Maroto [20] and Manjeshwar et al. [16]) to perform well for this type of problems and criteria (comparative evaluation of the performance of simulated annealing, genetic algorithms, and other types of meta-heuristics for the described problem is left as a fruitful direction for future work). The search combines the job sequencing and task reallocation heuristics. For a given solution, we consider two possible neighborhoods – the reallocation of tasks across the workstations and the reorganization of the job sequence using single job insertions.

Given an operations to workstation assignment F, the set of neighbors, ϑ, is generated by reallocating one operation at a time across those workstations that can accept an additional operation. Without eliminating duplications and infeasible assignments, a total of m^2 neighbors are generated, noting that the original assignment F is part of the neighborhood. The proposed procedure to generate the neighbors is as follows:

Step 0.   Let $h = 0$ and $\vartheta = \{\}$.
Step 1.   Let $g = 1$ and $h = h + 1$.
Step 2   If $f_h > 1$, then let $f_h = f_h - 1$; else, go to Step 1
Step 3.   If $f_g \leq f_{max}$, then let $f_g = f_g + 1$; else, go to Step 6.
Step 4.   Add F to set ϑ
Step 5.   Let $f_g = f_g - 1$.
Step 6   If $g < m$, then $g = g + 1$ and go to Step 3.
Step 7.   Let $f_h = f_h + 1$.
Step 8.   If $h < m$, then go to Step 1; else End.

Step 0 initializes h representing the workstation that will "give up" an operation and the set of neighbors, ϑ, to empty. Step 1 initializes the workstation that will receive an operation (g) and updates the one "giving up". Step 2 verifies that an operation can be removed from workstation h, and if this is the case it is removed, else it returns to Step 1. Step 3 increases the number of operations assigned to workstation g if this is feasible, else moves to Step 6. At Step 4, the process has generated a neighbor; thus, it is added to the set of neighbors. Step 5 sets workstation g to its original value, while Step 6 updates the value of the workstation receiving the additional operation. At Step 7, all neighbors considering a removal from workstation h have been generated; thus, the number of operations in h is set back to its original value. Step 8 repeats the process until all workstations have been considered (to give up one operation).

Given a sequence of jobs Π, a neighborhood is generated by iteratively selecting and removing jobs from the sequence, and then inserting this candidate job in all possible positions. Without eliminating duplications, a total of n^2 neighbors are generated, noting that the original sequence is part of the generated neighborhood. The value of F is constant in this process.

Step 1.   Let $x = 1$ and $\Xi = \{\}$.
Step 2.   Let γ' be the partial sequence of jobs when job $\Pi(x)$ is removed from Π.
Step 3.   Generate n schedules by inserting job $\Pi(x)$ in all possible positions of sequence γ'. Add the generated schedules to Ξ.
Step 4.   If $x < n$, then let $x = x + 1$ and go to Step 2.

Step 1 initializes the job position that will be removed and inserted in the generation process. Step 2 removes the job and creates a partial sequence. Step 3 generates the n possible complete sequences. Lastly, Step 4 calls for a repeat of the process until all positions have been considered.

The search algorithms used in this research are based on the principles of simulated annealing. The input to the process is a seed schedule S with an operations to workstation assignment F and a job sequence Π.

Step 1.   Select the seed schedule.
Step 2.   While stop criterion is not satisfied, continue.
Step 2.1   Perform the loop n times.
Step 2.1.1   Generate all neighbors.
Step 2.1.2   Select the neighbor that provides the 'best' improvement.
Step 2.1.3   If there is no improvement, select a neighbor with probability exp (-*Delta*/*Temp*).
Step 2.1.4   Change the current schedule to the neighbor.
Step 2.2   Modify Temp.

The seed schedule used in Step 1 is the best solution generated by the six combinations listed in Table 2. There are two versions of search algorithm, based on the evaluation of different neighborhoods (Step

2.1.1). For the first, denominated *SA-Job Search (SA-J)*, the neighborhood is based on generating set Ξ, therefore this heuristic only searches for alternative job sequences and keeps the operations to workstation assignment set *F* fixed. For the second, denominated *SA-Full (SA-F)*, the neighborhood is created by generating ϑ for *F* and for each member of ϑ generating Π from the best solution of NEH-D and NEHKK1. This is solution set θ, and for each member of θ, generate set Ξ. The size of the neighborhood for *SA-J* is n^2, while for *SA-F* is n^2m^2, although not all of these are feasible/unique.

The value of *Temp* is set as B^X, where *X* is the number of cycles (number of times Step 2 is repeated). *B* is the cooling parameter and was tested at three levels for each experiment (0.9, 0.95, 0.97) and the best result selected. The value of *Delta* is based on the difference between the best solution found and the current solution. The stop criterion is the failure to generate an improvement. The total complexity of the procedure is n^3m^2 multiplied by the number of times an improvement is made.

3. EXPERIMENTS

This section describes experiments aimed at: a) analyzing the impact of operations flexibility; b) demonstrating that schedules that are makespan optimal may not be *MU* optimal and vice versa; and c) analyzing the performance of the heuristics. The impact of operations flexibility will be demonstrated by comparing the performance of schedules when operations are allocated evenly (by number of operations) across the workstations versus when this allocation is based on the workloads. Thus, the number of operations to each workstation can be significantly different. Demonstrating that the schedules that are C_{max} optimal may not be *MU* optimal is based on the difference between the best solutions found for each criterion and indicates that multiple middle of the road solutions exist (a Pareto set of solutions). The larger the gap between the two extreme solutions (the C_{max} best solution versus the *MU* best solution), the larger we can expect the range of Pareto solutions to exist. Finally, analyzing the performance of the heuristic is based on comparison to optimal values generated by full enumerative search, and to relative performance for larger problems.

Two experiment sets are conducted, namely, optimal benchmark set and relative benchmark set. In the first set, the optimal solution is found through full enumeration, and then the performance of the heuristics is compared to the optimal results. In the second set, the performance of the heuristics is compared to the best heuristic's solution. Relative performance is commonly used (e.g. Vallada and Ruiz [23]) when an optimal solution cannot be determined due to computational requirements.

Previous research in the PFSP has demonstrated that the expected processing times of the jobs is correlated to the complexity of the problem (Lodree et al. [15]). Problems where some workstations have higher expected processing times (bottleneck machines) present more complex cases than problems where the expected load in all machines is similar. In these experiments, we consider two variations of processing time variability, the case where some of the operations are bottleneck operations (high load operations) and the case where some of operations are "low load" operations. As in a large portion of the discussed PFSP literature, the processing times are generated by a discrete uniform random variable in the range [1,99] for what are called "regular" operations in this paper. For *bottleneck* operations, the range is [75,99], while for *low-load* operations the range is [1,25]. Let *b* represent the percentage of bottleneck operations and *l* represent the percentage of low load operations, and let both of these parameters serve as experimental factors. It should be noted that the actual number of bottleneck and *low-load* operations is determined by $\lceil b\,t \rceil$ and $\lceil l\,t \rceil$, respectively. The particular operations that will be bottleneck or *low-load* are randomly selected for each problem instance. The levels considered for variables *b* and *l* are 0%, 20%, and 40%, and combined in the following five cases (*b* = 20%, *l* = 0%), (*b* = 40%, *l* = 0%), (*b* = 20%, *l* = 20%), (*b* = 0%, *l* = 20%), and (*b* = 0%, *l* = 40%). The remaining operations are "regular" operations with a processing time range of 1-99. For example, if *b* = 20%, *l* = 20% and *t* = 9 operations, then there would

be two bottleneck operations, two *low-load* operations, and five regular operations. For a problem instance, let's assume operations 2 and 6 were randomly selected as the bottleneck operations and operations 1 and 3 as the *low-load* operations. Consequently, p_{j4}, p_{j5}, p_{j7}, p_{j8}, and p_{j9} will be drawn from $DU[1,99]$; p_{j2} and p_{j6} will be drawn from $DU[75,99]$; while p_{j1} and p_{j3} will be drawn from $DU[1,25]$; for all jobs of that instance.

3.1 Optimal Benchmark

The optimal benchmark experiments consider three experimental variables; m, r, and (b,l). The number of machines is considered at 3 and 4, the ratio of operations to machines at 2 and 3, and the levels of b and l as described earlier. This results in 20 experimental combinations. One hundred replications are performed by experimental point, resulting in 2,000 total experiments. For these experiments, the number of jobs is fixed at 7 given that pilot experiments with fewer jobs provided no significant differences in the results; while full enumeration experiments with larger values of n were infeasible due to excessive computational times.

For the makespan criterion, results from the optimal benchmark experiments confirmed a relatively small improvement when operations to workstation flexibility was implemented. For the workstation utilization criterion, a relatively large improvement was obtained using the operations to workstation flexibility. The experimental results also indicate that the benefits obtained from implementing operations flexibility increase as problem complexity increases. Furthermore, the results demonstrated that if the objective is to minimize C_{max}, balancing the load of the workstations works well, while if the objective is to minimize MU, the unbalanced approach should be used. Both of the tested job sequencing approaches resulted in similar performance, but further improvements were realized when implementing a simulated annealing search only on the job sequencing part of the problem. This shows that job sequencing is still a factor in the considered problem environment. Also, the experiments proved the value of the combined simulated annealing search procedure to improve the initial solutions and to find close to optimal solutions.

3.2 Relative Benchmark

These experiments consider the experimental variable m at higher values than in the optimal benchmark experiments (i.e., $m = 5$ and $m = 10$) and also evaluate the effect on performance of changes in the number of jobs (i.e., $n = 20$ and $n = 50$). The factors (b, l) and r are considered at the levels described previously, resulting in 40 experimental combinations. Twenty-five replications are performed by experimental point, resulting in 1,000 total experiments.

In the case of the relative benchmark experiments, the results confirmed previous conclusions. Over all set of experiments, there were no solutions that simultaneously minimized makespan and workstation utilization. Moreover, schedules that are built to be good under the makespan criterion typically performed well for the workstation utilization performance (MU); but schedules that are built to be good under the workstation utilization criterion have poor makespan performance. As in the case of optimal benchmark experiments, the B approach performed well for the makespan criterion and the U approach performed well for the MU criterion. Furthermore, we can conclude from the experiments that as the problem size increases (variables m, n and r), the probability of a single schedule optimizing both performance criteria would be very low, and therefore the existence of multiple Pareto efficient solutions to each problem instance.

4. CONCLUSIONS

This paper investigates the flowshop scheduling problem with operations flexibility – the ability of the shop to organize the production operations along the production resources. Moreover, the paper

introduces the machine dominance principles for the flowshop scheduling problem with operations flexibility considering makespan as performance criterion. The operations flexibility is studied when the goals are to minimize the total completion time and the utilization of workstations. A two phase heuristic framework is proposed to solve this problem; the first procedure is aimed to allocate operations to workstation, and the second one is designed to generate the job sequencing. Furthermore, a two stage neighborhood search based on simulated annealing is implemented as improvement method. An extensive computational experiment is conducted to analyze the effect of the experimental factors, show the significance of the heuristic factor, and demonstrate the tradeoffs between maximum completion time and workstation utilization.

REFERENCES

[1] Burdett, R.L. & Kozan, E. *Sequencing and Scheduling in Flowshops with Task Redistribution*. The Journal of the Operational Research Society, 2001, 52, 1379–1389.

[2] Coffman, E.G. Jr. Computer and Job Shop Scheduling Theory. New York, NY: Wiley, 1976.

[3] Coffman, E.G. Jr., Garey, M.R. & Johnson, D.S. *An Application of Bin-packing to Multiprocessor Scheduling*. SIAM Journal on Computing, 1978, 7, 1–17.

[4] Dong, X., Huang, H. & Chen, P. *An Improved NEH-based Heuristic for the Permutation Flowshop Problem*. Computers & Operations Research, 2008, 35, 3962–3968.

[5] Fondrevelle, J., Oulamara, A. & Portmann, M-C. *Permutation Flowshop Scheduling Problems with Time Lags to Minimize the Weighted Sum of Machine Completion Times*. International Journal Production Economics, 2008, 112, 168–176.

[6] Framinan, J.M., Leisten, R. & Rajendran, C. *Different Initial Sequences for the Heuristic of Nawaz, Enscore and Ham to Minimize Makespan, Idletime or Flowtime in the Static Permutation Flowshop Sequencing Problem*. International Journal of Production Research, 2003, 41, 121–148.

[7] Garey, M., Johnson, D. & Sethi, R. *The Complexity of Flowshop and Jobshop Scheduling*. Mathematics of Operations Research, 1976, 1, 117–129.

[8] Gupta, J.N.D. & Stafford, E.F. Jr. *Flowshop Scheduling Research after Five Decades*. European Journal of Operational Research, 2006, 169, 699–711.

[9] Ho, J.C. & Chang, Y.L. *A New Heuristic for the n-job, m-machine Flowshop Problem*. European Journal of Operational Research, 1991, 52, 194–202.

[10] Johnson, S.M. *Optimal Two- and Three-stage Production Schedules with Set up Times Included*. Naval Research Logistics Quarterly, 1954, 1, 61–68.

[11] Kalczynski, P.J. & Kamburowski, J. *On the NEH Heuristic for Minimizing the Makespan in Permutation Flow Shops*. OMEGA, The International Journal of Management Science, 2007, 35, 53–60.

[12] Kalczynski, P.J. & Kamburowski, J. *An Improved NEH Heuristic to Minimize Makespan in Permutation Flow Shops*. Computers & Operations Research, 2008, 35, 3001–3008.

[13] Laha, D. & Sarin, S.C. *A Heuristic to Minimize Total Flow Time in Permutation Flow Shop*. OMEGA, The International Journal of Management Science, 2009, 37, 734–739.

[14] Li, X., Wang, Q. & Wu, C. *Efficient Composite Heuristics for Total Flowtime Minimization in Permutation Flow Shops*. OMEGA, The International Journal of Management Science, 2009, 37, 155–164.

[15] Lodree, E., Jang, W. & Klein, C.M. *A New Rule for Minimizing the Number of Tardy Jobs in Dynamic Flow Shops*. European Journal of Operational Research, 2004, 159, 258–263.

[16] Manjeshwar, P.K., Damodaran, P. & Srihari, K. *Minimizing Makespan in a Flowshop with Two Batch-processing Machines Using Simulated Annealing*. Robotics and Computer-Integrated Manufacturing, 2009, 25, 667– 679.

[17] Naderi, B. & Ruiz, R. *The Distributed Permutation Flowshop Scheduling Problem*. Computers & Operations Research, 2010, 37, 754–768.

[18] Nawaz, M., Enscore, E.E. & Ham, I. *A Heuristic Algorithm for the m-machine, n-job Flow-shop Sequencing Problem*. OMEGA, The International Journal of Management Science, 1983, 11, 91–95.

[19] Rad, S.F., Ruiz, R. & Boroojerdian, N. *New High Performing Heuristics for Minimizing Makespan in Permutation Flowshops*. OMEGA, The International Journal of Management Science, 2009, 37, 331–345.

[20] Ruiz, R. & Maroto, C. *A Comprehensive Review and Evaluation of Permutation Flowshop Heuristics*. European Journal Operational Research, 2005, 165, 479–494.

[21] Ruiz-Torres, A.J., Ablanedo-Rosas, J.H. & Ho, J.C. *Minimizing the Number of Tardy Jobs in the Flowshop Problem with Operation and Resourse Flexibility*. Computers & Operations Research, 2010, 37, 282–291.

[22] Sridhar, J. & Rajendran, C. *Scheduling in Flowshop and Cellular Manufacturing Systems with Multiple Objectives—A Genetic Algorithmic Approach*. Production Planning & Control, 1996, 7, 374–382.

[23] Vallada, E. & Ruiz, R. *Cooperative Metaheuristics for the Permutation Flowshop Scheduling Problem*. European Journal of Operational Research, 2009, 193, 365–376.

[24] Vallada, E. & Ruiz, R. *Genetic Algorithms with Path Relinking for the Minimum Tardiness Permutation Flowshop Problem*. OMEGA, The International Journal of Management Science, 2010, 38, 57–67.

A MATHEMATICAL MODELING APPROACH FOR THE SCHEDULING PROBLEM OF LOAD-HAUL-DUMP VEHICLES IN UNDERGROUND MINES

Irem OZKARAHAN
Troy University, Department of Computer Science, iozkarahan@troy.edu
Pinar Mizrak Ozfirat∗
Celal Bayar University, Department of Industrial Engineering, pinar.ozfirat@bayar.edu.tr
M. Kemal Ozfirat∗
Dokuz Eylul University, Department of Mining Engineering, kemal.ozfirat@deu.edu.tr

ABSTRACT

In underground mining, one of the major concerns is effective use of vehicles to produce and transport the mine ore out of the mine to the ground surface. In metal mining in most of the cases, Load-Haul-Dump (LHD) vehicles are employed to draw the mine ore from the drawing points and carry it to transfer points to load the ore onto trucks. From that point, trucks are in charge to transport the ore out of the mine. In this paper, a mathematical programming model is developed to solve the scheduling problem of LHD vehicles in underground mines. The problem is modeled as a parallel machine scheduling problem with unequal machines. The model is solved by OPL Studio 3.7 and used for a specific case solution. Optimum number of LHD to be employed and the optimum schedule of each vehicle is found.

Keywords: Load-Haul-Dump vehicle scheduling, underground mining, mixed integer programming

1. INTRODUCTION

Load-Haul-Dump (LHD) vehicles are ore transporting machines adapted for underground conditions. These vehicles are started to be used first in 1950s and are widely spread by 1960s. Today, they are employed in most of metal mines all around the world[1]. A typical LHD vehicle can be seen in Figure 1.

Figure 1. Standard LHD vehicle.

∗ Post doctorate researcher at Troy University, Department of Computer Science.

In underground mines, LHDs are used to load ore from the drawpoints in the mine and transport it to a truck. Then trucks are employed to take the ore out of the mine to a crusher machine on the surface. One of the main decision making problems that occur during this process is to decide on the number of LHDs and trucks as well as to schedule these vehicles. A similar problem occurs between trucks and shovels in open pit mines which is addressed by many researchers [2-4].

However, LHD scheduling problem is not addressed by many researchers. There are very few studies in this area. Scheding et al [5] describes the theoretical development and experimental evaluation of an autonomous LHD based on the results obtained by field trials. Saayman et. al. [6] examined the LHD and truck scheduling using simulation. To the best of our knowledge, this study is the first one which develops a mathematical programming model for LHD scheduling problem in underground mines. The model is used to find the effective number of LHDs and trucks for mine operation as well as the schedule of LHD vehicles for a specific case.

The rest of the paper is organized as follows. In section 2, problem overview is given. In section 3, the proposed approach is explained in detail. Computational results are given in Section 4. Finally section 5 is the conclusion.

2. PROBLEM OVERVIEW

In an underground metal mine, LHDs drive through the tunnels to the ore drawpoints. Then they load ore, move it to the transfer point where they dump the ore to a truck. Then trucks move the ore out of the mine. They take it to a crusher machine located on the ground surface. A schematic view of an underground mine can be seen in Figure 2.
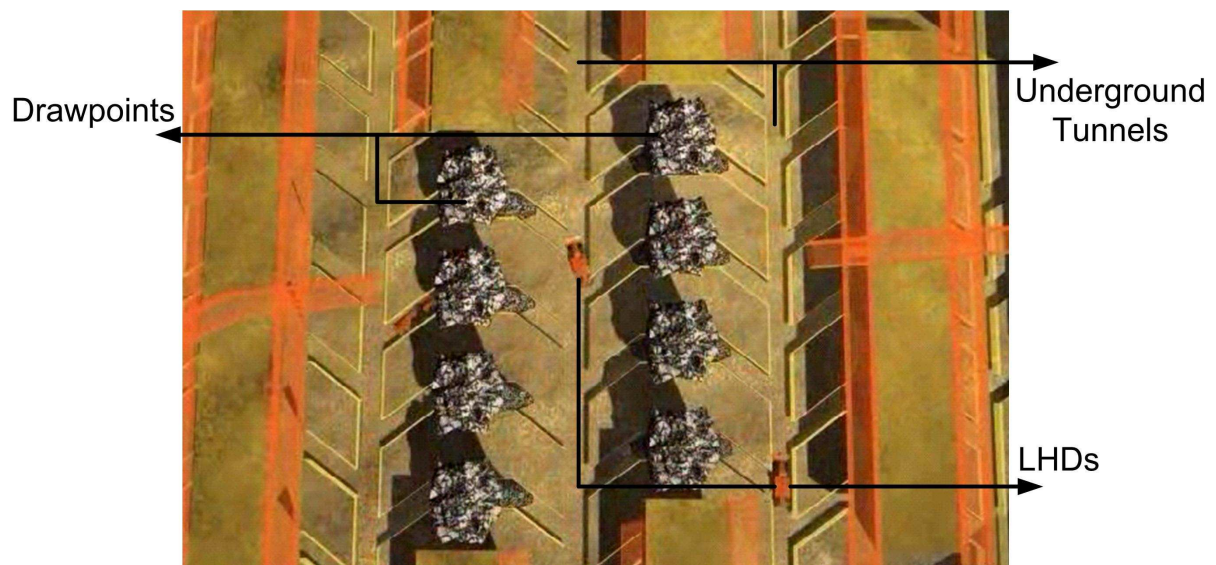
Figure 2. Top view of an underground mine which employs LHDs.

LHDs need to make several tours between the drawpoints and transfer point in order to fill the truck. Then the truck transports the ore out of the mine and comes back to the transfer point. Operations chart for LHD and truck in the underground mine can be seen in Figure 3. The ore reserve located in the mine is known ahead. Therefore, production amount is already known. There is nothing to be determined for production maximization. What we should determine is how to make this production in minimum time. In other words, all the ore reserve should be taken out of the mine in the minimum time. By this way, LHD and truck utilization would be maximized.
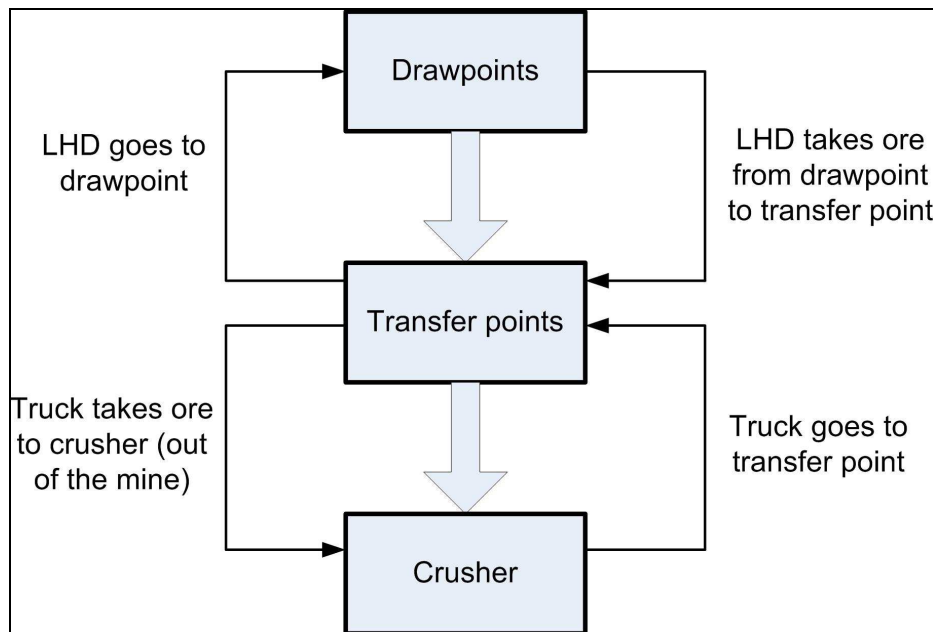
Figure 3. Operation chart for the LHD and the truck.

Another point in the problem is that there exist more than one transfer points in mines. The distance and hence the durations from drawpoints to each transfer point is different. Therefore in the optimal schedule LHDs should be assigned to drawpoints accordingly.

Considering these issues, a mixed integer programming model is developed in the next section in order to find the optimal schedule of LHDs.

3. PROPOSED APPROACH

The problem is modeled as a parallel machine scheduling problem where LHDs are the parallel machines and the tours of LHDs to drawpoints are jobs to be scheduled. Mixed integer programming is employed. There are certain assumptions made in the model:
- In order to prevent collision of LHDs within tunnels each LHD is associated with a transfer point. That is each LHD is in charge to serve one transfer point.
- Each truck is associated with a transfer point. That is each truck serves one transfer point.

The notation of the model is given in Table 1.

Table 1. Notation used in the model

Sets :

i : Number of jobs to be scheduled, i = 1..n

j : Number of positions, j : 1..n (all positions of all LHDs need not to be filled)

k : Number of LHDs, k : 1..m

Parameters :

P_{ik} *: Duration of job i if it is processed by LHD k.*

TTT : Truck tour time

Decision Variables :

C_{max} *: Makespan*

C_{jk} *Ending time of the job in position j of LHD k.*

$X_{ijk} : \begin{cases} 1 \text{ if job i is scheduled in position j of LHD k.} \\ 0 \text{ otherwise} \end{cases}$

The model is given with Equations 1 to 11 below.

$$min. \quad C_{max} \qquad (1)$$

Subject to

$$C_{max} \geq C_{jk} \quad \forall j,k \qquad (2)$$

$$\sum_j \sum_k X_{ijk} = 1 \quad \forall i \qquad (3)$$

$$\sum_i X_{ijk} \leq 1 \quad \forall j,k \qquad (4)$$

$$\sum_i X_{ij+1k} \leq \sum_i X_{ijk} \quad \forall j,k \qquad (5)$$

$$C_{jk} = C_{j-1k} + \sum_i P_{ik} \cdot X_{ijk} \quad \forall j \in \{1..n\} - \{6,12,18,...\} \qquad (6)$$

$$C_{jk} \geq C_{j-1k} + \sum_i P_{ik} \cdot X_{ij-1k} \quad \forall j = \{6,12,18,...\} \qquad (7)$$

$$C_{jk} \geq (C_{j-1k} + TTT) \cdot \sum_i X_{ij-1k} \quad \forall j = \{6,12,18,...\} \qquad (8)$$

$$C_{0k} = 0 \quad \forall k \qquad (9)$$

$$X_{ijk} \in 1 \qquad (10)$$

$$C_{jk} \geq 0 \qquad (11)$$

Equation 1 is the objective function which is to minimize makespan. Equation 2 defines makespan (All jobs precede makespan). Equation 3 states that all jobs must be scheduled exactly once. Equation 4 states that at most one job can be scheduled in one position and equation 5 makes sure that all positions are filled consecutively (i.e. no in-between position is empty). Equation 6 defines the ending time of the job at the jth position on LHD k. However,

for the jobs at 6th, 12th, 18th,... positions, LHD needs to wait for the truck to return from crusher (every five tour of an LHD leads to one tour of a truck). This is stated by equation 7 and 8. Also, in order to define the first position jobs on LHDs, a dummy "*job 0*" is assigned to each LHD. Equation 9 states the ending time of "*job 0*" is 0. Finally, equations 10 and 11 define the decision variables domains.

The model is written and solved in OPL Studio 3.7 [7]. Computational results are given in the next section.

4. COMPUTATIONAL RESULTS

In the mine under study, there are 15 drawpoints each of which should be visited by an LHD five times. Therefore, there exist 75 jobs to be performed totally. Jobs to be performed are referred with their numbers as *1*, *2*, *3*, *4*, …, *75* to represent *Job 1*, *Job 2*, *Job 3*, *Job 4*, …, *Job 75* respectively.

In the solution of the model, initially it is assumed that there are three LHD vehicles and three trucks in the mine. Once the model is solved with these parameters, different number of trucks and LHDs are tested in order to find the most effective number of LHDs and trucks for the mine operation. As number of LHDs increase (i.e. machines increase), makespan would surely decrease. On the other hand as the number of trucks increase, the truck tour time, *TTT*, (time it takes a truck to go to the surface and come back) decreases.

The experiments carried out with different number of LHDs and trucks and the corresponding makespan values are given in Table 2 below.

Table 2. Computational results of the model with different parameter levels.

Experiment Number	Number of LHDs	Number Trucks	Makespan (minutes)
1	3	3	180
2	4	3	141
3	5	3	115
4	*6*	*3*	*90*
5	3	4	177
6	4	4	141
7	5	4	115
8	6	4	89
9	3	5	177
10	4	5	141
11	5	5	113
12	6	5	89
13	3	6	177
14	4	6	136
15	5	6	112
16	6	6	88
17	7	3	80
18	8	3	72
19	9	3	65
20	10	3	61
21	11	3	55
22	12	3	52

The change in makespan values according to different number of LHDs can be seen in Figure 4 below. The tradeoff between the benefits obtained by the decrease in makespan and the cost of an additional LHD vehicle should be analyzed carefully and is beyond the scope of this paper. However, at first glace, as number of LHD increases from 3 up to 6, the decrease in makespan is quite sharp. But, as number of LHD increases from 6 to 12 the decrease in the makespan is not very considerable. Therefore it may be beneficial to operate 6 LHDs in the mine.
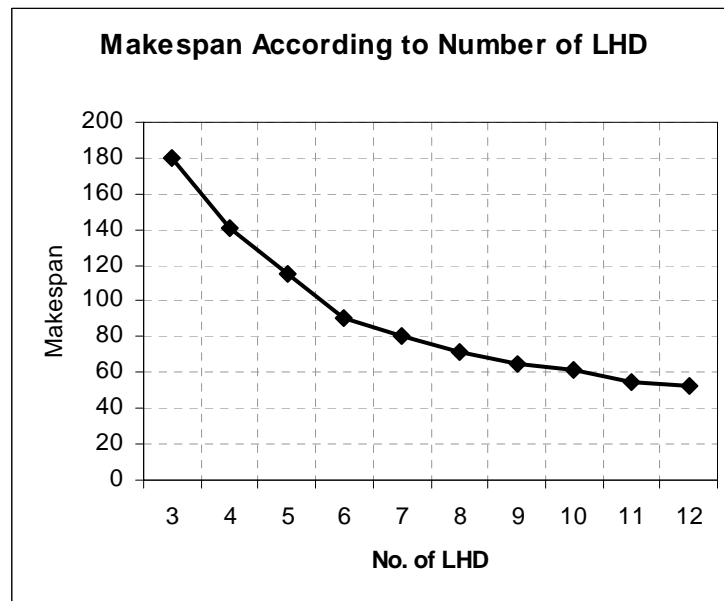
Figure 4. Change in makespan according to number of LHD vehicles employed.
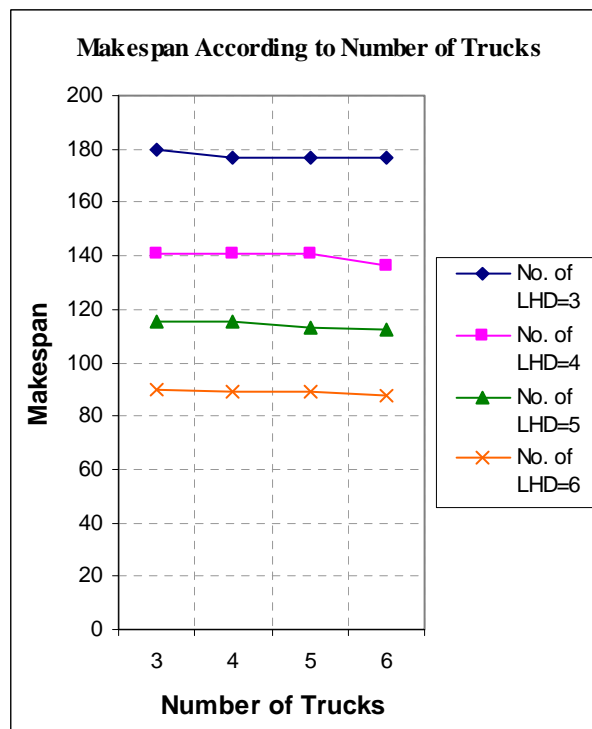
Figure 5. Change in makespan according to number of trucks employed.

On the other, the change in makespan values according to the number of trucks can be seen in Figure 5. As it is seen from the figure, number of trucks have very little or no impact on the makespan value. Therefore it may be advised to have only 3 trucks in mine operation. From Table 2, the makespan value for 6 LHDs and 3 trucks (experiment 4) is 90 and the corresponding schedule is given in Figure 6. The numbers in the figure correspond to the numbers of jobs where there are 75 jobs to be scheduled totally.
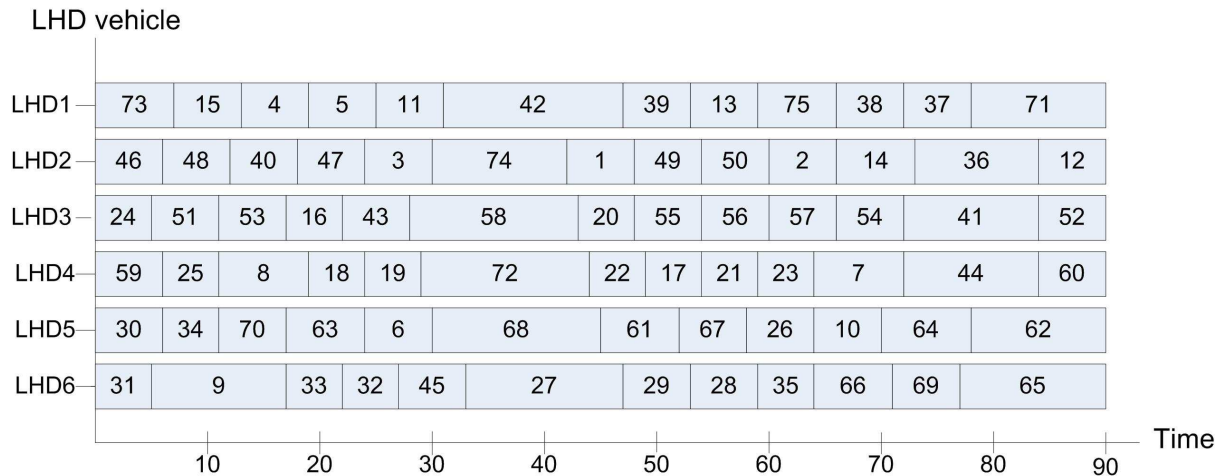
Figure 6. Gantt chart of the schedule belonging to 6 LHD- 3 truck operation

5. CONCLUSION

In this study, the LHD scheduling problem of underground mining is handled. A mixed integer programming model is developed for the solution of the problem. The model is run to test different number of LHDs and trucks. The results revealed that operating six LHD vehicles and 3 trucks may be the most effective and efficient solution for the mine under study.

The model developed in this paper is used to solve a case specific problem. However, it can be easily modified for different assumptions and operations of different mines. Therefore it can be a powerful tool for the vehicle scheduling problem of underground mines.

REFERENCES

[1] Tatar, C. & Ozfirat, M.K. Underground Mine Machines and Mechanization, Izmir: Dokuz Eylul University Engineering Faculty Publications, 2003 (In Turkish).

[2] Nasuf E. & Kırmanlı, C. *Computer Aided Truck Dispatching Systems in Open Pit Mines and World Applications*, Mining Journal, 1993, 32(2), 5-11 (In Turkish).

[3] Ta, C.H., Kresta, J.V., Forbes, J.F., Marquez, H.J. *A Stochastic Optimization Approach to Mine Truck Allocation*. International Journal of Mining, Reclamation & Environment, 2005, 19(3), 162-175.

[4] Aksoy, C.O., Kose, H., Malli, T., Ozfirat, K., Gonen, A., Gurgen, S. *Truck-Shovel Optimization Accounting For Cost, Safety, and Operational Practice*, CIM Bulletin, 2005, 98, (1091).

[5] Scheding S., Dissanayake G., Nebot E. M., and Durrant-Whyte H. *An Experiment in Autonomous Navigation of an Underground Mining Vehicle*, IEEE Transactions on Robotics & Automation, 15(1), 1999.

[6] Saayman, P., Craig, I.K., and Camisani-Calzolari, F.R. *Optimization of an Autonomous Vehicle Dispatch System in an Underground Mine*, The South African Institute of Mining & Metallurgy, 2006, 106, 77-86.

[7] *OPL Studio 3.7 Language Manual*, 2003, ILOG S.A., France.

THE BEHAVIOUR OF UNPACED PRODUCTION LINES WITH UNEQUAL MEAN PROCESSING TIMES, VARIABILITY, OR BUFFER CAPACITIES

Sabry Shaaban[1], Tom McNamara[2] and Ahmed Atil[2]

1 Department of Economics, Strategy and Organization
ESC La Rochelle
102 Rue de Coureilles
17024 La Rochelle, France
Email: shaabans@esc-larochelle.fr

2 Department of Finance and Operations
ESC Rennes School of Business
2 Rue Robert d'Arbrissel
35065 Rennes, France
Email: tom.mcnamara@esc-rennes.fr
Email: ahmed.atil@esc-rennes.fr

ABSTRACT

In this paper we study the operating behaviour and performance of reliable, unpaced and unbalanced serial production lines with either imbalanced service time means, unequal coefficients of variation, or uneven buffer capacities. The lines were simulated with various values of line length, buffer storage size, degree of imbalance, coefficient of variation, along with a number of imbalance configurations. The primary measures of efficiency were idle time and average buffer level. Output data from the discrete event simulation of such lines under their steady-state mode of operation were analyzed using a set of statistical methods. Various relationships between the independent and response variables, rankings of configurations and comparisons with balanced lines were obtained. For the mean processing times imbalance, it turned out that a bowl-shaped arrangement provides smaller idle time amounts and lower average buffer levels than those of a balanced line counterpart. As regards the variability imbalance, it was found that the best configurations are respectively, a bowl allocation and a monotone decreasing order, with the first resulting in decreased idle times and the second leading to lower average buffer levels than those of a balanced line. As far as the buffer size imbalance is concerned, it was concluded that the most advantageous patterns that generate lower idle times and average buffer levels as compared to a balanced line are to respectively distribute total available buffer capacity as evenly as possible along the buffers and to allocate more buffer capacity towards the end of the line.

Keywords: unpaced serial production lines; simulation; one source imbalance; imbalanced operation time means; unequal coefficients of variation; uneven buffer sizes; idle time; average buffer level

1

1. INTRODUCTION

When setting up an unpaced production line (with no form of mechanical pacing), how you design it is going to impact its efficiency quite considerably. For instance, where to place operators who work at different speeds, or vary in the speed they work at or where to keep unfinished items along the production line are just some of the problems facing the line manager.

Buffer Allocation
One factor that needs thinking about is determining the size of the storage space (buffer) in between workstations where partly finished products are kept, awaiting the next step of the process. Lines unbalanced with respect to their buffer capacities are of great interest as technical considerations often restrict the amount of space available in the line, thereby making it difficult to allocate total buffer capacity evenly amongst individual buffers.

Average Working Time of an Operator
Not all the operators are going to be able to complete their tasks at the same time. People work at different average speeds for several reasons, some are personal, their physical capacity, their motivation and some are inherent to the task, it might be a complex task or just simply that the amount of work along the line just can't be distributed evenly in terms of time. It is clear that some tasks cannot be completed until the preceding steps have taken place; a very simplistic example being we can't pack a product until it's made.

Worker Variability
Not only do different operators work at different speeds, the one and same person can vary in the rate at which he or she works over the day for example. A person's working speed can vary considerably from his or her average. This can be for different reasons: fatigue, boredom and tasks that are complex or changing.

Figure 1 shows a 5-station serial line with 4 buffers, where the squares depict the stations and the diamonds represent the buffers:
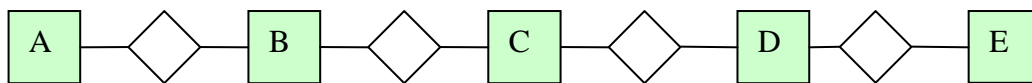
Figure 1. An unpaced production line

This paper is organized as follows. First, the relevant literature is reviewed. Next the objectives, methodology and experimental design of the study are presented. Subsequent sections give the simulation results, their analysis and compare the performance of balanced and unbalanced lines. The two last sections provide summary of the results, discussion and conclusions.

2. LITERATURE REVIEW

Unpaced serial production lines are a vitally important type of production system. A multitude of industries rely on them to provide a whole host of goods and services. Because of this critical role, they have deservedly received a great deal of attention in the literature.

2

2.1 Unbalanced Operation Time Means

Numerous studies have been reported on the performance and behaviour of unbalanced lines with unequal mean service times (MTs). These investigations can be classified into three general groups; bowl phenomenon, algorithmic and theory of constraint studies. These will be briefly reviewed in turn:

The Bowl Phenomenon
One of the most important investigations is that of [1], who analysed lines having up to 4 stations, exponential work time distributions, and equal buffer capacities. They found that the optimal throughput rate (TR) is achieved by assigning more work (higher MT) to the end stations and less work (smaller MT) to the middle stations, resulting in a gain of 0.54% in TR over that of the balanced line. They called this discovery the "bowl phenomenon". It was found, however, that as the size of the buffers increases, this phenomenon quickly disappears, resulting in an optimal balanced line configuration. Later on, [2] & [3] extended their analysis to lines of up to six stations with Erlangian distributions and found that the improvement in TR due to unbalancing rises to 1.37% for a six-station line.

Since then several other papers have been published on various aspects of the bowl phenomenon (see for example [4] [5] [6]).

Algorithmic Approaches
A second line of research focused on the development of mathematical and algorithmic approximation methods to obtain various performance measures (mainly TR). Several investigations have been conducted, resulting in predictive formulas being developed and applied to various MT unbalanced conditions (see e.g. [7] [8] [9]).

Theory of Constraints
A third research area is in line with the Theory of Constraints (TOC). Briefly, TOC is concerned with the identification of the slowest station (the bottleneck or constraint), and adding more resources to it so that it will never be starved of work. TOC is based on the drum-buffer-rope (DBR) concept. The drum represents the bottleneck station, which dictates the overall movement and TR of the line. The buffers are used to provide protection for the bottleneck against statistical fluctuations. The rope is a signalling device from the bottleneck to drive all stations to work in harmony with the pace of the bottleneck, pulling work into it. Another TOC control system is constant work in process (CONWIP), under which a limit is imposed on the line's total work in process (WIP).

One of the earliest investigations into this domain was that of [10]. They modelled 6-station uniform as well as exponential single bottleneck lines, with buffer capacities of up to 20 units. They found that the average buffer content is highest in the buffer that immediately precedes the bottleneck.

[11] studied three types of line design: a traditionally balanced line, a just-in-time line, and a TOC line. He concluded that the TOC line performed better than the other two designs. Similar findings were obtained by [12] for a six-station line with the bottleneck being located at the last station.

[13] simulated two line designs; a CONWIP line with protective capacity and another without

it (balanced). The results indicate that a line with protective capacity achieves lower cycle times (higher TR) than a balanced line.

[14] studied a 4-station TOC line. Their results show that both protective capacity and location of the constraint lead to significantly improved output, smaller idle time, and less work in process, but with diminishing returns. They also found that it is beneficial to position the constraint at the first station.

[15] used simulation to investigate CONWIP lines with a single constraint located at the first, middle and last station. Some generalizations concerning the protective capacity amount and location were obtained.

[16] simulated single bottleneck lines and found that a DBR line outperforms a CONWIP line by 15%, with the percentage gain increasing when the bottleneck is placed near the beginning of the line. However, as the difference in capacity between the bottleneck and non-bottleneck stations falls below 2%, CONWIP starts to outperform the DBR.

[17] reported the same finding on the superiority of DBR line as compared to CONWIP, but at the cost of accumulating large WIP at the constraint station. [18] studied over 80 firms that instituted TOC methods and in each case they found significant improvements in operational performance.

[19] used a drum development method for manufacturing environments with bottleneck re-entrant flows, in an attempt to implement an effective DBR management system.

2.2 Unbalanced Variability of Service Times

[20] simulated a 4-station line with buffer capacity of 6 or more and found that interspersing the 2 deterministic stations with the 2 variable resulted in higher IT levels than those of the balanced line.

Other early studies, employing mostly simulation, seemed to show that placing stations incrementally, with the highest CVs towards the end of the line gave good results in terms of getting lower idle time (IT) than the balanced line, or a slight increase in output rate, or sometimes both (see e.g. [21] [22] [23]).

[24] found that the bowl phenomenon also existed for lines unbalanced in terms of their coefficients of variation (CVs). Their simulation of 4 and 10 station lines showed that a bowl-shaped line (stations with lowest CVs placed in the middle) gave lower IT and higher production rates than an inverted bowl-shaped line. They found that the gains due to unbalancing the CVs were higher than those seen when unbalancing the MTs. [25] confirmed the results of [24] concerning the favourable performance of the bowl arrangement.

In a study of 3, 4 and 12 station lines, [26] found that for the shorter 3 and 4 station lines, slight reductions in IT were obtained for the bowl pattern over the balanced line. For the longer 12 station line however, ITs were significantly higher than for the control, suggesting that the superiority of the bowl pattern only existed for shorter lines. This tendency was also observed by [27], with the best patterns for shorter lines being bowl-shaped, while for longer lines ($N \geq 9$), the bowl pattern gave results inferior to the balanced line.

4

Other approaches include the use of heuristic or optimization methods to assess the Performance of CV-unbalanced lines, mostly in terms of throughput rate (see for instance [28] [29] [30] [31]). These algorithms have been applied to lines having unequal CVs, squared CVs or standard deviations under a variety of operating conditions and for differing CV patterns.

2.3 Uneven Buffer Capacities

There is a significant body of literature on the issue of buffer allocation in production lines and its effects on performance.

[32] found no substantial improvement in throughput from uneven buffering. In addition, if unbalancing was unavoidable, larger buffer capacity (BC) should be allocated to the central stations and smaller BC to the end stations.

[33] developed an algorithm to determine the optimal buffer allocation with regard to throughput. Results showed that for a limited number of occasions, a line with unbalanced buffers can provide a superior throughput to that of a balanced line. [34] indicated that BC should be placed as evenly as possible along the line, but any small additional amount of BC should be allocated to the line's centre.

[35] stated that for a limited total buffer capacity, a balanced allocation is best. As more BC becomes available, preferential treatment should be given to the centre of the line.

[36] concluded that there is no single optimal buffer allocation policy for all operating conditions, but that preference should be given to the central locations of the line. Should an optimal buffer configuration be unknown; it is best then to strive for an inverted bowl arrangement. This was referred to as the 'storage bowl phenomenon'.

[37] studied a three station line with two buffer locations, namely B1 and B2. He found it best to alternate between the two placed (i.e. assigning the first unit to B1, the second unit to B2, the third unit again to B1 and so on). This policy was termed the 'Alternation Rule'.

[38] advised that the central stations should be given priority when allocating BC. [39] found that the optimal placement of a single unit of buffer is toward the central location, resulting in an increase in throughput.

Other authors developed a variety of buffer allocation algorithms. [29] used dynamic programming to determine an optimal BC allocation. [40] developed a solution procedure for optimizing profit within buffer and production constraints. [41] worked out a cash-flow oriented buffer allocation method. [42] employed a simulated annealing approach to find an optimal BC policy. [43] developed an algorithm for allocating buffer capacity to minimize average work-in-process.

[44] stated that if the interest is to maximize throughput, then an inverted bowl or a close approximation is generally the preferred arrangement. On the other hand, if the objective was a reduction in average WIP level, then it is better to allocate more buffers towards the end of the line.

[45] in a comparison study of three rules for buffer allocation using search methods found that

5

lines where buffer space was evenly distributed should be avoided when combined with ascending or descending mean-time operation patterns in terms of throughput.

[46] utilized a dynamic programming algorithm for the efficient distribution of BC, with the stated objectives of maximizing throughput and minimizing cycle time, WIP, and the probability of a station being blocked.

[47] tested the bowl storage hypothesis in balanced and unbalanced lines. In the case of a line with buffer imbalance, optimal solutions in terms of cost showed that the bowl storage pattern performed best for lines with more buffer space, and that more evenly allocated buffers gave the optimal solution when buffer space was limited. These results were confirmed in a later study by [48] with the same patterns being seen for limited or plentiful buffer space availability in lines unbalanced in terms of their mean operation times.

[49] looked at buffer allocation in terms of whether and when a push or pull strategy was best at different line lengths. For lines where the buffer was allocated evenly, throughput was maximised, whereas unbalancing the buffer allocation gave better results in terms of WIP, ascending buffer patterns being superior to descending patterns.

More recently, [50] considered buffer allocation in balanced and unbalanced lines as part of a study investigating both reliable and automated lines.

3. STUDY OBJECTIVES, METHODOLGY AND EXPERIMENTAL DESIGN

This research aims at studying the operating behaviour of reliable unpaced lines, with unequal mean service times, variability, or buffer sizes. The main objectives of the investigation are:
- To assess the merits of various patterns of imbalance and identify the best ones.
- To compare and contrast the performance of the unbalanced lines studied with that of a balanced line counterpart and determine possible performance improvement.
- To gauge the effects of line design factors – line length, buffer capacity / mean buffer capacity and degree of imbalance on the dependent measures of performance; idle time and average buffer level.
- To identify the most important factors influencing the dependent variables.

As no mathematical procedure is presently capable of handling the unbalanced steady-state characteristics of the more realistic lines, computer simulation was utilized as the most suitable technique for this kind of study.

3.1 Factorial Design

The most efficient and powerful of the many experimental designs is the complete factorial design. This type of design has been chosen for the current investigation. In the context of the particular lines being studied, the independent variables were:
- Total number of stations in the line (N).
- Total amount of buffer capacity for the line (TB).
- Buffer capacity, BC / mean buffer capacity (MB), where MB = TB divided by the number of buffers.
- Degree of unbalanced service time means (DI).
- CV value range
- Pattern of mean work time (MT) imbalance.

6

- Pattern of coefficients of variation (CV) imbalance.
- Pattern of buffer capacity (BC) imbalance.

3.2 Work Times Distribution and Performance Measures

[51] undertook a detailed investigation of published histograms of operation times experienced in real life and concluded that processing times follow a Weibull distribution (positively skewed), with an average CV value of around 0.274. This probability distribution was utilized in this study.

One way to measure how efficiently a line is working is through the calculation of the average buffer level for the whole line (ABL); obviously, we want to keep the number of unfinished pieces in storage as low as possible. Another approach is to compute the time that the line is not functioning (idle time or IT) as a percentage of total working time. This needs to be kept as low as possible as well in the interests of keeping labour costs down.

3.3 Simulation Run Parameters and Model Assumptions

The following parameter values were employed:
- Initial conditions: start the simulation run with all the buffers being nearly half-full.
- Length of the transient period: discard all accumulated statistics produced during the 5,000 time units (TU) start-up period.
- Length of the simulation run and number of observations of performance measures: use a steady-state run length of 30,000 TU, divided into 12 batches of 2,500 TU each, i.e. the mean dependent variable values are recorded every 2,500 TU and then the grand mean, representing the average of these 12 mean values, is computed. Results obtained from a trial procedure confirmed that these figures were sufficient to reduce the IT and ABL autocorrelations to the negligible values of 0.001 and 0.000 respectively.

The basic operating assumptions for the reliable unpaced serial flow line simulated are as follows:
- The first station is never starved and the last station is never blocked.
- No breakdowns occur and no defective parts are produced.
- Only one type of product flows in the system, with no changeovers.
- Time to move the work units in and out of the storage buffers is very small, hence negligible.

3.4 Unbalanced Lines Investigated

Three types of unbalanced lines with only one source of imbalance were studied. Their designs are summarized in Table 1 on the next page.

In the investigations into the effects of imbalance from single sources, the independent variables not directly under study were kept constant, whereas the variable of interest was simulated in different patterns along the line, and the outcome variable computed for each of these configurations.

7

	Source of Imbalance	MT	CV	BC / MB	Line Length (N)
SINGLE SOURCE	**a. Mean Operation Time (MT Investigation)**	**Degree of Imbalance: 2%, 5% and 12%**	CV = 0.274	1, 2 and 6 units equal for each station	5 and 8 stations
	b. Variability (CV Investigation)	10 units/station	**CV = 0.08; 0.27 and 0.50**	1,2 and 6 units equal for each station	5 and 8 stations
	c. Buffer Capacity (BC Investigation)	10 units/station	CV = 0.274	**MB 2 and 6 allocated unevenly**	5 and 8 stations

Table 1. Line designs studied (source of variability **in bold)**

For all three of the investigations, the line length was specified at two values, a shorter 5-station line (N = 5) and a longer 8-station line (N = 8).

The single source lines were those imbalanced in terms of:

a. Unequal service time means (MT investigation): each station in this study has the same CV value of 0.274 and all the buffer capacities are equal. Buffer capacity (BC) was set at 1, 2 and 6 units. Mean time imbalance patterns were simulated with degrees of imbalance at 2%, (very low imbalance), 5% (relatively low imbalance), and 12% (higher imbalance).

b. Unbalanced coefficients of variation (CV investigation): each station has identical mean operation time value (MT), held constant at 10 units per workstation for both line lengths, and buffer capacity was 1, 2 and 6 units. The coefficients of variation were defined as steady, S (CV = 0.08), medium, M (CV = 0.27) and variable, V (CV = 0.5).

c. Uneven buffer sizes (BC investigation): each station has the same mean processing time and CV values (10 time units and 0.274, respectively), but buffer sizes are unequal. Mean buffer values were 2 and 6 units, allocated unevenly between stations.

3.5 Patterns of Imbalance

In this section we will go into more detail with respect to the specific patterns of imbalance that are studied here.

In the study where mean operation times are imbalanced (MT Investigation), four patterns of imbalance are considered:
- A monotone increasing order (/).
- A monotone decreasing order (\).
- A bowl arrangement (V).
- An inverted bowl shape (^).

When coefficients of variability were imbalanced (CV Investigation), four CV imbalance policies were simulated; these were as follows:

8

- Separating the variable stations from one another by steadier stations (patterns P1- P3 portray this policy).
- Assigning steadier stations to the line centre, i.e. a bowl arrangement (patterns P4 and P5 depict this policy).
- The stations with medium variability are allocated to the middle of the line. This policy represents both a decreasing order (pattern P6) and an increasing sequence (pattern P7) of CVs along the line.
- The most variable stations are assigned to the centre of the line centre - an inverted bowl arrangement (pattern P8).

These CV imbalance policies are illustrated in Table 2 below:

Pattern (Pi) of Unbalanced CVs	Line Length (N)	
	5	8
P1	MSVMS	MSVMSVMS
P2	VMSVM	VMSVMSVM
P3	SMVSM	SMVSMVSM
P4	MSSSV	MMSSSSVV
P5	MSSSV	MMSSSSVV
P6	VMMMS	VVMMMMSS
P7	SMMMV	SSMMMMVV
P8	MVVVS	MMVVVVSS

Table 2: Unbalanced CV patterns S: (CV = 0.08), M (CV = 0.27), V: (CV = 0.50)

In the case of buffer capacity imbalance (BC Investigation), four policies were also explored for total buffer capacity allocation – these can be described as:
- Concentrating available capacity closer to the end of the line. This policy displays an increasing order of BC (pattern A).
- Concentrating buffer capacity nearer the middle of the line. This policy portrays an inverted bowl BC sequence (pattern B).
- Concentrating capacity towards the beginning of the line. This policy shows a decreasing order of BC (pattern C).
- No concentration. This policy is broken into three main sub-policies:
    - General (pattern D1).
    - Alternating BC between high and low along the line (pattern D2).
    - Positioning smaller BC towards the centre - a bowl shape (pattern D3).

These policies are displayed on the next page in Table 3.

9

Line Length (N)		5		8	
Mean Buffer Size (MB)		2	6	2	6
P1	A	1,1,1,5	3,3,3,15	1,1,1,1,6,2,2	3,3,3,3,18,6,6
P2	B	1,1,5,1	3,3,15,3	1,1,6,2,2,1,1	3,3,18,6,6,3,3
P3	C	5,1,1,1	15,3,3,3,	6,2,2,1,1,1,1	18,6,6,3,3,3,3
P4	D1	2,2,3,1	6,6,9,3	2,2,2,3,3,1,1	6,6,6,9,9,3,3
	D2	2,3,2,1	6,9,6,3	2,2,3,3,2,1,1	6,6,9,9,6,3,3
	D3	2,1,3,2	6,3,9,6	2,2,1,1,3,3,2	6,6,3,3,9,9,6
Total Buffer Capacity (TB)		8	24	14	42

Table 3. Unequal buffer size patterns (Pi = policy of buffer capacity imbalance)

4. RESULTS

Due to space limitations, only IT and ABL results for the best, second best or good, and the worst patterns will be shown.

4.1 Idle Time Results

4.1.1 IT Data

Tables 4-7 exhibit IT data for a number of unbalanced and balanced line configurations for the MT, CV and BC investigations:

Buffer Size		1			2			6		
% Imbalance Degree		2	5	12	2	5	12	2	5	12
Pattern of Unbalanced Means	\	9.975	11.181	14.978	5.065	7.302	13.247	2.586	5.443	12.010
	/	9.740	10.221	14.705	5.396	6.611	12.585	2.378	5.349	11.815
	∧	9.996	10.141	12.335	5.442	6.116	9.335	2.381	2.928	7.902
	V	**9.401**	**9.551**	**10.132**	**4.941**	**5.074**	**6.528**	**2.060**	**2.537**	**4.995**
Balanced Line		9.522			4.985			2.066		

Table 4. MT investigation: IT data for patterns \, /, ∧, V and the balanced line (5 stations)

Buffer Size		1			2			6		
% Imbalance Degree		2	5	12	2	5	12	2	5	12
Pattern of Unbalanced Means	\	11.780	12.691	16.551	6.368	7.486	13.479	2.969	6.003	12.802
	/	11.469	12.429	16.173	6.413	7.943	13.068	3.293	5.366	11.902
	∧	11.251	12.269	13.205	5.837	7.154	9.347	2.589	3.421	7.909
	V	**11.123**	**11.239**	**11.774**	**5.772**	**6.080**	**7.357**	**2.156**	**2.584**	**5.658**
Balanced Line		11.522			5.935			2.174		

Table 5. MT investigation: IT data for patterns \, /, ∧, V and the balanced line (8 stations)

10

Line Length		5			8		
Buffer Size		1	2	6	1	2	6
Pattern of Unbalanced CVs	P2	11.019	6.656	2.634	14.737	8.409	2.901
	P4	**8.554**	**4.409**	**1.176**	**12.937**	**8.406**	**2.896**
	P5	8.704	4.679	2.036	12.977	9.302	3.799
	P6	18.295	11.486	5.283	20.578	12.728	6.624
Balanced Line		9.522	4.985	2.066	11.522	5.935	2.174

Table 6. CV investigation**:** IT data for patterns P2, P4-P6 and the balanced line.

Line Length		5		8	
Mean Buffer Size		2	6	2	6
Pattern of Buffer Size Imbalance	A1	9.104	3.499	10.982	4.123
	A2	7.978	3.094	9.315	3.571
	B1	7.532	2.696	9.143	3.599
	C1	9.009	3.149	10.872	4.088
	D1	5.707	2.082	**7.112**	**2.212**
	D2	**5.598**	**1.733**	7.497	3.492
Balanced Line		4.985	2.066	5.935	2.174

Table 7. BC investigation: IT data for patterns A1-A2, B1, C1, D1-D2 and   the balanced line

4.1.2 Ranking of Policies and Patterns

From Tables 4-7, the following observations can be made:

MT investigation:
- The results indicate that the bowl arrangement (V) is the best unbalanced pattern, followed by the inverted bowl pattern (^).
- A decreasing mean pattern (\) is the worst.
- Pattern (/) is generally better than pattern (\), with the difference in IT being slight.

CV investigation:
- It is not possible to discern one overall policy as being the best or the worst with respect to all constituent patterns.
- Pattern P4, a bowl arrangement, is the best overall pattern, followed by P5 (the second bowl pattern)
- Configuration P6 is the worst.

BC investigation:
- None of the four policies can be regarded as the best or the worst in terms of all its constituent patterns.

11

- D2 (N = 5) and D1 (N = 8) are the best unbalanced patterns, i.e. the best configuration is one where the available capacity is distributed as uniformly as possible along the line.
- D1 (N = 5) and D2 (N = 8) can be deemed as good configurations.
- A1 (the increasing order arrangement) is the worst.
- Within policies 1, 2, and 3, patterns A1, B1, C1 are respectively the worst; lending support to the strategy of avoiding extreme allocation of TB (i.e. most TB is assigned to one buffer and the rest to the other buffers).
- The descending order policy may be considered as generally the best alternative if a balanced, or close to balanced buffer arrangement is not feasible.

4.1.3 Effects of the Independent Variables on IT

The simulation results show the following relationships between the design factors and idle time:

MT investigation:
- An increase in N causes a corresponding rise in IT, particularly for smaller levels of BC in the case of the best pattern.
- As BC increases IT decreases, with the rate of decrease for the best pattern slowing down as BC and N continue to increase.
- IT goes up as DI becomes higher, with the rate of increase accelerating for larger BC values.

CV investigation:
- IT increases with N especially the lower BC is in the case of the best pattern.
- When BC is expanded IT goes down.

BC investigation:
- As N increases, IT tends to increase. For the best pattern, the IT tendency to increase with N is more substantial the lower MB is.
- IT decreases as MB goes up. The marginal decrease in IT for the best pattern falls as both MB and N increase.

4.1.4 ANOVA

ANOVA outcomes for the MT, CV and BC investigations show that all of the main effects are highly significant at the 99% confidence level and that all of the interactions are significant at the 95% level or above. The sub-run (batch size) effect appears to be non-significant, lending support to the contention that all the data represent the steady state condition. The ranking of the variables influencing IT are as follows:

MT investigation: the most important factor affecting IT is DI. The second and third factors are BC and MT pattern, respectively.

CV investigation: the strongest effect on IT comes from BC, whereas the pattern of CV imbalance has a lesser impact.

BC investigation: The main independent variable influencing IT is MB, followed by the

12

imbalance pattern

4.2 Average Buffer Level Results.

4.2.1 ABL Data

Tables 8-11 show ABL data for various unbalanced and balanced line configurations for the MT, CV and BC investigations:

Buffer Size		1			2			6		
% Imbalance Degree		2	5	12	2	5	12	2	5	12
Pattern of Unbalanced Means	/	0.591	0.673	0.814	1.143	1.420	1.724	4.197	4.953	5.047
	\	**0.468**	**0.398**	**0.249**	**0.920**	**0.536**	**0.291**	**1.672**	**0.991**	**0.413**
	V	0.533	0.560	0.588	1.032	1.024	1.046	2.504	2.690	2.666
	∧	0.542	0.526	0.542	1.272	1.018	1.005	2.976	2.946	2.961
Balanced Line		0.526			1.033			3.321		

Table 8. MT investigation: ABL data for patterns \, /, ∧, V and the balanced line (5 stations)

Buffer Size		1			2			6		
% Imbalance Degree		2	5	12	2	5	12	2	5	12
Pattern of Unbalanced Means	/	0.614	0.698	0.812	1.280	1.433	1.683	4.324	4.804	5.545
	\	**0.503**	**0.406**	**0.284**	**0.827**	**0.641**	**0.347**	**2.186**	**1.049**	**0.451**
	V	0.569	0.554	0.575	1.014	1.003	0.971	2.760	2.752	3.250
	∧	0.552	0.548	0.544	1.089	1.084	1.046	3.021	3.020	2.992
Balanced Line		0.559			0.970			2.601		

Table 9. MT investigation: ABL data for patterns \, /, ∧, V and the balanced line (8 stations)

Line Length		5			8		
Buffer Size		1	2	6	1	2	6
Pattern of Unbalanced CVs	P4	**0.397**	**0.638**	**2.433**	**0.285**	**0.464**	**1.203**
	P5	0.668	1.373	3.618	0.722	1.575	4.756
	P6	0.503	0.992	2.883	0.506	0.941	2.816
	P8	0.674	1.271	3.719	0.755	1.452	4.548
Balanced Line		0.526	1.033	3.321	0.559	0.970	2.601

Table 10. CV investigation: ABL data for patterns P4, P5, P6, P8 and the balanced line.

13

Line Length		5		8	
Mean Buffer Size		2	6	2	6
Pattern of Buffer Size Imbalance	A1	**0.531**	**1.457**	0.531	1.444
	A2	0.534	1.679	**0.513**	**1.417**
	A3	0.615	1.733	0.514	1.429
	D2	1.218	2.792	1.411	4.248
Balanced Line		1.033	3.321	0.970	2.601

Table 11. BC investigation: ABL data for patterns A1-A3, D2 and the balanced line

4.2.2 Ranking of Policies and Patterns

Based on Tables 8-11 it is interesting to note the following:

MT investigation:
- The best pattern turns out to be the descending (\) order.
- The second and third best patterns being respectively, an inverted bowl (∧) and a bowl (V) arrangements.
- The worst pattern is the increasing (/) configuration.

CV investigation:
- Since each policy includes a number of patterns with varying performance, it is impossible to label one specific policy as either the best or the worst. Specific configurations within the policies however, may be regarded as superior or inferior.
- The bowl-shaped pattern P4 is the best for ABL. This is the same pattern that gave the best results for IT.
- The worst patterns are generally P5 and P8.

BC investigation:
- Again, no one policy can be said to be the best or the worst.
- The best pattern is A1 for N = 5, and A2 for N = 8, i.e. the best pattern has its buffer capacity concentrated towards the end of the line (an ascending order).
- When N = 5, A2 is the second best pattern and when N = 8, A3 is the second best configuration.
- In general, configuration D2 (the random arrangement) can be viewed as the worst.

4.2.3 Effects of the Independent Variables on ABL

As regards the impact of the exogenous variables on average buffer level, the simulation data indicate the following relationships:

MT investigation:
- ABL becomes higher as BC is increased. This increase in ABL continues at a diminishing rate as BC rises and as N is reduced.
- As DI increases ABL falls, with the drop in ABL becoming less marked as DI continues to go up and as BC decreases.

14

CV investigation:
- As BC rises, so does ABL.

BC investigation:
- ABL increases with MB.

It should be noted that in the three investigations the influence of line length seems not to be important; there is no directly observable pattern of change of ABL levels with N.

4.2.4 ANOVA.

The analysis of variance of the simulated ABL data for the MT, CV and BC investigations produced the same findings as those found in section 4.1.4 on IT results. The rankings of the independent parameters influencing ABL are as follows:

MT investigation: the most important factor affecting ABL is BC, followed respectively by MT pattern and DI.

CV investigation: as was the case for idle times, the most salient variable influencing ABL is BC, followed by the CV pattern.

BC investigation: the most significant exogenous factor impacting ABL is MB, followed by the pattern of buffers.

5. BEST UNBALANCED PATTERNS' SAVINGS OVER THE BALANCED LINE

In the MT investigation, the most favourable unbalanced MT pattern in terms of IT was a bowl configuration (V). Table 12 (see next page) shows the percentage differences in this pattern's IT vis-a-vis the balanced line (the control).

From Table 12, the following can be observed:
- The highest improvement in IT for the most favourable configuration over the balanced line is 3.46%
- An increase in DI either reduces or immediately eliminates the advantage in IT, especially for higher BC values.
- As BC increases, this advantage either instantly or gradually disappears.
- When N becomes larger, the advantage goes up in magnitude.

On the other hand, the best pattern in ABL is a decreasing MT order (\). Table 13 (see next page) exhibits the percentage savings in the best configuration's ABL over the balanced line:

The following can be discerned from Table 13:

- The highest saving in ABL for the best pattern over the balanced line is 87.56%. So, placing the fastest operators at the end of the line can bring considerable advantages in terms of ABL performance.
- The best pattern outperformed the balanced line for all line lengths, buffer capacities and degrees of imbalance in mean times.

- As DI goes up, the superiority of the best pattern over the control increases.

15

- Increasing BC has the effect of raising the best pattern's advantage.

Line Length = 5			Line Length = 8		
Buffer Size	% Imbalance Degree	% Difference	Buffer Size	% Imbalance Degree	% Difference
1	2	-1.27	1	2	**-3.46**
1	5	0.31	1	5	-2.46
1	12	6.41	1	12	2.18
2	2	-0.88	2	2	-2.74
2	5	1.79	2	5	14.50
2	12	30.95	2	12	23.90
6	2	-0.31	6	2	-0.82
6	5	22.80	6	5	18.86
6	12	141.77	6	12	160.26

(-) indicates saving
Table 12. MT investigation: % differences in the best pattern's IT over the control

Line Length = 5			Line Length = 8		
Buffer Size	% Imbalance Degree	% Saving	Buffer Size	% Imbalance Degree	% Saving
1	2	-11.30	1	2	-10.02
1	5	-25.29	1	5	-27.37
1	12	-52.66	1	12	-49.20
2	2	-10.94	2	2	-14.74
2	5	-48.11	2	5	-33.92
2	12	-71.83	2	12	-64.23
6	2	-49.65	6	2	-15.96
6	5	-70.16	6	5	-59.67
6	12	**-87.56**	6	12	-82.66

(-) indicates saving
Table 13. MT investigation: % savings in the best pattern's ABL over the control

For the CV investigation, the best overall results for reducing both idle times and average buffer levels came from pattern 4, one of the two bowl shaped patterns considered, with the steadier workers in the middle. A summary of the % differences in IT and ABL are shown for the best pattern in Table 14 (see next page).

16

Line Length	5			8		
Buffer Size	1	2	6	1	2	6
% Difference in IT	-10.17	-11.56	**-43.08**	12.28	41.63	33.21
% Saving in ABL	-24.53	-38.24	-26.74	- 49.02	-52.17	**-53.75**

(-) indicates saving

Table 14. CV investigation: % differences in the best pattern IT and ABL compared to the control

From Table 14, the following points can be noted:

- The best IT saving over the balanced line (-43.08%) and the best ABL superiority (-53.75%) represent considerable savings.
- The improvements in IT disappear as the line lengthens, whereas the savings in ABL increase with the number of workstations.
- The unbalanced line is consistently superior to the balanced line for the ABL results across the board.
- There is no consistent trend which appears with the increase in buffer levels.

As regards the BC investigation, Table 15 summarises the % differences in IT and ABL for the best unbalanced patterns in comparison with those of the balanced line:

Line Length	5		8	
Mean Buffer Capacity	2	6	2	6
% Difference in IT	12.30	**-16.14**	19.83	1.75
% Saving in ABL	-48.60	**-56.13**	-47.11	-45.52

(-) indicates saving

Table 15. BC investigation: % savings in the best pattern's ABL over the control

As is exhibited in Table 15 above, the following can be concluded:

- Pattern D2 (no concentration of available buffer capacity) shows a reduction in IT of 16.14% as compared to the balanced line, whereas for configuration A1 (buffer capacity is concentrated towards the end of the line), the savings obtained in ABL are considerable (over 56%).
- As N increases, any saving in IT disappears while ABL's advantage declines.
- The best pattern consistently exhibits significantly lower ABL levels over the balanced line for all factor levels considered.

6. SUMMARY

Several unbalancing policies and methods were examined in three single-source imbalance investigations. None of the policies were noticeably better or worse than any of the others in broad terms, but there were particular patterns within each policy that showed improvements of performance either in idle time or in average buffer level when compared to the balanced line counterpart. Table 16 (see next page) summarizes the best performing configurations in terms of IT and ABL for the three investigations:

17

Best Performance Configurations	Idle Time	Average Buffer Level
Mean Operation Times (MT)	Bowl pattern	Descending order
Variability (CV)	Bowl pattern	Bowl pattern
Buffer Capacity (BC)	Close to balance	Ascending order

Table 16. The influence of single source imbalance patterns on idle time and average buffer levels

It should be noted from Table 16 above that for the CV investigation the best pattern in terms of both idle time and average buffer level results is the bowl shaped pattern. This is of great interest to those manufacturers who, within the constraints of lean buffering need to keep down buffer levels and increase output rates at the same time.

It was found that for all the three investigations, as BC/MB increases, IT goes down but ABL rises, suggesting that BC/MB exerts an opposite influence on IT and ABL. In the case of the MT investigation, increasing DI raises IT, but ABL declines. Also, when DI is increased, the % saving in IT over the balanced line decreases, but in terms of ABL it increases.

In addition, it was observed that the best configurations result in a significantly smaller ABL levels than those of corresponding balanced lines at all the factor levels considered.

As regards the influence of the various factors on IT and ABL, Table 17 below summarizes ANOVA findings:

Investigation	Performance Measure	Factor 1	Factor 2	Factor 3
MT	IT	DI	BC	MT pattern
CV	IT	BC	CV pattern	----
BC	IT	MB	BC pattern	----
MT	ABL	BC	MT pattern	DI
CV	ABL	BC	CV pattern	----
BC	ABL	MB	BC pattern	----

Table 17. Ranking of the independent variables in their effects on IT and ABL

Moreover, the best patterns found have resulted in various degrees of savings over the balanced line counterpart for the three investigations, as is shown below in Table 18.

Investigation	% Saving in IT	% Saving in ABL
MT	-3.46	-87.56
CV	-43.08	-53.75
BC	-16.14	-56.13

(-) indicates saving

Table 18. Highest obtained % savings in IT and ABL

7. DISCUSSION AND CONCLUSIONS

The main purpose of this study was to assess the effects that unbalancing service mean times, coefficients of variation, or buffer sizes have on the efficiency of a production line. One of the main conclusions of this research is that the decision of how to allocate different sized buffers between workstations, where to place operators with different average working times and variability will depend on the particular conditions of the production facilities.

It may be a priority to keep the amount unfinished goods in storage as low as possible, for example fresh produce where hygiene and safety issues are important. In this case, a manager would opt for reductions in average buffer levels. To do this, one might allocate more buffer capacity to the end of the line. If worker average times are known to differ, it could be advantageous to put the fastest workers towards the end, and when workers vary in their average speeds to a great degree, one might consider placing the steadiest workers in the middle. This is especially the case where just-in-time and lean buffering strategies are in place where operations managers are facing enormous pressure to reduce expensive inventory, and so to decrease production lead times.

In contrast, if we are looking at a sector where labour costs are high, for example the automobile industry, then it could be advantageous to move towards bringing idle time down and either distributing buffer capacity as evenly as possible along the line or again considering placing faster workers towards the middle.

We should remember, however, that the best and other good patterns are specific patterns among numerous possibilities, and that imbalance directed in the wrong way could lead to the opposite effect, i.e. increases in average buffer levels and/or idle times.

Companies spend billions of dollars every year on the design, installation, operation, and maintenance of production lines. Even the slightest improvement in efficiency or reduction in inventory costs can result in substantial savings over the lifespan of a line. Since a balanced line is virtually unattainable in practice and that most lines suffer from a certain degree of imbalance, it would make sense for production managers to examine the benefits of deliberately unbalancing their lines in the right way, particularly as unbalancing can be done at no extra cost.

The study showed that in many cases substantially superior performance to that of the balanced line in terms of IT or ABL, or sometimes both is attainable (see Table 18 above).

The scale of the potential reductions in IT and ABL, when calculated over the lifecycle of a production line means that purposely unbalancing the buffer sizes and operators with different variability and speeds could lead to real benefits for the manufacturer and so might be a strategy to take into account when designing the production line.

It is hoped that this research has contributed to the total body of knowledge of production lines in giving new insights into how to fine tune unbalanced lines in order to improve performance. There is still scope for a considerable amount of research based on this study, for example experiments on the effects of the single source imbalance for merging (assembly) lines as well as for unreliable lines, will enrich knowledge in this area and give managers a more accurate picture when designing their production lines.

19

REFERENCES

[1] Hillier, F.S. & Boling, R.W. (1966) *The Effect of Some Design Factors on the Efficiency of Production Lines with Variable Element Times*. Journal of Industrial Engineering, 1966, Vol. 17, No. 12, pp. 651-658.

[2] Hillier, F. S. & Boling, R. W. *Toward Characterising the Optimal Allocation of Work in Production Line Systems with Variable Operation Times*. Advances in Operations Research, Proceedings of EURO II, 1977, Roubens, M. (ed.), North-Holland, pp.649-658.

[3] Hillier, F.S. & Boling, R.W. *On the Optimal Allocation of Work in Symmetrically Unbalanced Production Line Systems with Variable Operation Times*. Management Science, 1979, Vol. 25, No. 8, pp.721-728.

[4] Yamazaki, G., Sakasegawa, H. & Shanthikumar, J. G. *On Optimal Arrangement of Stations in a Tandem Queuing System with Blocking*. Management Science, 1992, Vol. 38, No.1, pp. 137-153.

[5] Hillier, F. S. & So, K. C. (1996), *On the Robustness of the Bowl Phenomenon*. European Journal of Operational Research, 1996, Vol. 89, No. 3, pp. 496-515.

[6] Pike, R. & Martin, G. E. *The Bowl Phenomenon in Unpaced Lines*. International Journal of Production Research, 1994, Vol. 32, No. 3, pp. 483-499.

[7] Baker, K. R., Powell, S. G. & Pyke, D. F. *A Predictive Model for the Throughput of Unbalanced, Unbuffered Three-Station Serial Lines*. IIE Transactions, 1994, Vol. 26, No. 4, pp. 63-71.

[8] Liu, C. M. & Lin, C. L. *Performance Evaluation of Unbalanced Serial Production Lines*. International Journal of Production Research, 1994, Vol. 32, No. 12, pp. 2897-2914.

[9] Tempelemeier, H. & Burger, M. *Performance Evaluation of Unbalanced Flow Lines with General Distributed Processing Times, Failures and Imperfect Production*. IIE Transactions, 2001, Vol. 33, No. 4, pp. 293-302.

[10] Lambrecht, M. and Segaert, A., (1990), 'Buffer Stock Allocation in Serial and Assembly Type of Production Lines'. *International Journal of Production Management*, Vol. 10, No. 2, pp. 47-61.

[11] Cook, D. P. *Simulation Comparison of Traditional, JIT, and TOC Manufacturing Systems in a Flow Shop with Bottlenecks*. Production and Inventory Management Journal, 1994, Vol. 35, No. 1, pp. 73-78.

[12] Chakravorty, S. S. & Atwater, J. B. *A Comparative Study of Line Design Approaches for Serial Production Systems*. International Journal of Operations & Production Management, 1996, Vol. 16, No. 6, pp. 91-108.

[13] Atwater, B. J. & Chakravorty, S. S. *Does Protective Capacity Assist Managers in Competing Along Time-Based Dimensions?* Production and Inventory Control Management Journal, 1994, Vol. 35, No.3, pp. 53-59.

20

[14] Kadipasaoglu, S. N., Xiang, W., Hurley, S. F. & Khumawala, B. M. *A Study on the Effect of the Extent and Location of Protective Capacity in Flow Systems*. International Journal of Production Economies, 2000, Vol. 63, No. 3, pp. 217-228.

[15] Sloan, W. L. '*A Study on the Effects of Protective Capacity on Cycle Time in Serial Production Lines*'. M.Sc. Thesis, Mississippi State University, 2001

[16] Gilland, W. G. *A Simulation Study Comparing Performance of CONWIP and Bottleneck – Based Release Rules*. Production Planning & Control, 2002, Vol. 13, No. 2, pp. 211-219.

[17] Koh, S. G. & Bulfin, R. L. *Comparison of DBR with CONWIP in an Unbalanced Production Line with Three Stations*. International Journal of Production Research, 2004, Vol. 42, No. 2, pp. 391-404.

[18] Mabin, V. J. & Balderstone, S. J. *The Performance of the Theory of Constraints Methodology*. International Journal of Operations and Production Management, 2003, Vol. 23, No. 6, pp. 568-595.

[19] Wu, H. H. & Yeh, M. L. *A DBR Scheduling Method for Manufacturing Environments with Bottlenecks Re-entrant Flows*. International Journal of Production Research, 2006, Vol. 44, No. 5, pp. 883-902.

[20] Anderson, D. R. *Transient and Steady State Minimum Cost In-Process Inventory Capacities for   Production Lines*. Ph.D. Thesis, Purdue University 1968.

[21] Payne, S., Slack, N. & Wild, R. *A Note on the Operating Characteristics of Balanced and Unbalanced Production Flow Lines*. International Journal of Production Research, 1971, Vol. 9, No. 2, pp. 93-98.

[22] Tembe, S.V. & Wolff, R.W. *The Optimal Order of Service in Tandem Queues. Operations Research*, 1974, Vol. 22, No. 3, pp. 824-832.

[23] Shimshak, D.G. *A Study of Queues in Series*. PhD. Thesis, The City University of New York. 1976

[24] Carnall, C.A. & Wild, R. *The Location of Variable Work Stations and the Performance of Production Lines*. International Journal of Production Research, 1976, Vol. 4, No. 6, pp.703-710.

[25] El-Rayah, T. *The Efficiency of Balanced and Unbalanced Production Lines*. International Journal of Production Research, 1979b, Vol.17, No.1, pp.61-75.

[26] De La Wyche, P. & Wild, R. *The Design of Imbalanced Series Queue Flow Lines*. Operational Research Quarterly, 1977, Vol. 28, No. 3, ii, pp.695-702.

[27] Lau, H-S. *On Balancing Variances of Station Processing Times in Unpaced Lines*. European Journal of Operational Research, 1992, No. 56, pp. 345-356.

[28] Liu, C.M., SU, S.-F. & Lin C.L. *Predictive Models for Performance Evaluation of Serial*

21

Production Lines. International Journal of Production Research, 1996, Vol. 34, No. 5, pp. 1279-21291.

[29] Yamashita, H. & Altiok, T. *Buffer Capacity Allocation for a Desired Throughput in Production Lines*. IIE Transactions, 1998, Vol. 30, No. 10, pp. 883-891.

[30] Harris, J.H. & Powell, S.G. *An Algorithm for Optimal Buffer Placement in Reliable Serial Lines*. IIE Transactions, 1999, Vol. 31, No. 4, pp. 287-302.

[31] Chiadamrong, N. & Limpasontipong, P. *Using Storage Buffer to Improve Unbalanced Asynchronous Production Flow Line's Performance*. International Journal of Manufacturing Technology and Management, 2003, Vol. 5, Nos. 1/2, pp. 149-161.

[32] El-Rayah, T. *The Effect of Inequality of Inter-stage Buffer Capacities and Operation Time Variability on the Efficiency of Production Line Systems*. International Journal of Production Research, 1979a, Vol. 17, No. 1, pp.77-89.

[33] Chow, W.M. *Buffer Capacity Analysis for Sequential production Lines with Variable Process Times*. International Journal of Production Research, 1987, Vol. 25, No. 8, pp. 1183-1196.

[34] Conway, R., Maxwell, W., McClain, J. & Thomas, L. J. *The Role of Work in Progress Inventory in Serial Production Lines*. Operations Research, 1988, 36(2), 229-241.

[35] Hillier, F. S. & So, K.C. *The Effect of the Coefficient of Variation of Operation Times on the Allocation of Storage Space in Production Line Systems*. IIE Transactions, 1991, Vol. 23, No. 2, pp. 198-206.

[36] Hillier, F. S., So, K. C. & Boling, R. W. *Notes: Toward Characterizing the Optimal Allocation of Storage Space in Production Line Systems with Variable Processing Times*. Management Science, 1993, Vol. 39, No. 1, pp. 126-133.

[37] Powell, S. G. *Buffer Allocation in Unbalanced Three – Station Serial Lines*. International Journal of Production Research, 1994, Vol. 32, Nos. 9, pp. 2201-2217.

[38] Hillier, F. S. & So, K. C. *On the Optimal Design of Tandem Queuing Systems with Finite Buffers*. Queuing Systems, 1995, Vol. 21, pp. 245–266.

[39] Powell, S. G. & Pyke, D. F. *Allocation of Buffers to Serial Production Lines with Bottlenecks*. IIE Transactions, 1996, Vol. 28, No. 1, pp. 18-29.

[40] Gershwin, S.B. & Schor, J.E. *Efficient Algorithms for Buffer Space Allocation*. Annals of Operations Research, 2000, Vol. 93, pp. 117–144.

[41] Helber, S. *Cash-Flow Oriented Buffer Allocation in Stochastic Flow Lines*. International Journal of Production Research, 2001, Vol. 39, pp. 3061-3083.

[42] Spinellis, D. D. & Papadopoulos, C. T. *A Simulated Annealing Approach for Buffer Allocation in Reliable Production Lines*. Annals of Operations Research, 2000, Vol. 92, pp. 373-384.

22

[43] Kim, S. & Lee, H. *Allocation of Buffer Capacity to Minimize Average Work in Progress.* Production Planning & Control, 2001, Vol. 12, No. 7, pp. 706-716.

[44] Papadopoulos, H. T. & Vidalis, M. I. *Minimizing WIP Inventory in Reliable Production Lines*. International Journal of Production Economics, 2001, Vol. 70, pp. 185-197.

 [45] Chan, F. T. S. & Ng, E. Y. H. *Comparative Evaluations of Buffer Allocation Strategies in a Serial Production Line*. International Journal of Advanced Manufacturing Technology, 2002, Vol. 19, No. 11, pp. 789-800.

[46] Huang, M. G., Chang, P. L. & Chou, Y. C. *Buffer Allocation in Flow-Shop-Type Production Systems with General Arrival and Service Patterns*. Computers and Operations Research, 2002, Vol. 29, pp. 103–121.

[47] Hillier, M. S. *Characterizing the Optimal Allocation of Storage Space in Production Line Systems with Variable Processing Times*. IIE Transactions, 2000, Vol. 32, pp. 1-8.

[48] Hillier, M. S. & Hillier, F. S. (2006), *Simultaneous Optimization of Work and Buffer Space in Unpaced Production Lines With Variable Processing Times*. IIE Transactions, 2006 Vol. 38, pp. 39-51.

[49] Grosfeld-Nir, A. & Magazine, M. *A Simulation Study of Pull Systems with Ascending/Descending Buffers and Stochastic Processing Times*. International Journal of Production Research, 2005, Vol. 43, pp. 3529–3541.

 [50] Vergara, H.A. & Kim, D.S. *A New Method for the Placement of Buffers in Serial Production Lines*. International Journal of Production Research, 2009, Vol. 47, pp. 4437-4456.

[51] Slack, N. *Work Time Distributions in Production System Modelling*, Research Paper, Oxford Centre for Management Studies, 1982

THE OPTIMAL PARAMETERS OF YIELD DISTRIBUTION IN AN EMQ MODEL WITH BACKORDERS

Javad Paknejad, Frank G. Zarb School of Business, Hofstra University, Hempstead, NY 11590, 516-463-4244, M.J.Paknejad@hofstra.edu

Farrokh Nasri, Frank G. Zarb School of Business, Hofstra University, Hempstead, NY 11590, 516-463-4244, Farrokh.Nasri@hofstra.edu

John F. Affisco, Frank G. Zarb School of Business, Hofstra University, Hempstead, NY 11590, 516-463-4244, John.F.Affisco@hofstra.edu

ABSTRACT

This paper extends the EOQ results of a recent paper [7] to the case of an EMQ model with backorders and random yield. Specifically, the paper studies the economic trade-offs associated with investment decisions aimed at improving the production yield, through changing the parameters of yield distribution. The optimal values of policy variables in a yield-adjusted EMQ model with backorders are obtained. For the purpose of this paper, yield is defined as the proportion of conforming items in a production lot. Assuming that the yield probability density function is uniform with location parameter C and Scale parameter (1-C), the paper presents explicit results for the optimal values of decision variables as well as the optimal expected annual cost for the special case of linear investment function.

Keywords: Inventory Theory, Quality

1. INTRODUCTION

The publicity of stockless production or zero inventories [5] provoked researchers to give meticulous thought to the possible relationship between lot size and yield, that is, producing or procuring order quantities of imperfect quality, when developing new lot sizing paradigms. In the beginning, Rosenblatt and Lee [14] investigated the effect of process quality on lot size in the classical economic manufacturing quantity (EMQ) model. Subsequently, Porteus [13] introduced a modified economic order quantity (EOQ) model to highlight the relationship between quality and lot size. In both [13] and [14] the optimal lot size is shown to be smaller than that of the classical EMQ and EOQ models, respectively. Furthermore, in both papers, the authors consider the demand to be deterministic. Moinzadeh and Lee [6] relaxed the deterministic demand assumption and studied the impact of defective items on the policy variables of a continuous-review inventory model with Poisson demand and constant lead time. Paknejad, Nasri, and Affisco [12] extend this work to consider general stochastic demand and constant lead time in the continuous review (s,Q) model. Paknejad, Nasri, and Affisco [11] develop a quality-adjusted EOQ model for the case where both backorders and stockouts are allowed.

Cheng [3] develops a modified Economic Production Quantity (EPQ) model by integrating quality considerations with the lot size. Using classical optimization techniques, the author presents closed form expressions for the optimal lot size and optimal expected fraction acceptable. The optimal lot size is intuitively appealing since it indicates an inverse relationship between lot size and process capability. Goyal et al [4] provides an excellent survey of the early literature on integrating lot size and quality control policies. Yano and Lee [15] present an enlightening classification and description of the early research on lot sizing in the presence of random yield.

In [11] and [12] the authors assume that the number of conforming units in each lot follows a binomial distribution, implying that the manufacturer operates a process that is in statistical control. That is, the production yield, defined as proportion of conforming items in each lot, is known and constant. Such an assumption is also made in Affisco, Paknejad, and Nasri [1] for the case of the EOQ and Affisco, Paknejad, and Nasri [2] for the case of the joint economic lot size model.

Nasri, Paknejad, and Affisco [9] investigated the relationship between order quantity and quality for processes that have not yet achieved the state of statistical control and developed a yield-adjusted EOQ model with shortages when the procurement yield is random and not constant. In [8], the authors do the same in the context of EMQ with backorders model. The results developed in [8] and [9] reinforce the wide-spread awareness that yield is a significant concern for all procurement and manufacturing organizations. However, the models in both [8] and [9] , consider the parameters of yield distribution as constants and, therefore, do not account for the possible benefits of efforts by these organizations directed at yield improvement programs, which ultimately alter the parameters of yield distribution. Nasri, Paknejad, and affisco [7] extend the work in [9] and study the impact of efforts devoted to yield improvement programs on the policy variables of a yield-adjusted EOQ model with planned shortages assuming that the shortage cost is levied based on the number of shortages per unit of time. Paknejad, Nasri, and Affisco [10] extended the results in [7] by combining two distinct shortage costs: (1) based on the number of shortages per unit time, and (2) based on the average number of shortages, irrespective of duration of shortage.

One of the implicit assumptions of the models developed in [7] and [10], for the purpose of this paper, is that items are ordered from an outside vendor, which implies instantaneous replenishment of items. When items are produced internally, rather than being ordered from an outside vendor, the Economic Manufacturing Quantity (EMQ) model with finite production rate is frequently used in determining the optimal production lot size. In this paper we consider the case where items are manufactured internally at a finite production rate and extend the EOQ results in [7] to the case of the classic EMQ model. The paper presents the optimal values of decision variables in closed forms for the specific case of uniform probability density function for yield distribution and linear investment function for changing the location and scale parameters of yield distribution.

2. MODEL AND ASSUMPTIONS

Consider the classical undiscounted, deterministic, single item Economic Manufacturing Quantity (EMQ) model that allows shortages with the following total annual cost function

$$TC(S,Q) = \frac{D}{Q}K + \frac{\left[Q\left(1-\frac{D}{P}\right)-S\right]^2}{2Q\left(1-\frac{D}{P}\right)}C_h + \frac{S^2}{2Q\left(1-\frac{D}{P}\right)}C_b \qquad (1)$$

where
D = Annual Demand in units,
P = Annual production rate in units, where $(1-\theta)P - D > 0$,
Q = Lot size per setup,
S = Number of units backordered,
K = Setup cost per setup,
c_h= Non-defective holding cost per unit per year,
c_b= Backordering cost per unit per year.

The results of classical optimization yields the following well-known expressions for the optimal values for the lot size, Q^*, units backordered, S^*, and the annual cost, $C^*(S,Q)$

$$Q^* = \sqrt{\frac{2DK}{c_h\left(1-\frac{D}{P}\right)}\left(\frac{c_h+c_b}{c_b}\right)} \qquad (2)$$

$$S^* = \sqrt{\frac{2DK}{c_b}\left(1-\frac{D}{P}\right)\left(\frac{c_h}{c_h+c_b}\right)} \qquad (3)$$

and

$$C^*(S,Q) = \sqrt{2DKc_h\left(1-\frac{D}{P}\right)\left(\frac{c_b}{c_h+c_b}\right)} \qquad (4)$$

Implicit in these derivations is that all units produced by the production process are of acceptable quality. Now assume this is not the case. Specifically, assume that each production lot size contains a random proportion of defective units. Using a hundred percent inspection policy, each item is inspected immediately after production at a fixed inspection cost per unit. We further assume that the inspection process is perfect, and that all rejected items are sold at a secondary market that will pay salvage equivalent to manufacturing plus inspection costs. Based on this scenario, we now adjust the EMQ with planned shortages model for the quality factor as follows (see Figure 1.)

Let:

λ = Yield, being defined as the proportion of non-defective items in an order lot, $\lambda \in [0, 1]$, a continuous random variable,

f(λ) = Probability density function of λ,

E(λ) = First moment of λ,

E(λ^2) = Second moment of λ,

y = λQ = Number of non-defective items in a lot,

c(y) = Total cost per cycle given that there are y non-defective items in a lot of size Q,

T = y/D = Cycle time, time between two successive placement of orders,

E(.) = Mathematical expectation,

Figure 1: The behavior of the EMQ model with Random Defective units
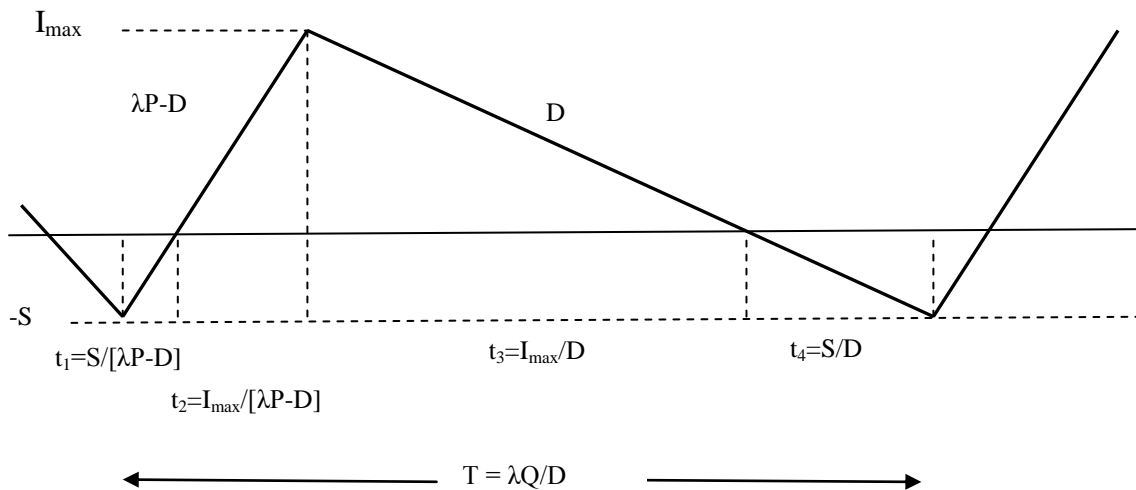
The total cost per cycle is

$$c(y) = K + \frac{\left[Q\left(\lambda - \dfrac{D}{P}\right) - S \right]^2}{2Q\left(\lambda - \dfrac{D}{P}\right)} . c_h T + \frac{S^2}{2Q\left(\lambda - \dfrac{D}{P}\right)} . c_b T$$

(5)

$$= K + \left[\frac{\lambda^2 Q^2}{2D} - \frac{\lambda Q}{2P} - \frac{\lambda Q S}{D} \right] c_h + \frac{(c_h + c_b)\lambda S^2}{2D\left(\lambda - \dfrac{D}{P}\right)}$$

The average cycle time and cycle cost are

$$E(T) = \frac{E(y)}{D} = \frac{E[\lambda Q]}{D} = \frac{Q E(\lambda)]}{D}$$

(6)

and

$$E(c) = K + \left\{ \frac{Q^2 E(\lambda^2)}{2D} - \frac{E(\lambda)Q}{2P} - \frac{E(\lambda)QS}{D} \right\} c_h + \frac{(c_h + c_b)E(\lambda)}{2D\left(E(\lambda) - \frac{D}{P}\right)} \cdot S^2 \tag{7}$$

The expected total annual cost is

$$c(S,Q) = \frac{DK}{QE(\lambda)} + \left\{ \frac{QE(\lambda^2)}{2E(\lambda)} - \frac{D}{2P} - S \right\} c_h + \frac{(c_h + c_b)S^2}{2Q\left[E(\lambda) - \frac{D}{P}\right]} \tag{8}$$

In what follows we assume that the probability density function of λ is uniform with location parameter C and scale parameter $C^c = 1\text{-}C$. That is,

$$f(\lambda) = \begin{cases} \dfrac{1}{1-C} & \text{for } C \le \lambda \le 1, \text{ where } 0 \le C < 1 \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

In this case

$$E(\lambda) = \frac{1+C}{2} \quad , \tag{10}$$

and

$$E(\lambda^2) = \frac{1+C+C^2}{3} \quad . \tag{11}$$

Substituting (9), (10), and (11) into (8) and using calculus, the optimal values for the order quantity, Q^*_{adj}, units backordered, S^*_{adj}, and expected total annual cost, $EAC^*_{adj}(S,Q)$, are found as follows

$$Q^*_{adj,u} = \frac{2}{1+C} \sqrt{\frac{2DK}{c_h \left\{ \dfrac{4}{3} \dfrac{1+C+C^2}{(1+C)^2} - \dfrac{c_h}{c_h+c_b}\left[1 - \dfrac{D}{P}\left(\dfrac{2}{1+C}\right)\right] \right\}}}, \tag{12}$$

$$S^*_{adj,u} = \left(\frac{1+C}{2} - \frac{D}{P}\right)\left(\frac{c_h}{c_h + c_b}\right)Q^*_{adj}, \tag{13}$$

and

$$EAC^*_{adj,u} = \sqrt{2DKc_h\left\{\left(\frac{1}{3}\right)\left(\frac{1-C}{1+C}\right)^2 + \frac{c_b}{c_h+c_b}\left[1 - \frac{D}{P}\left(\frac{2}{1+C}\right)\right]\right\}} \quad . \tag{14}$$

Please note that in (12) through (14), if the yield location parameter C = 1, then the yield scale

parameter $C^c = 0$. In such case, the quality is perfect and the quality-adjusted model of this paper simply reduces to the traditional EMQ model with backorders expressed in (2) through (4).

3. THE OPTIMAL YIELD PARAMETER MODEL

The decision variables in the model of previous section are Q and S for a fixed location parameter of yield distribution, C. The value of this parameter determines the values of both mean and variance of yield distribution. As C approaches one, yield rate increases and yield variability diminishes, hence quality improves. In this paper, we assume that the option of investing to increase C is available. To evaluate the economic trade-offs associated with this investment option, we introduce a companion yield parameter, Ω, as follows:

$$\Omega = \frac{1-C}{1+C} \qquad \text{for} \qquad 0 < \Omega \le 1 \ , \tag{15}$$

Please note that as C increases from 0 to 1, Ω decreases from 1 to zero. Thus, reducing Ω implies increasing C and, hence, improving quality.

Now, we consider Ω to be a decision variable and seek to minimize the average annual cost composed of, investment cost of reducing yield parameter to a new level, ordering, shortage, and holding costs. Specifically, we seek to minimize

$$C(Q,S,\Omega) = i.a_\Omega(\Omega) + EAC^*_{adj,u}(Q,S) \ , \tag{16}$$

Subject to

$$0 < \Omega \le 1 \ , \tag{17}$$

where i is the cost of capital, $a_\Omega(\Omega)$ is a convex and strictly decreasing function of Ω representing the cost of reducing the yield parameter to the level Ω, $EAC_{adj,u}(Q,S)$ is the sum of all inventory related costs given in equation (8) for the case of uniform yield distribution, and Ω_0 is the original quality parameter.

One reasonable way of dealing with this optimization problem is to use a sequential approach, suggested by Porteus [13]. In this case, we hold Ω fixed, optimize over Q and S to obtain $Q^*_{adj}(\Omega)$ and $S^*_{adj}(\Omega)$, given by equations (12) and (13) with $C = \frac{1-\Omega}{1+\Omega}$, and then optimize over Ω. That is, we seek to minimize

$$w(\Omega) = i.a_\Omega(\Omega) + EAC^*_{adj}(\Omega) \ , \tag{18}$$

where $EAC^*_{adj}(\Omega)$ is given by equation (14) modified for Ω as follows:

$$EAC^*_{adj}(\Omega) = \sqrt{2DKc_h\left\{\frac{1}{3}\Omega^2 + \frac{c_b}{c_h+c_b}\left[1 - \frac{D}{P}\,1+\Omega\,\right]\right\}}. \tag{19}$$

Of course if the optimal Ω obtained in this way does not satisfy restriction (17), we should not make any investment and the results of the quality adjusted EOQ model with planned shortages

hold. Please note that it may not always be possible to carry out the above minimization except for some special cases of $a_\Omega(\Omega)$. The following section considers the case of linear investment function.

4. THE LINEAR INVESTMENT FUNCTION

In this case the yield parameter, Ω, declines linearly as the investment amount, a_Ω, is increased. That is

$$a_\Omega(\Omega) = a - b\Omega \qquad for \qquad 0 < \Omega \le \Omega_0 < 1, \tag{20}$$

where $a = b\Omega_0$ and Ω_0 is the original companion yield parameter before investment. Here our main objective is to minimize $w(\Omega)$ after substituting (19) and (20) into (18).

Theorem: If Ω_0 and b are strictly positive and $3\left(\dfrac{D}{P}\right)^2 + 2DKC_h > 3i^2b^2 + 4\left(\dfrac{C_h + C_b}{C_h}\right)\left(1 - \dfrac{D}{P}\right)$, then

the following hold:
a) The optimal value of the yield parameter is given by

$$\Omega^{**} = \min\ \Omega_0, \Omega_{imp,u} \ , \tag{21}$$

where
Ω_0 = the original companion yield parameter and

$$\Omega_{imp,u} = \frac{3}{2}\left(\frac{C_b}{C_h + C_b}\right)\left[\frac{D}{P} = ib\sqrt{\frac{3\left(\dfrac{D}{P}\right)^2 - 4\left(\dfrac{C_h + C_b}{C_h}\right)\left(1 - \dfrac{D}{P}\right)}{3i^2b^2 - 2DKC_h}}\right], \tag{22}$$

(b) The resulting optimal yield location and optimal yield scale parameters are

$$C_u^{**} = \max\ C_0, C_{imp,u} \ , \tag{23}$$

$$C_u^{c**} = \min(C_0^c, C_{imp,u}^c), \tag{24}$$

where C_0 and C_0^c are the original location and scale parameters, and

$$C_{imp,u} = \frac{1 - \Omega_{imp,u}}{1 + \Omega_{imp,u}}, \tag{25}$$

$$C_{imp,u}^c = \frac{2\Omega_{imp,u}}{1 + \Omega_{imp,u}}. \tag{26}$$

(c) The optimal values for the order quantity, $Q_{adj,u}^{**}$, units backordered, $S_{adj,u}^{**}$, and

expected total annual cost, $EAC_{adj,u}^{**}$, are as follows

$$Q_{adj,u}^{**} = \begin{cases} Q_{adj,u}^{*} & if\ \Omega_{imp,u} \geq \Omega_0 \\ Q_{imp,u} & if\ \Omega_{imp,u} < \Omega_0 \end{cases}, \tag{27}$$

$$S_{adj,u}^{**} = \min(S_{adj,u}^{*}, S_{imp,u}) = \begin{cases} S_{adj,u}^{*} & if\ \Omega_{imp,u} \geq \Omega_0 \\ S_{imp,u} & if\ \Omega_{imp,u} < \Omega_0 \end{cases}, \tag{28}$$

$$EAC_{adj,u}^{**} = \min(EAC_{adj,u}^{*}, EAC_{imp,u}) = \begin{cases} EAC_{adj,u}^{*} & if\ \Omega_{imp,u} \geq \Omega_0 \\ EAC_{imp,u} & if\ \Omega_{imp,u} < \Omega_0 \end{cases}, \tag{29}$$

where $Q_{adj,u}^{*}$, $S_{adj,u}^{*}$, and $EAC_{adj,u}^{*}$ are given by (12), (13), and (14),

$$Q_{imp,u} = \left(1+\Omega_{imp,u}\right) \sqrt{\frac{2DK}{c_h\left\{\left(\dfrac{\Omega_{imp,u}}{\sqrt{3}}\right)^2 + \left(\dfrac{c_b}{c_h+c_b}\right)\left[1-\dfrac{D}{P}\left(1+\Omega_{imp,u}\right)\right]\right\}}}, \tag{30}$$

$$S_{imp,u} = \left(1+\Omega_{imp,u}-\frac{D}{P}\right)\left(\frac{c_h}{c_h+c_b}\right)Q_{imp,u}, \tag{31}$$

and

$$EAC_{imp,u} = \sqrt{2DKC_h\left\{\left(\frac{\Omega_{imp,u}}{\sqrt{3}}\right)^2 + \left(\frac{c_b}{c_h+c_b}\right)\left[1-\frac{D}{P}\left(1+\Omega_{imp,u}\right)\right]\right\}}. \tag{32}$$

It is interesting to note that $\Omega_{imp,u}$, $C_{imp,u}$, $C_{imp,u}^{c}$, $Q_{imp,u}$, $S_{imp,u}$, and $EAC_{imp,u}$ do not depend on Ω_0. Furthermore, when $\Omega_{imp,u} \geq \Omega_0$, then no investment is made and the results of the quality adjusted model with uniformly distributed λ hold. In such case Ω_0 will be used in place of $\Omega_{imp,u}$ and equations (30), (31), and (32) will be replaced by equations (12) through (14). In addition, when quality is perfect (i.e., $\Omega=0$, $C=1$, and $C^c=0$), then the results of this paper simply reduce to the corresponding results of the traditional EMQ with backorders given in equations (2) through (4).

Details of the proofs are omitted.

5. CONCLUSION

This paper extended the results for the optimal parameters of yield distribution in a recently developed yield-adjusted EOQ with planned shortages model with random production yield [7] to the case of Economic Manufacturing Quantity (EMQ) . The paper studied the economic trade-offs associated with investment efforts aimed at improving the production yield through changes in both the location parameter as well as the scale parameter of yield distribution. Assuming that the yield probability density function is uniform, the paper presented explicit relationships for the optimal values of location and scale parameters, order quantity, backorder level, and expected total annual cost of the yield-adjusted EMQ model for the case of linear investment function.

REFERENCES

[1]   Affisco, J.F., Paknejad, M.J., and Nasri, F. "Quality improvement in the economic order quantity model," *Proceedings of the Northeast Decision Sciences Institute*, pp. 256-258, Atlantic City, New Jersey, 2000.

[2]   Affisco, J.F., Paknejad, M.J., and Nasri, F. "Quality improvement and setup reduction in the joint economic lot size model," *European Journal of Operational Research*, 2002, 142, 497-508.

[3]   Cheng, T.C.E., 1991, "EPQ with process capability and quality assurance considerations," *Journal of Operational Research Society*, 42(8), 713-720.

[4]   Goyal, S.K., Gunasekaran, A., Martikainen, T. & Yli-Olli, P., 1993, "Integrating production and quality control policies: A Survey," *European Journal of Operational Research*, 69, 1-13.

[5]   Hall, R.W. *Zero Inventories.* Dow Jones-Irwin, Homewood, IL, 1983.

[6]   Moinzadeh, K., and H.L., Lee, "A continuous review inventory model with constant resupply time and defective items," *Naval Research Logistics*, 1987, 34, 457-467.

[7]   Nasri, F., Paknejad, J., & Affisco, J. F., "The Optimal parameters of yield distribution in an EOQ model with planned shortages," *2010 Western Decision Sciences Institute Proceedings*, pp. 4571-4577, April 2010.

[8]   Nasri, F., Paknejad, J., & Affisco, J. F., "An EMQ model with planned shortages and random defective units", *2009 Decision Sciences Institute Proceedings*, pp 10541-10546, New Orleans, November 2009.

[9]   Nasri, F., Paknejad, J., & Affisco, J. F., "lot size determination in an EOQ model with planned shortage and random defective Units," *Northeast Decision Sciences Institute Proceedings*, pp. 403-408, on CD-ROM, April 2009.

[10]   Paknejad, J., Nasri, F. & Affisco, J.F., "The optimal parameters of yield distribution in an EOQ model with two types of shortage," *Proceedings of the 2010 Decision Sciences Institute*, pp. 2221-2228, San Diego, November 2010.

[11]   Paknejad, J., Nasri, F. & Affisco, J.F., "Flexibility and Quality Improvement in the EOQ with Planned Shortage Model," *Proceedings of the Northeast Decision Sciences Institute*, pp. 257-259, Pittsburgh, Pennsylvania, 2001.

[12]   Paknejad, J., Nasri, F. & Affisco, J.F., "Defective units in a continuous review (s,Q) system," *International Journal of Production Research*, 1995, 33(10), 2767-2777.

[13]   Porteus, E., "Optimal lot sizing, process quality improvement and setup cost reduction," *Operations Research*, 1986, 34(1), 137-144.

[14]   Rosenblatt, M.J. and Lee, H.L., "Economic production cycles with imperfect production processes," *IIE Transactions*, 18(3), 1986, 48-55.

[15]   Yano, C.A. and Lee, H.L., "Lot Sizing with random yields: A review," *Operations Research*, 43(3), 1995, 311-334.

AN EXPLORATORY STUDY:
IS THERE ANYTHING GOOD ABOUT RATEMYPROFESSOR?

Richard L. Peterson
Department of Management & Information Systems
Montclair State University, Montclair, NJ 07043, 973-655-7038
petersonr@mail.montclair.edu

Mark L. Berenson
Department of Management & Information Systems
Montclair State University, Montclair, NJ 07043, 973-655-6857
berensonm@mail.montclair.edu

Risha Aijaz
Department of Management & Information Systems
Montclair State University, Montclair, NJ 07043, 973-655-4335
aijazr1@mail.montclair.edu

ABSTRACT

Professor rating services such as RateMyProfessor.com are frequently dismissed as invalid due to their inherent nature of self selection. Only raters at the extremes, so the theory goes, contribute to the site as they have axes to grind or high praises to deliver. The middle, supposedly more even handed raters don't bother to offer ratings or comments. While rater bias on these sites seems logical in theory, is it actually true? RateMyProfessor provides an opportunity to test for bias in the comments offered by the raters. If these verbal evaluations are biased the language used would be extreme and outside the boundaries of normal written discourse. The work of Hart [8] and others provides both a software tool for textual analysis and normative data on nine "dimensions of language". The exploratory research reported here examines written comments for individual professors. We evaluate the comments of positive and negative raters to the norms of discourse across a variety of genres and identify language dimensions where raters truly are extreme and where they are not.

KEYWORDS: Professor Ratings, Text Analysis, Content Analysis

INTRODUCTION

Ten years after being created, RateMyProfessor.com (RMP) today is the most popular and widely used website by students to provide ratings and comments about their professors; with more than 10 million opinions of over 1 million professors. The ratings cover more than 6,000 schools across the United States, Canada, England, Scotland, and Wales*. It's appreciated by students "shopping" for professors, but dismissed by empiricists for its self-selection of raters violates the bedrock principle of random selection

RMP offers two sources of data: numeric ratings and textual comments. Raters are presented with a 1-5 scale:

Easiness

◉ Hard   ○   ○   ○   ○ Easy

Helpfulness:

◉ Useless   ○   ○   ○   ○ Extremely Helpful

Clarity:

○ Incomprehensible   ○   ○   ○   ○ Crystal

Interest level prior to attending class:

○ None at all   ○   ○   ○   ○ It's my world!

Appearance: (just for fun)

○ Hot   ○ Not

An Overall Quality rating is calculated from Helpfulness and Clarity. Other information solicited—but not displayed--includes: Textbook Use (5-point scale from "Low" to "High," Textbook Used (by name or ISBN), Grade, Attendance (Mandatory or Not Mandatory), Professor Status (Still Teaching, Retired/Gone), and Class and Section (both fill-in). Finally, RMP includes a 350 character text box for students to type comments. The text box includes a

prominent note "Please keep comments clean. Libelous comments will be deleted." Also included are warnings "Remember, **YOU ARE RESPONSIBLE** for what you write here. Submitted data become the property of RateMyProfessors.com. IP addresses are logged." Adjacent to the Comments field is a label "Guidelines" that links to page of "Dos" and "Don'ts" for raters which is reproduced as Appendix A.

Despite the best intentions of those responsible for the development of RateMyProfessors.com, one must seriously question the value of the numerical ratings provided because the raters "self-select," violating the tenet of random selection needed for drawing overall inferences from any survey or designed experiment. It has long been surmised that the majority of those students who decide to participate in the evaluation process hold distinctly bipolar views of the faculty member they are rating. Students with a legitimate or perceived gripe are more likely to participate in the ratings, as are students who very much appreciate or value what the faculty member has contributed to the course. A third group of raters, asserted to be far fewer in number, believe the faculty member being rated is "okay/average" but feel obligated to participate in the rating process because of their responsibility to fraternity, sorority or fellow classmates. Thus it is conjectured that the distribution of RMP evaluations will be U-shaped for most faculty members, the majority of ratings being "Good" or "Poor," with a minority of ratings being "average." On the other hand, had universal/mandatory ratings been required, as may be the case at various institutions of higher learning, whether a 3-point rating scale is used, or more popular 5 or 7-point Likert-type rating scales are used, one would hypothesize the distribution of ratings for most faculty would be unimodal with one of the tails (highest rating or lowest rating) being very limited in frequency count.

Interestingly, and despite the problems of self-selection bias in the RMP ratings, Jaschik [12] has shown that there is a significantly positive correlation between the numerical RMP quality rating and the average rating computed from the student evaluations of the faculty by a whole class. As outlined in the Methodology, using the RMP quality rating data for the faculty in the Management and Information Systems Department in the School of Business at Montclair State University along with the corresponding student evaluation ratings on campus, one aspect of this overall study will be to attempt to corroborate the findings reported in the study by Jaschik [12].

If we can't trust the ratings per se due to bias, is it also the case that we can't trust the comments? Are the comments also biased such that the language of the comments is somehow different from "normal" language? Are raters somehow different from the "general" population in terms of language usage? To make this determination we might compare the language of the comments of the raters on RMP to a variety of language samples.

Content analysis of text is a research methodology that uses a set of procedures to analyze and categorize communication [20]. The methodology offers a number of potential benefits including the identification of individual differences among communicators [20], avoidance of recall biases [1], and the ability to obtain otherwise unavailable information [13]. In the business disciplines content analysis has been used in accounting [17], management [2], marketing [19], and corporate strategy [15] [21].

There are three types of approaches to language analysis 16]: human-scored procedures, artificial intelligence systems, and individual word count systems. With the first approach, coding rules are established, human coders are trained, and then the coders classify selected aspects of the text. Artificial intelligence approaches consider the lexicon, syntax, and semantics of text [18]. With respect to individual word count methodology, individual words in the text are counted and the frequency of each word is compared to the frequency of these same words in other communication samples. Word frequencies outside of the range of frequencies in these comparative samples are an indication of differences between the samples.

DICTION [8] is one of a number of word frequency programs. In DICTION the frequency of word usage in the analyzed text is compared to the frequency of word usage across various genres studied by Hart [7] [4] [5] [9] [10]. The genre(s) to which the analyzed text is compared may be selected from business, daily life, entertainment, journalism, literature, politics, and scholarship. Hart [11] analyzed from one to six sub-genres to derive word frequency norms for each genre. In total, the norms are based on the analysis of 22,027 texts of various genres written between 1948 and 1998.

The words from these genres are arranged in 33 dictionaries or word lists ranging in size from 10 to 745 words. No word appears in more than one dictionary. Brief descriptions of each dictionary may be found in Appendix B.

In addition to the absolute frequency counts, DICTION calculates four variables based on word ratios. These calculated variables are:

Insistence, a measure of "code-restriction" that indicates a "preference for a limited, ordered world";

Embellishment, a measure of the ratio of adjectives to verbs;

Variety, a measure of conformity to, or avoidance of, a limited set of expressions (different words/total words); and

Complexity, a measure of word size based on the Flesch[3] method.

Frequency counts from the various dictionaries along with the four calculated compose five master variables. Hart's [11] master variables, intended to capture the tonal features of the text, are defined and formulated as follows:

Certainty is a measure of language "indicating resoluteness, inflexibility, and completeness and a tendency to speak ex cathedra."

Certainty = [Tenacity. + Leveling. + Collectives. + Insistence] - [Numerical Terms + Ambivalence. + Self Reference + Variety]

Activity is a measure of "movement, change, [and] the implementation of ideas and the avoidance of inertia;

Activity = [Aggression. + Accomplishment. + Communication. + Motion] - [Cognitive Terms. + Passivity. + Embellishment]

Optimism is a measure of "language endorsing some person, group, concept or event or highlighting their positive entailments."

Optimism = [Praise + Satisfaction + Inspiration] - [Blame + Hardship +Denial]

Realism is a measure of language "describing tangible, immediate, recognizable matters that affect people's everyday lives."

Realism = [Familiarity + Spatial Awareness. + Temporal Awareness. +Present Concerns. + Human Interest + Concreteness] - [Past Concern + Complexity]

Commonality is a measure of language "highlighting the agreed-upon values of a group and rejecting idiosyncratic modes of engagement."

Commonality = [Centrality. + Cooperation. + Rapport] - [Diversity + Exclusion + Liberation]

SPECIFIC RESEARCH QUESTIONS

In RMP an Overall Quality rating is calculated from the raters' numerical evaluations of ratings of professors' Helpfulness and Clarity. For each professor rated, RMP categorizes the professor's Overall Quality as "Good," "Average," or "Poor." In this exploratory study of the textual comments of raters we questioned whether there would be significant differences among the comments created by raters classified in each category of Overall Quality compared to the general population as defined by the dictionaries included in DICTION. Specifically, the questions addressed were:

- Are there significant differences in the commentary provided by RMP evaluators who rate professors "Good" versus those who rate professors "Poor" with respect to the master variables?

- Are there significant differences in the commentary provided by RMP evaluators who rate professors "Good" versus those who rate professors "Poor" with respect to the five content-analysis master variables?

- Are there significant differences in the commentary provided by RMP evaluators who rate professors "Good" versus those who rate professors "Poor" with respect to the four content-analysis calculated variables?

For the Master variables of DICTION, we hypothesized as follows:

Certainty. As a measure of resoluteness, the group of raters rating professors as "Poor" will be more certain in their language than those rating the professor as "Good." This certainty will be revealed in higher scores for tenacity, leveling, collectives, insistence, numerical terms, ambivalence and self-references. "Poor" raters will also show this certainty by a reduced variety of words.

Activity. This measure of movement will show "Poor" raters with greater activity. Words of aggression, accomplishment, and communication will be higher for this group. They will also use more terms of passivity, and more embellishments. "Good" and "Poor" raters will use more cognitive terms than the "Average"

Optimism. "Good" and "Poor" raters will reflect their ratings in their choice of words in their comments. "Good" raters will use more terms of praise, satisfaction, and inspiration. "Poor" raters will do the opposite; fewer terms of praise, satisfaction, and inspiration. In addition, while :Good" raters will not avoid blame, hardship, and denial, "Poor" raters will use more of these terms than would be usual.

Realism. There will be no significant differences in the frequency of word use in any dictionary making up this master variable either for "Good" or "Poor" raters.

Commonality. ""Good" raters will use above average numbers of words of centrality, cooperation and rapport while "Poor" raters will use fewer of these words. These raters will indicate their feelings of exclusion and liberation by more frequent use of these terms.

METHODOLOGY

For this exploratory study we restricted the data set to all full-time, tenure/tenure-track faculty teaching in a department within a school of business at a public university who had entries on RateMyProfessor. Of the 23 faculty members in the department over the period, 100 percent had ratings on RMP. As we were interested in studying raters who were or planned to be business students, given that some professors also had expertise in other disciplines and taught some courses outside the school of business, we eliminated all ratings and comments for any professor where the reported experience with the professor was in a course not offered by the department. This resulted in the elimination of 18 RMP records of the 700 total. Ratings without comments (a total of 17) were also eliminated as these comments were the focus of current study. Finally as our interest was on ratings and comments at the extremes we eliminated 106 raters whose Overall Quality rating was "Average."

The comments of the remaining 559 raters were then cleaned to correct misspellings and abbreviations that would impact the word frequency counts. No other changes were made to the corpus. Text files of all the comments from "Good" and "Poor" raters and a combined file were created and submitted to DICTION. All words were processed (DICTION allows sampling of the corpus) and no custom dictionaries were created for the analysis.

AN EXPLORATORY ANALYSES OF SPECIFIC RESEARCH HYPOTHESIS

Table 1 displays the DICTION reported results for the two groups of ratings, "Good" versus "Poor," for each of the five master variables and their corresponding component variables. Also indicated is whether or not our hypothesized directions of results were confirmed. The results indicate that for the five master variables of DICTION our combined hypotheses were not consistently confirmed, nor were they for the component variables.

Table 1 – Dictionary, Constructed, and Master Variable Results for "Good" and "Poor" Raters

Variable	Frequency	Frequency2	Hypothesis: Poor vs. Good
	Good	Poor	
Numerical Terms	6.09	7.25	>
Ambivalence	18.09	17.30	>
Self-reference	7.80	8.58	>
Tenacity	36.01	46.70*	>
Leveling Terms	11.52	15.21*	>
Collectives	12.57	8.85	>
Praise	17.01*	5.99	<
Satisfaction	2.82	3.67	<
Inspiration	1.22*	1.31*	<
Blame	6.20*	15.53*	>
Hardship	1.16*	3.32	>
Aggression	1.28	0.05*	>
Accomplishment	3.65*	3.87*	>
Communication	10.20	11.97*	>
Cognition	10.27	19.70*	?
Passivity	2.23	2.82	>
Spatial Terms	2.09*	4.37	?
Familiarity	80.85*	89.63*	?
Temporal Terms	4.56*	8.56	?
Present Concern	13.09	19.66*	?
Human Interest	41.20	58.05*	?
Concreteness	10.24*	16.01	?
Past Concern	3.03	2.44	?
Centrality	0.26*	1.11*	<
Rapport	0.17*	1.56	<
Cooperation	15.09*	11.09*	<
Diversity	0.39	0.52	?

Exclusion	0.29	0.31	>
Liberation	0.53	0.30	>
Denial	9.43*	20.95*	>
Motion	2.44	1.61	?

Calculated Variables	Good	Poor	Hypothesis: Poor vs. Good
Insistence	73.04	2.51	>
Embellishment	1.55*	1.00	>
Variety	0.47	0.46	<
Complexity	3.93*	4.25*	?

Master Variables	Good	Poor	Hypothesis: Poor vs.Good
Activity	47.60	6.32	>
Optimism	50.20	9.31	<
Certainty	51.97	2.87	>
Realism	46.00	9.92	?
Commonality	52.53	2.53	<

The first surprising result occurred with the master variable certainty where the "Good" raters scored higher than the "Poor" raters. However, for the eight component variables we correctly hypothesized the result five times and were wrong three times.

The second surprising result occurred with the master variable activity. The "Good" raters again scored higher than the "Poor" raters. However, for the seven component variables we correctly hypothesized the result three times and were wrong twice. We did not hypothesize and difference in results for two component variables.

For the master variable optimism our hypothesis was confirmed. The "Good" raters again scored higher than the "Poor" raters. On the other hand, for the six component variables we correctly hypothesized the result four times and were wrong twice.

For the master variable realism we did not hypothesize any difference in direction of results for "Good" versus "Poor" raters, nor did we make any hypotheses for the eight component variables.

For the master variable commonality our hypothesis was confirmed. The "Good" raters again scored higher than the "Poor" raters. On the other hand, for the six component variables we correctly hypothesized the result only two times and were wrong three times. We did not hypothesize the direction of the results for one of the component variables.

DISCUSSION

From Table 1 it is observed that the "Good" raters outscored the "Poor" raters on four of five master variables and our hypothesized results were only confirmed twice. For one master variable, realism, we did not specify a preconceived difference in direction and that was the only master variable that demonstrated higher scores for the "Poor" raters. Breaking these results down by the components of the five master variables, we correctly hypothesized results 14 times, we were incorrect 10 times and on 11 occasions we did not attempt to predict the direction of the results.

The question that must be pondered is why such unexpected results? A few possibilities must be thoroughly examined.

1 – As asked rhetorically in the Introduction section, if we can't trust the numerical ratings per se due to bias, is it also the case that we can't trust the comments? Are the comments also biased such that the language of the comments is somehow different from "normal" language? Are raters somehow different from the "general" population in terms of language usage? To make this determination we may need to compare the language of the comments of the raters on RMP to a variety of language samples, not just the DICTION program..

2 – Although much has been written about the DICTION program in various articles by Hart [4] [5] [6] [7] [8] [9] [10] and others who have used it for research one must question its validity reliability which have not been reported. Furthermore, there is no readily found description of the computation of mean and standard deviation, or computation of the standard Z scores and Hart fails to demonstrate why he breaks with long-held convention and describes absolute Z scores greater than 1.0 as outside the normal range and thereby significant.

3 – Hart's [10] five master variables appear to be independent constructs arising from a factor analysis – there is little to no correlation among these constructs. The component variables comprising the five master variables did not seem to consistently display the direction of difference expected by our hypotheses, perhaps a misunderstanding on our part of the definition of the involved terms?

4 – It is impossible for us to determine how the DICTION program searches for discrepancies in various commentary that could result in misclassification. For example, it is not known whether the DICTION program can properly classify a comment about teacher performance that says "the teacher is easy" versus " the teacher is not easy" versus "I was told the teacher was easy but I don't think this is so." If the DICTION program cannot properly distinguish among such responses both its reliability and validity as a measuring instrument can be questioned.

Further exploration will provide answers to the above.

CONCLUSIONS

Despite the surprising findings which indicated discrepancies with several of our hypotheses we remain encouraged by the results of this exploratory study. Once we satisfactorily address the aforementioned dilemmas described in the Discussion section we plan to extend the study and address additional questions in two phases:

:Phase I Questions

- Is there a significantly positive correlation between the biased RMP quality ratings of the faculty in the Management and Information Systems Department in the School of Business at Montclair State University and the corresponding campus student evaluations received by these faculty? That is, do the MGIS Department data corroborate the findings reported in the study conducted by Jaschik [12]?

- Omitting "Average" ratings, are the percentage of "Good" to "Good" or "Poor" ratings significantly higher from the class evaluations than from the RMP evaluations? An affirmation of this hypothesis is proof of the negative self-selection bias in the RMP ratings.

Phase II Questions

- Given the self-selection bias issues, using the Management and Information Systems Department faculty ratings in the School of Business at Montclair State University as a base, is there a statistically significant difference between the RMP quality ratings of these faculty and those given to a randomly selected sample of similar faculty from corresponding/similar AACSB-International public universities?

- Are there significant differences in the commentary provided by RMP evaluators who rate professors "Good" versus those who rate professors "Poor" with respect to the 33 content-analysis dictionary variables between the aforementioned Montclair State University faculty and the randomly selected faculty?

- Are there significant differences in the commentary provided by RMP evaluators who rate professors "Good" versus those who rate

professors "Poor" with respect to the five content-analysis master variables between the aforementioned Montclair State University faculty and the randomly selected faculty?

- Are there significant differences in the commentary provided by RMP evaluators who rate professors "Good" versus those who rate professors "Poor" with respect to the four content-analysis calculated variables between the aforementioned Montclair State University faculty and the randomly selected faculty?

REFERENCES

[1]   Barr, P. S., Stimpert, J. L., & Huff, A. S. (1992). Cognitive Change, Strategic Action, and Organizational Renewal. *Strategic Management Journa*l, 13, 15-36.

[2]   Fiol, C. M. (1989). A semiotic analysis of corporate language: Organizational boundaries and joint venturing. *Administrative Science Quarterly*, 34, 277-303.

[3]   Flesch, R. (1951). *The Art of Clear Thinking*. New York: Harper.

[4]   Hart, R. P. (1984a). Systematic analysis of political discourse: The development of DICTION. K. Sanders, L. Kaid, & D. Nimmo (Eds.), *Political Communication Yearbook* (pp. 97-134). Carbondale: Southern Illinois University Press.

[5]   Hart, R. P. (1984b). *Verbal Style and the Presidency: A Computer-based Analysis*. New York: Academic.

[6]   Hart, R. P. (1987). *The Sound of Leadership: Presidential Communication in the Modern Age*. Chicago, IL: University of Chicago Press.

[7]   Hart, R. P., and S. Jarvis. (1997). Political Debate: Forms, Styles, and Media. *American Behavioral Scientist* 40: 1095-1122.

[8]   Hart, R. P. (2000a). *DICTION 5.0: The Text-analysis Program*. Thousand Oaks, CA: Scolari/Sage Publications.

[9]   Hart, R. P. (2000b). *Campaign Talk: Why Elections are Good for Us*. Princeton, NJ: Princeton University Press.

[10]   Hart, R. P. (2001). Redeveloping DICTION: Theoretical Considerations. In M. West (Ed.), *Theory, Method, and Practice of Computer Content Analysis* (pp. 26-55). New York: Ablex.

[11]   Hart, R. P. (2003). DICTION http://rhetorica.net.htm

[12]   Jaschik, S. (2007). Could RateMyProfessors.com be Right? *Inside Higher Education.* June:1–2.

[13]   Kabanoff, B., Waldersee, R., & Cohen, M. (1995). Espoused values and organizational change themes. *Academy of Management Journal*, 38, 1075-1104.

[14]   McCain, J. (2010) 4 Great Sites to Rate and Review Teachers & Professors   http://www.makeuseof.com/tag/4-great-sites-rate-review-teachers-professors/

[15]   Merchant, H. (2004). Revisiting Shareholder Value Creation via International Joint Ventures: Examining Interactions among Firm- and Context-specific Variables. *Canadian Journal of Administrative Sciences*, 21, 129-145.

[16]   Morris, R. (1994). Computerized Content Analysis in Management Research: A Demonstration of Advantages & Limitations. *Journal of Management*, 20, 903-931.

[17]   Rogers, R. K., Dillard, J., & Yuthas, K. (2005). The Accounting Profession: Substantive Change and/or Image Management. *Journal of Business Ethics*, 58, 159-176.

[18]   Rosenberg, S. D., Schnurr, P. P., & Oxman, T. E. (1990). Content Analysis: A Comparison of Manual and Computerized Systems. *Journal of Personality Assessment*, 54, 298-310.

[19]   Simon, M., & Houghton, S. M. (2003). The Relationship Between Overconfidence and the Introduction of Risky Products: Evidence from a Field Study. *Academy of Management Journal*, 46, 139-149.

[20]   Weber, R. P. (1990). *Basic Content Analysis*. Newbury Park, CA: Sage.

[21]   Yuthas, K., Rogers, R., & Dillard, J. F. (2002). Communicative Action and Corporate Annual Reports. *Journal of Business Ethics*, 41, 141-157.

APPENDIX A

RateMyProfessor.com Posting Guidelines

As a user of RateMyprofessors.com, you agree and accept the terms and conditions of the site. This site is a resource for students to provide and receive feedback on professor's teaching methods and insight into the courses. Comments should only be posted by students who have taken a class from the professor. Please limit one comment per person per course.

The following guidelines are intended to protect all users-students and professors. Please review before posting on RateMyProfessors.com

- DOs:
    - Be honest.
    - Be objective in your assessment of the professor.
    - Limit your comments to the professor's professional abilities. Do not to get personal.
    - Proof your comments before submitting. Poor spelling WILL NOT cause your rating to be removed; however, poor spelling may result in your rating being discredited by those who read it.
    - Leave off your Name, Initials, Pseudo Name, or any sort of identifying mark when posting.
    - Refer to the Rating Categories to help you better elaborate your comments.
    - Remember that negative comments that still offer constructive criticism are useful. Comments that bash a professor on a personal level are not.
    - Submit helpful comments that mention professor's ability to teach and/or communicate effectively, course load, type of course work and course topics.
- DO NOTs:
    - State something as a fact if it is your opinion.
    - Post a rating if you are not a student or have not taken a class from the professor.
    - Post ratings for people who do not teach classes at your college or university.
    - Input false course or section codes for a class that does not exist.
    - Rate a professor more than once for the same class.
    - Make references to other comments posted.
    - Professors : Do not rate yourselves or your colleagues.

Comments will be deemed inappropriate that are libelous, defamatory, indecent, vulgar or obscene, pornographic, sexually explicit or sexually suggestive, racially, culturally, or ethnically offensive, harmful, harassing, intimidating, threatening, hateful, objectionable, discriminatory, or abusive, or which may or may appear to impersonate anyone else.

COMMENTS THAT CONTAIN THE FOLLOWING WILL BE REMOVED :

- Profanity, name-calling, vulgarity or sexually explicit in nature
- Derogatory remarks about the professor's religion, ethnicity or race, physical appearance, mental and physical disabilities.
- References to professor's sex life (Including sexual innuendo, sexual orientation or claims that the professor sleeps with students).
- Claims that the professor shows bias for or against a student or specific groups of students.
- Claims that the professor has been or will be fired, suspended from their job, on probation.
- Claims that the professor engages or has previously engaged in illegal activities (drug use, been incarcerated.)
- Includes a link/URL to a webpage or website that does not directly pertain to the class.
- Any piece of information including contact info that enables someone to identify a student.
- Any piece of information about the professor that is not available on the school's website and allows someone to contact them outside of school. This also includes remarks about the professor's family and personal life.
- Accusations that the professors is rating themselves or their colleagues.
- Is written in a language other than English? Unless you attend a French-Canadian school.

The Do Nots of these Posting Guidelines will be enforced and violations will result in either the rating's comment being removed, or the entire rating being deleted. If you see a rating that you believe violates Posting Guidelines, please click the red flag and state the problem. It will be evaluated by RateMyProfessors moderators.

Please do not flag a rating just because you disagree with it.

Comments containing a threat of violence against a person or any other remark that would tend to be seen as intimidating or intends to harm someone will deleted. RateMyProfessors will notify the authorities of your

IP address and the time you rated. This is enough information to identify you. IP addresses will also be turned over to the proper authorities when presented with a subpoenas or court orders from a government agency or court.

Multiple Ratings

Multiple ratings / comments from the same IP in a short amount of time are automatically deleted on our backend to fight rating abuse. There is no differentiation between positive and negative comments. Please give only one comment per person per course.

New Professors

Requests to add a new professor can be submitted on the school page and will be added once approved by a moderator. Please only submit professors who currently teach a course at your college or university.

http://www.ratemyprofessors.com/rater_guidelines.jsp

APPENDIX B

Descriptions of the Dictionaries and Scores

Accomplishment: Words expressing task-completion (establish, finish, influence, proceed) and organized human behavior (motivated, influence, leader, manage). Includes capitalistic terms (buy, produce, employees, sell), modes of expansion (grow, increase, generate, construction) and general functionality (handling, strengthen, succeed, outputs). Also included is programmatic language: agenda, enacted, working, leadership.

Aggression: A dictionary embracing human competition and forceful action. Its terms connote physical energy (blast, crash, explode, collide), social domination (conquest, attacking, dictatorships, violation), and goal-directedness (crusade, commanded, challenging, overcome). In addition, words associated with personal triumph (mastered, rambunctious, pushy), excess human energy (prod, poke, pound, shove), disassembly (dismantle, demolish, overturn, veto) and resistance (prevent, reduce, defend, curbed) are included.

Ambivalence: Words expressing hesitation or uncertainty, implying a speaker's inability or unwillingness to commit to the verbalization being made. Included are hedges (allegedly, perhaps, might), statements of inexactness (almost, approximate, vague, somewhere) and confusion (baffled, puzzling, hesitate). Also included are words of restrained possibility (could, would, he'd) and mystery (dilemma, guess, suppose, seems).

Blame: Terms designating social inappropriateness (mean, naive, sloppy, stupid) as well as downright evil (fascist, blood-thirsty, repugnant, malicious) compose this dictionary. In addition, adjectives describing unfortunate circumstances (bankrupt, rash, morbid, embarrassing) or unplanned vicissitudes (weary, nervous, painful, detrimental) are included. The dictionary also contains outright denigrations: cruel, illegitimate, offensive, miserly.

Centrality: Terms denoting institutional regularities and/or substantive agreement on core values. Included are indigenous terms (native, basic, innate) and designations of legitimacy (orthodox, decorum, constitutional, ratified), systematicity (paradigm, bureaucratic, ritualistic), and typicality (standardized, matter-of-fact, regularity). Also included are terms of congruence (conformity, mandate, unanimous), predictability (expected, continuity, reliable), and universality (womankind, perennial, landmarks).

Cognitive Terms: Words referring to cerebral processes, both functional and imaginative. Included are modes of discovery (learn, deliberate, consider, compare) and domains of study (biology, psychology, logic, economics). The dictionary includes mental challenges (question, forget, re-examine, paradoxes), institutional learning practices (graduation, teaching, classrooms), as well as three forms of intellection: intuitional (invent, perceive, speculate, interpret), rationalistic (estimate, examine, reasonable, strategies), and calculative (diagnose, analyze, software, fact-finding).

Collectives: Singular nouns connoting plurality that function to decrease specificity. These words reflect a dependence on categorical modes of thought. Included are social groupings (crowd, choir, team, humanity), task groups (army, congress, legislature, staff) and geographical entities (county, world, kingdom, republic).

Communication: Terms referring to social interaction, both face-to-face (listen, interview, read, speak) and mediated (film, videotape, telephone, e-mail). The dictionary includes both modes of intercourse (translate, quote, scripts, broadcast) and moods of intercourse (chat, declare, flatter, demand). Other terms refer to social actors (reporter, spokesperson, advocates, preacher) and a variety of social purposes (hint, rebuke, respond, persuade).

Complexity: A simple measure of the average number of characters-per-word in a given input file. Borrows Rudolph Flesch's (1951) notion that convoluted phrasings make a text's ideas abstract and its implications unclear.

Concreteness: A large dictionary possessing no thematic unity other than tangibility and materiality. Included are sociological units (peasants, African-Americans, Catholics), occupational groups (carpenter, manufacturer, policewoman), and political alignments (Communists, congressman, Europeans). Also incorporated are physical structures (courthouse, temple, store), forms of diversion (television, football, CD-ROM), terms of accountancy (mortgage, wages, finances), and modes of transportation (airplane, ship, bicycle). In addition, the dictionary includes body parts (stomach, eyes, lips), articles of clothing (slacks, pants, shirt), household animals (cat, insects, horse) and foodstuffs (wine, grain, sugar), and general elements of nature (oil, silk, sand).

Cooperation: Terms designating behavioral interactions among people that often result in a group product. Included are designations of formal work relations (unions, schoolmates, caucus) and informal associations (chum, partner, cronies) to more intimate interactions (sisterhood, friendship, comrade). Also included are neutral interactions (consolidate, mediate, alignment), job-related tasks (network, détente, exchange), personal

involvement (teamwork, sharing, contribute), and self-denial (public-spirited, care-taking, self-sacrifice).

Denial: A dictionary consisting of standard negative contractions (aren't, shouldn't, don't), negative functions words (nor, not, nay), and terms designating null sets (nothing, nobody, none).

Diversity: Words describing individuals or groups of individuals differing from the norm. Such distinctiveness may be comparatively neutral (inconsistent, contrasting, non-conformist) but it can also be positive (exceptional, unique, individualistic) and negative (illegitimate, rabble-rouser, extremist). Functionally, heterogeneity may be an asset (far-flung, dispersed, diffuse) or a liability (factionalism, deviancy, quirky) as can its characterizations: rare vs. queer, variety vs. jumble, distinctive vs. disobedient.

Exclusion: A dictionary describing the sources and effects of social isolation. Such seclusion can be phrased passively (displaced, sequestered) as well as positively (self-contained, self-sufficient) and negatively (outlaws, repudiated). Moreover, it can result from voluntary forces (secede, privacy) and involuntary forces (ostracize, forsake, discriminate) and from both personality factors (small-mindedness, loneliness) and political factors (right-wingers, nihilism). Exclusion is often a dialectical concept: hermit vs. derelict, refugee vs. pariah, discard vs. spurn).

Familiarity: Consists of a selected number of C.K. Ogden's (1968) "operation" words which he calculates to be the most common words in the English language. Included are common prepositions (across, over, through), demonstrative pronouns (this, that) and interrogative pronouns (who, what), and a variety of particles, conjunctions and connectives (a, for, so).

Hardship: This dictionary contains natural disasters (earthquake, starvation, tornado, pollution), hostile actions (killers, bankruptcy, enemies, vices) and censurable human behavior (infidelity, despots, betrayal). It also includes unsavory political outcomes (injustice, slavery, exploitation, rebellion) as well as normal human fears (grief, unemployment, died, apprehension) and incapacities (error, cop-outs, weakness).

Human Interest: An adaptation of Rudolf Flesch's notion that concentrating on people and their activities gives discourse a life-like quality. Included are standard personal pronouns (he, his, ourselves, them), family members and relations (cousin, wife, grandchild, uncle), and generic terms (friend, baby, human, persons).

Inspiration: Abstract virtues deserving of universal respect. Most of the terms in this dictionary are nouns isolating desirable moral qualities (faith, honesty, self-sacrifice, virtue) as well as attractive personal qualities (courage, dedication, wisdom, mercy). Social and political ideals are also included: patriotism, success, education, justice.

Leveling: Words used to ignore individual differences and to build a sense of completeness and assurance. Included are totalizing terms (everybody, anyone, each, fully), adverbs of permanence (always, completely, inevitably, consistently), and resolute adjectives (unconditional, consummate, absolute, open-and-shut).

Liberation: Terms describing the maximizing of individual choice (autonomous, open-minded, options) and the rejection of social conventions (unencumbered, radical, released). Liberation is motivated by both personality factors (eccentric, impetuous, flighty) and political forces (suffrage, liberty, freedom, emancipation) and may produce dramatic outcomes (exodus, riotous, deliverance) or subdued effects (loosen, disentangle, outpouring). Liberatory terms also admit to rival characterizations: exemption vs. loophole, elope vs. abscond, uninhibited vs. outlandish.

Motion: Terms connoting human movement (bustle, job, lurch, leap), physical processes (circulate, momentum, revolve, twist), journeys (barnstorm, jaunt, wandering, travels), speed (lickety-split, nimble, zip, whistle-stop), and modes of transit (ride, fly, glide, swim).

Numerical Terms: Any sum, date, or product specifying the facts in a given case. This dictionary treats each isolated integer as a single "word" and each separate group of integers as a single word. In addition, the dictionary contains common numbers in lexical format (one, tenfold, hundred, zero) as well as terms indicating numerical operations (subtract, divide, multiply, percentage) and quantitative topics (digitize, tally, mathematics). The presumption is that Numerical Terms hyper-specify a claim, thus detracting from its universality.

Passivity: Words ranging from neutrality to inactivity. Includes terms of compliance (allow, tame, appeasement), docility (submit, contented, sluggish), and cessation (arrested, capitulate, refrain, yielding). Also contains tokens of inertness (backward, immobile, silence, inhibit) and disinterest (unconcerned, nonchalant, stoic), as well as tranquillity (quietly, sleepy, vacation).

Present Concern: A selective list of present-tense verbs extrapolated from C.K. Ogden's list of "general" and "picturable" terms, all of which occur with great frequency in standard American English. The dictionary is not topic-specific but points instead to general physical activity (cough, taste, sing, take), social operations (canvass, touch, govern, meet), and task-performance (make, cook, print, paint).

Past Concern: The past-tense forms of the verbs contained in the Present Concern dictionary.

Praise: Affirmations of some person, group, or abstract entity. Included are terms isolating important social qualities (dear, delightful, witty), physical qualities (mighty, handsome, beautiful), intellectual qualities (shrewd, bright, vigilant, reasonable), entrepreneurial qualities (successful, conscientious, renowned), and moral qualities (faithful, good, noble). All terms in this dictionary are adjectives.

Rapport: This dictionary describes attitudinal similarities among groups of people. Included are terms of affinity (congenial, camaraderie, companion), assent (approve, vouched, warrants), deference (tolerant, willing, permission), and identity (equivalent, resemble, consensus).

Satisfaction: Terms associated with positive affective states (cheerful, passionate, happiness), with moments of undiminished joy (thanks, smile, welcome) and pleasurable diversion (excited, fun, lucky), or with moments of triumph (celebrating, pride, auspicious). Also included are words of nurturance: healing, encourage, secure, relieved.

Self-Reference: All first-person references, including I, I'd, I'll, I'm, I've, me, mine, my, myself. Self-references are treated as acts of "indexing" whereby the locus of action appears to reside in the speaker and not in the world at large (thereby implicitly acknowledging the speaker's limited vision).

Spatial Awareness: Terms referring to geographical entities, physical distances, and modes of measurement. Included are general geographical terms (abroad, elbow-room, locale, outdoors) as well as specific ones (Ceylon, Kuwait, Poland). Also included are politically defined locations (county, fatherland, municipality, ward), points on the compass (east, southwest) and the globe (latitude, coastal, border, snowbelt), as well as terms of scale (kilometer, map, spacious), quality (vacant, out-of-the-way, disoriented) and change (pilgrimage, migrated, frontier.)

Temporal Awareness: Terms that fix a person, idea, or event within a specific time-interval, thereby signaling a concern for concrete and practical

matters. The dictionary designates literal time (century, instant, mid-morning) as well as metaphorical designations (lingering, seniority, nowadays). Also included are calendrical terms (autumn, year-round, weekend), elliptical terms (spontaneously, postpone, transitional), and judgmental terms (premature, obsolete, punctual).

Tenacity: All uses of the verb "to be" (is, am, will, shall), three definitive verb forms (has, must, do) and their variants, as well as all associated contractions (he'll, they've, ain't). These verbs connote confidence and totality.

Source: Hart [11]

Coaching research field:

cognitive structure and its evolution over time

Sybil PERSSON

ICN Business School – CEREFIGE

Pôle Lorrain de Gestion

13, rue Michel Ney, 54000 Nancy - FRANCE

E-mail: sybil.persson@icn-groupe.fr

Tel: +33- 670-343-804; Fax: +33-383-173-080

Silvester IVANAJ

ICN Business School – CEREFIGE

Pôle Lorrain de Gestion

13, rue Michel Ney, 54000 Nancy - FRANCE

E-mail: silvester.ivanaj@icn-groupe.fr

Tel: +33-383-173-774; Fax: +33-383-173-080

Extended abstract:

Coaching is an increasing field of research in the occidental world. Within the "developmental interactions" which is the expression chosen by D'Abate et al. (2003) to embrace the different ways for employee development in a popular approach, cousin relational practices exist: mentoring, coaching, apprenticeship and tutoring. If these developmental interactions have received attention from practitioners, consultants and researchers, the differences between them are not clear.

In the management sciences, coaching appears mainly as a tool for learning, a support for managing people and a way to increase performances within an organization. It appears that different outcomes can be highlighted because different points of view exist, in the methodological meaning of the term, but also in a cultural perspective. We considered that

these differences need to be assessed by taking into account the cultural background of the practitioners involved in the coaching stream, but also the disciplinary background of the scholars in the social sciences who have lead research on this controversial issue.

This study seeks to provide a response to concerns regarding the construction of knowledge and future choices by studying the intellectual structure of coaching research and asking the following question: "What is the overall structure of the coaching research field, and how has it has evolved over time"?

To answer this question and analyze the content of the field, we use a scientometric approach, the method of co-word analysis (COA). This method allows us to perform a quantitative analysis of the literature to identify the main areas of research in a particular scientific field and their development over time (Callon, Courtial & Laville, 1991; Callon, Law & Rip, 1986; Coulter, Monarch & Konda, 1998; Courtial, 1994). The COA is based on the theoretical assumption that the COA in scientific texts can adequately capture their content. The method thus measures the strength of the links between two documents by calculating the co-occurrence of the same words in these scientific or technical documents (which are representative of the field under consideration). These words, often known as "descriptors", are scientific concepts, ideas, or pieces of knowledge that represent a scientific field or area. They are recognized by the scientific community working in that field as required points of passage for all authors who ultimately seek to work in the field (Callon, Law & Rip, 1986). COA has been successfully used by a number of authors to explore the evolution of several scientific fields, including ecology (Neef & Corley, 2008), robotics (Lee & Jeong, 2008), information security (Lee, 2008), economics (Cahlik, 2000 ; Cahlik & Jirina, 2006), polymer chemistry (Callon, Courtial & Laville, 1991), software engineering (Coulter et al., 1998), information retrieval (Ding, Chowdhury & Foo, 2001)**,** informatics systems (Larsen &

Levine, 2005), biotechnology (Rodriguez, Janssens, Debackere & De Moor, 2007), and fuel cells (Hassan, 2005).

This paper is composed of two major parts. First, we describe our choice of methodologies and techniques as used in COA. This choice essentially consists of processes for collecting and preparing data and the process of analysis used to generate clusters and understand their contents. Then, we present the main results obtained from the cluster analysis and their place in the structure of the field.

The data analyzed were collected from a set of 1500 papers published in academic journals that are recognized by the research community as the most representative of the field. Some of these areas are well developed and represented by central concepts. Other areas that have been tackled intensively in the past are currently losing their central position and becoming peripheral subjects. This drop-off has been countered by the emergence of other central areas with strong potential for future development. Finally, several dimensions remain obscure and occupy a marginal position in the field.

Keywords: COA, coaching, science mapping

References

Cahlik, T. & Jirina, M. (2006). Law of cumulative advantages in the evolution of scientific field. *Scientometric*s, 66(3): 441-449.

Cahlik, T. (2000), Search for fundamental articles in economics. *Scientometric*s, 49(3), 389-402.

Callon, M., Courtial, J.-P. & Laville, F. (1991). Co-word analysis as a toll for describing the network of interactions between basic and technological research: The case polymer chemistry. *Scientometric*s, 22, 153-203.

Callon, M., Law, J. & Rip, A. (1986). Mapping of the dynamics of science and technology. MaCM illan, London, UK.

Coulter, N., Monarch, I. & Konda, S. (1998). Software engineering as seen through its research literature: a study in co-word analysis. *Journal of American Society for Information Science*, 49 (13), 1206-1223.

Courtial, J.P., (1994), A co-word analysis of Scientometrics. Scientometrics, 31(3), 251-260.

D'Abate, C., Eddy, E., Tannenbaum, S. (2003) What's In A Name ? A Literature-Based Approach To Understanding Mentoring, Coaching, and Other Constructs That Describe Developmental Interactions, *Human Resource Development Review*, Vol 2 No 4, pp 360-384

Ding, Y., Chowdhury, G.G. & Foo S. (2001), Bibliometric cartography of information retrieval research by using co-word analysis. *Information Processing and Management*, 37, 817-842.

Hassan, E. (2005), The evolution of the knowledge structure of fuel cells. *Scientometrics*, 62(2), 223-238.

Larsen T.J. & Levine, L. (2005). Searching for management information systems: coherence and change in the discipline. *Information Systems Journal*, 15, 357-381.

Lee, B. & Jeong Y.I. (2008), Mapping Korea's national R&D domain of robot technology by using the co-word analysis. *Scientometric*s, 77(1), 3-19.

Lee, W.H. (2008). How to identify emerging research fields using scientometrics: an example in the field of Information security. *Scientometric*s, 76(3), 503-525.

Neef, M.W. & Corley, E.A. (2008). 35 years and 160,000 articles: A bibliometric exploration of the evolution of the ecology. *Scientometric*s, DOI: 10.1007/s11192-008-2099-3.

Rodriguez, V., Janssens, F., Debackere, K. & De Moor, B. (2007), Do material agreements affect the choice of research agendas? The case of biotechnology in Belgium. *Scientometrics*, 71(2), 239-269.

**Addressing Category Management Dilemmas and Integrating Shopper Marketing
to Enhance General Management Performance**

Sophie Schwartz
1421 boulevard du Mont Royal
H2V 2J5 Outremont Montreal, Quebec Canada
sophieschwartz.qc@gmail.com
EyeLevel North America

January 2011

ABSTRACT

Category management is well recognized and reputed to be a valuable concept, however, this role is still difficult to integrate into existing organizations. Category management is often controversial internally, although companies, both suppliers and retailers, rarely admit this fact. Theoretically, category management is valued as supporting a company's collaborative way of driving go-to-market initiatives. In practice, operational category management resources tend to be restricted to continuous analytical tasks or to be mobilized for massive ad-hoc initiatives. Increased access to shopper data is changing the landscape in today's business world. Consumer Product Goods (CPG) companies must review the position and attributes of their category management roles.

Various studies have been carried out examining the collaboration between suppliers and retailers as well as investigating industry best practices. There is no or little research, however, on the role of category management as a distinct department within organizations, nor on its managerial impact.

In this paper, we suggest ways in which category management could be used to develop key competencies in a manager's career path. Focusing on the extension of category management skills toward shopper marketing, we identify how CPG companies can reach new performance areas for their business via customer-oriented mindset. By acknowledging today's market reality using the Shopper concept, there is an opportunity for manufacturers to review their internal organization and to strengthen cohesion within their company.

Our methodology was to conduct guided interviews and discussions with both category managers and with industry leaders. Combining practice in CPG companies and published managerial concepts, we provide action-oriented guidelines to be used as managerial tools. Our aim is to facilitate initiatives in organizations and for individuals while embedding Category Management skills, especially Shopper Marketing. This approach, when applied in a way that includes all players within a company, will provide market success.

KEY WORDS: CATEGORY MANAGEMENT / CONSUMER PACKAGED GOODS / SHOPPER MARKETING / CUSTOMER / LEADERSHIP PIPELINE

INTRODUCTION

Moving toward Category Management capabilities

The need for Category Management, and therefore the role of category manager, first appeared in retail companies when they moved to larger-size supermarkets or hypermarkets. Their actions needed to be driven at the category level. Category managers needed to master operational tasks in order to solve the complexity coming from larger assortment of products in stores.

Retailers expanded their sales surface through new store openings and acquisitions and through new store formats. They quickly faced a new stage of complexity: broader assortment, in more outlets, managed by more store-managers, under numerous banners, and across new channels. Mass-market retailers continued to use size increase to deliver growth and needed to operate with high volumes and revenues to offset fixed costs like storage, distribution and IT structure.

In the 1980's, as retailers were growing, they were offered a wider choice of products from manufacturers. Major suppliers were establishing their brand names and increasing the number of their products. At the same time, distribution banners provided additional retail spaces.

A new field of practice emerged, known as Merchandising in Europe and Shelf Management in North America. Merchandising gave answers to questions such as: where should we put this product in stores? Which layout should we choose and how would we make it effective? How do we facilitate the management of Point-of-Sales (POS) for staff? How do we make purchase easier for shoppers? How do we increase revenues or margins per shelf meter?

A number of merchandising publications and sophisticated tools were developed, dedicated to shelf management methods and quantitative modeling. They extended to various practices and solutions supporting promotional merchandising (e.g. promotional displays and innovative promotional layout) and their related analytics. Merchandising could be described as the discipline covering everything that related to the contact between items and shoppers. This would include strategic guidelines, logical organization, in-store furniture and the allocation of space per unit. The objective was to increase the sales performance of a specific category in order to fulfill its assigned role.

Competition increased between banners. Retailers reacted by further investing in communication and promotions - temporary price discounts to stimulate purchases. At the same pace, manufacturers increased their own promotional offers. They introduced new roles handling Trade Marketing, also referred as Trade Promotion in some companies. Trade Marketing covered all aspects of promotion including POS considerations, trade leaflets, impact analysis, branding extensions and in-store events. Depending on business scale and strategy, manufacturers could create customized promotional activities for a single retailer.

As a consequence, most modern CPG companies had two functional areas bringing their expertise to sales: the merchandising team and the trade marketing team. These were usually part of the Sales Department as they addressed implementation at the store level, providing solutions to retailers and guidance to the supplier sales force. These two areas of expertise are still key for

both CPG companies and retailers and they mobilize specific skill-sets. They require solid technical support as well as data quality and consistency in order to build experience on benchmarks and well-understood ratios. Performance indicators are usually specific per category.

Over time, Merchandising and Trade Marketing tended to be consolidated in the role of Category Manager. As market analysis and solutions became more sophisticated Merchandising and Trade Marketing have reemerged as distinct roles in some businesses. Category management retained its primacy, supported by the other two fields of expertise. Depending on scale, Merchandising and Trade Marketing could either be provided by dedicated resources or third parties.

Evolution of the Category Management role:
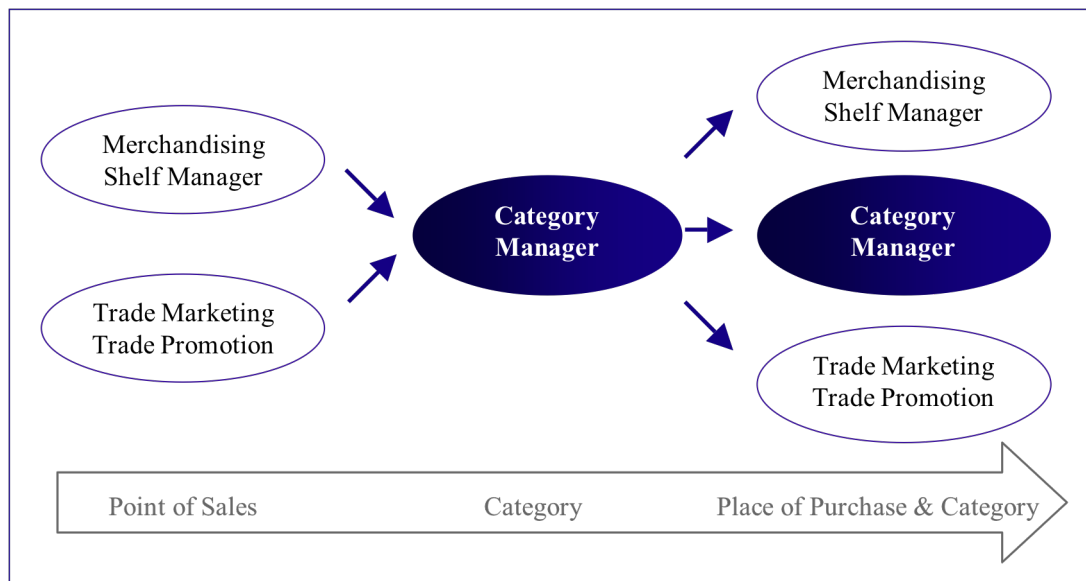
The essence of the category management role comes from the need to support in-store activities. This dimension remains fundamental today and should be understood as a "Place of Purchase" (POP) rather than a 'Point of Sales" (POS). It applies to virtual as well as physical points of contacts with shoppers, across all channels.

In addition, the management of product assortment required specific skills and tools. These were used to take decisions for listing new products, to keep or remove items from the shelf and to extend products' lines. These often combine a large set of criteria, such as profitability (in percentage and in value), revenues (unit/category), on-shelf rotations, storage weight, shopper's loyalty, category's attractiveness and prices coverage. The knowledge of these indicators has been traditionally the responsibility of the category manager.

A product's Retail Sales Price (RSP) is still a sensitive subject in most companies. Because of its complexity, as well as for legal requirements, RSP has often been managed by marketing departments who advise recommended selling prices.

Another dimension that has solidified category management as a distinct role is the understanding of the Shopper. In the 1990's, the power of the shoppers' decision was

acknowledged. Today, we often talk about shoppers' empowerment. A restricted growth in mature markets, the rise of purchase intelligence at the household level, and day-to-day access to a larger range of shopping "places" and times exacerbates horizontal competition between banners and between brands. The development of Shopper Marketing in the category management skill-set at the strategic and operational level is currently on the agenda of most companies. We will explore the consequences of and opportunities offered by this new development in Part Two.

Depending on the region and organization, business jargon surrounding this topic varies. Here, we make a distinction between the Shopper (one who makes purchases, whatever the medium of purchase) and the Consumer (one who uses the purchased product). The Shopper always makes purchase to respond to the needs of a Consumer. These can be the same person or not.

We avoid the word "Customer" as it can be used to name different actors. A customer could be the client who lists a product from a Supplier (*case 1*), a client at POS or online store to purchase an item (*case 2*) or an end user who actually consumes the product (*case 3*).

Language used - the Go-to-Market Chain of Actors:
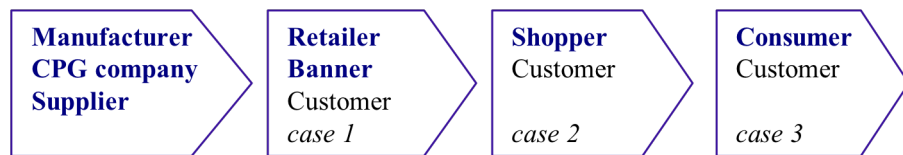
Documented Category Management Processes

In the early 1990's, Dr Brian Harris, Founder and Co-Chairman of The Partnering Group, defined and communicated the 8-step Category Management Process.[1] Practitioners applied the full process either to launch a category strategy or to lead category management projects. Category managers should be familiar with the 8-step method and associated concepts, such as category definition, category roles and category drivers.

In addition to the intrinsic value of the model, which has proven itself over time, the publication of the 8-step model was a key event. It provided a common vocabulary for category management. The value of documented concepts, public presentations and professional networks dedicated to category management is great. These are the first steps to creating a professional community and to providing references to business and human resources defining the role of category manager.

The turnover of category managers in operational business units is quite high. When newcomers assume these roles, documentation supported by expert training and mentoring helps to save time by focusing on the acquisition of skills instead of defining what category management means. A professional support environment is necessary to grow the category management's overall capability level.

Harris' 8-step formal process was created for practice between retailers and their key suppliers. Furthermore, some complementary approaches have been developed, each having a different perspective on implementation. As an example, major data providers like IRI and Nielsen provide methodology on store clustering based on their dominant shoppers' profile combined with local criteria. Some geo-marketing approaches are led with experts in this field. In another direction, a recent initiative was the Jointly Agreed Growth (JAG) Process published by ECR Europe (2008). The working sessions gathered a group of practitioners of retailers and suppliers that came up with a 5-step process, proposed to focus on joint planning.

The presentations made by category managers were often dedicated to best practice examples rather than theory. These combined success stories from leading brands and initiatives on key categories. When experts published category management cases, they often illustrate their company's involvement in strategic collaborative partnerships with retailers. For example, the presentations that were made during the Category Management Conference in Atlanta (October 2010) by Kellogg's, Pepsico, Johnson & Johnson, Georgia Pacific or during LSA ECR France conference (March 2010) by L'Oreal.[2]

In Europe, some manufacturers and retailers formalized concepts on category management up to the mid 1990's. Their work contributed to the content of the ECR Europe Bluebooks.[3] These documents remain key references for the profession. During recent years, market downturns and the internal restructuring of companies resulted in a shift of focus from demand-side projects to supply-side projects in most regions. Supply-side initiatives, like the need for increased effectiveness and cost reduction, were high on the agendas of many companies. ECR publications were more widely dedicated to supply management.

Recently, CPG companies have begun reinvesting resources into studying the role of shopper for their business. Previously, Shopper understanding was rarely managed at a board level or was limited to a few experts in a company. The results of these new avenues of study have produced results, such as growing shopper intelligence that will push the industry into new way of working supported by publications.

Structure of this Paper

The purpose of this research is to describe the role and the attributes of category management in manufacturer organization from a managerial perspective. Our methodology combined guided interviews with professionals, managers who had experience as category managers, and industry leaders. We also referenced several published studies. In our discussions with practitioners of category management, a number of common difficulties and point of failure arose which we discuss here. We discuss potential solutions to these issues while proposing some ideas that will work to increase market performance.

Part One of this paper explores how handling category management dilemmas can open new opportunities to managers in their professional career and serve as a preparation for future general management responsibilities.

We propose a framework for evaluating competency. It provides a simple practical reference scheme to be used by managers as well as by human resource experts, to help identify the potential benefits of a category management role based on the background of an individual. We adapt the Leadership Pipeline model developed by Charan, Drotter and Noel in 2001 to category management.[4] This model is a useful reference tool for managing career pathways and preparing business leaders, opening new perspectives for cross-functional roles as category management.

In Part Two, we identify aspects of the growing field of shopper expertise that are important for category managers. Shopper Marketing forms a large part of our discussions. We suggest that driving shopper understanding into action can strengthen the internal cohesion across organization and therefore contribute to the overall leadership performance.

This paper does not include a detailed roadmap of category management nor does it explore the implications of category management in the business context or market maturity. These are areas of interest beyond the scope of this publication.[5] Other areas for future study include distribution channel management, integrated business planning (especially national category planning), and the use of shopper insights. These areas need expert support at the retailer-level as well as at the manufacturer-level to simplify process and align practical contributions in business.

PART I. DEVELOPING MANAGER' KEY COMPETENCIES
VIA A CATEGORY MANAGEMENT ROLE

The objective of this section is to describe a framework for evaluating category management competency. There is a distinction between professional skills and competencies. Skills are specific to a profession and are acquired through practice, specific development programs or inherited from experienced practitioners. One's skill level determines their ability to carry out their role. Competencies, rather, operate within a frame of reference that considers all professions within a company. Their level is evaluated to be more, or less advanced per role, depending on the maturity and the expected contribution of the role in the company's activity. The level of competency does not vary significantly with company size and the acquisition of competencies proceeds at a pace related to each individual's learning curve. Competencies guide behaviors and they determine the way an individual practices his skills. Competencies are often referred to as "soft skills". When endorsed by company leadership, Action Learning programs are a highly effective method for developing both skills and competencies.[6]

Professional **Skills**	Professional **Competencies**
Specific skill set per functional area (eg Finance) -> define WHAT you are doing -> specific skills are required to accomplish tasks and deliver your goals	Cross-functional framework in a company -> define HOW you are working -> specific competencies that are mobilized to enable your skills practice

We will focus on category management competencies rather than skills. They allow us to better understand the integration of category management into a company structure and to identify potential for career development for a given category manager.

1.1. - PREDICAMENT #1
What is Category Management about? Could you describe your job in a simple way?

Describing one's job in a simple, concise way is a basic requirement for self-motivation and appropriation of a role. One needs to be able to define it in simple words to non-experts, to family, friends or colleagues within a company. Because of the nature of the category management role, this can be a challenge.

In our interviews, this difficulty people had in responding to this question was evident. In addition, some professionals expressed that they were feeling a lack of recognition for their contributions as few people could understand and value what their role was in the larger picture.

One notes that category managers working at international level did not indicate a self-esteem concern. Some had few contacts with retailers and worked more as category experts for marketing teams. These people managed manufacturer's large portfolio of products, integrating shoppers and markets trends, they were essentially partnering with brand's marketing teams and had a limited external exposure. Others were exclusively dedicated to a given international key-account. These people coordinated local teams' initiatives to meet the strategy agreed with the retailer's international team. In both cases, category mangers worked at a strategic level.

The Category Management Association defines Category Management as:

> Today's expanded definition of Category Management: Trading partners collaborating to determine the point of optimization in pricing, promotion, shelving, and assortment to maximize profitability and shopper satisfaction. Successful Category Management draws on the latest industry trends, leverages available data, and utilizes best-in-class technology from the leading solution providers.[7]

Key competencies:
Management of complexity, to draw simple insights from various sources, sophisticated data, and numerous performance indicators.
Entrepreneurship, to maintain internal and external momentum, adjusting to changing business priorities

For a CPG company, category managers define and conduct go-to-market initiatives, primarily partnering with marketing and sales teams. This role includes everything that deals with understanding the company's market dynamics, defining strategic business role for market categories, and integrating shoppers' insights. They work to adapt products or services to be sold in the most effective way to each retailer as well as purchased from retailers by shoppers in stores.

Making with category managers a simple exercise of formulating their role and just sharing it at board level would already contribute to a better understanding across the business.

1.2 PREDICAMENT #2
Content manage as an attractive role, but only for a limited time. Is it a role intended to serve as a pathway to sales or to marketing positions?

In CPG companies, both sales people and brand managers show high willingness to take on category management roles but an equal high tendency to quit them, seeing them only as temporary positions that could serve to help them reach higher responsibilities in their core profession or to move in a secured way from one area of expertise to another. Category managers rarely see a space for long-term evolution within category management.

Although this is not a quantitative survey, we noticed that it was quite different for category managers previously working in areas other than sales and marketing. In manufacturing companies, for example, professionals having a background in supply-chain, research analytics or with retailers did not consider their category management role as a temporary step in heir career path.

Below are some examples of career paths that include category management. Note that a category management role can include trade marketing, promotion, sales development responsibilities. This is related to a company's size, sophistication and internal language.

Simplified sales path:
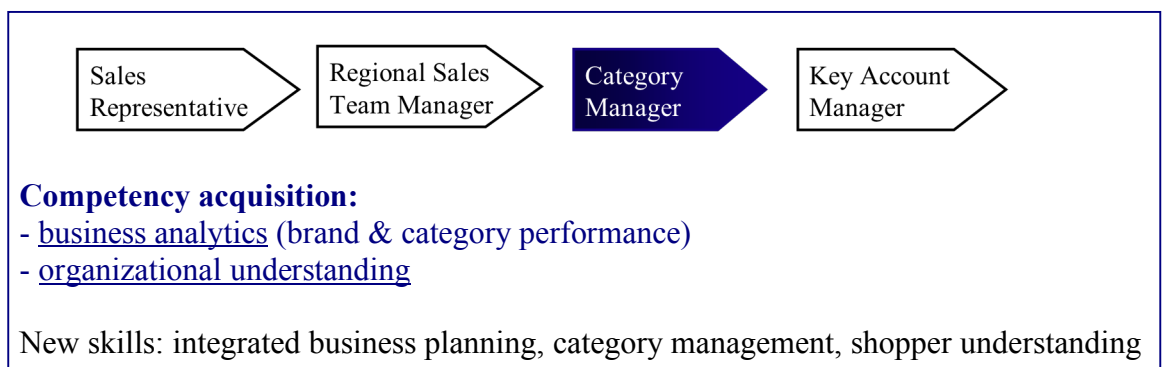
This excerpt from one of our discussions provides a good representation of the role's added value:

Why KAMs Need Category Management Skills Today More Than Ever?

When I first started as a KAM at Unilever in 1997, Category Management was just emerging in Canada and quite frankly was the domain of a few people in the sales team that were gathered in one corner of the office surrounded by mountains of AC Nielsen data books - that's right - literally huge binders of panel data that were sent to the offices once per month and stored in rows upon rows of cabinets.

With the arrival of Workstation software on my computer things changed forever. Unilever was very forward thinking in having all their KAM teams to be proficient in Workstation and have at least the fundamental Cat Man skills so that they could find growth opportunities with their clients and be full participants in Category Management Projects. Cat Man analysis was firmly embedded in the KAM role whether it be for business planning, business reviews, opportunity gap analysis or full blown CatMan projects.

Of course, the true spirit of CatMan is an objective approach to the category that is not influenced by one particular supplier's view on the category and brands within. But for a KAM, CatMan skills can be used offensively to find opportunities to drive their brand growth. In today's tough economy where every sale at retail involves an intense battle, KAMs need to have these analytical skills to keep ahead of the competition and find opportunities for them to grow their market share within their customer.

We have developed a list of core competencies of a world-class KAM and Category Management Skills transcend several of these competencies including: Thorough Business Planning, In-Touch With the Retail Reality and Fact Based Selling.[8]

Simplified marketing path:
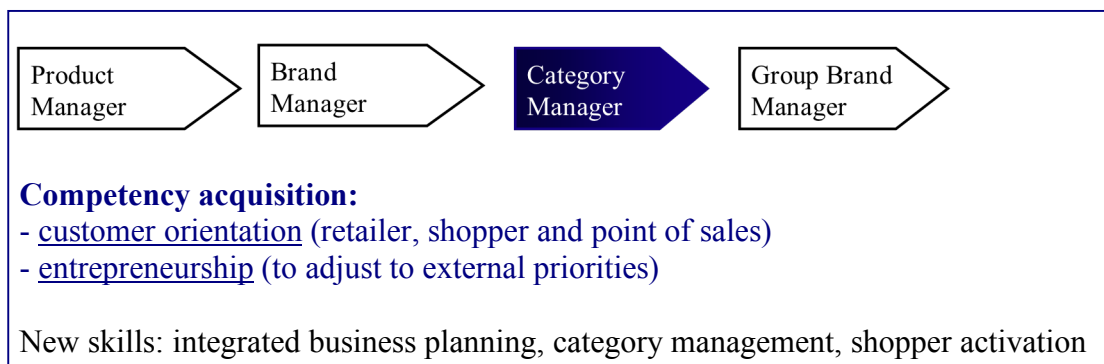
In both the simplified sales and marketing career paths, category manager role was seen as a way to reach higher responsibilities while returning to one's previous profession.

Category Management following a board confirmed experience
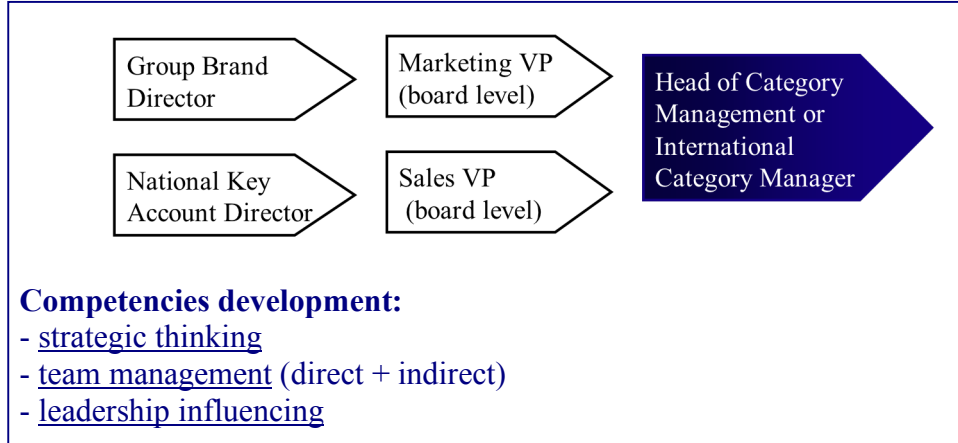
When moving to category management at a more advanced stage in their career path, people tend to remain in this position for a relatively long period (six years or more). In continuity with their previous position at the board level, the company often acknowledges the strategic level of their category management position.

In the cases we studied, category manager VP's were often integrated into strategic meetings, even if they were not on the company board. Keeping category management leaders in close relationship with other local boards members, general managers and marketing VP's increases the effectiveness of the category management role. In some instances, the approval of the head of category management is required for strategic initiatives to be accepted.

Category management roles offer opportunities to develop some competencies that are not traditionally accessible without experience as a business unit manager. There are similarities between those two roles related to new levels of complexity that need to be managed.[9]

Transition step between Sales and Marketing:
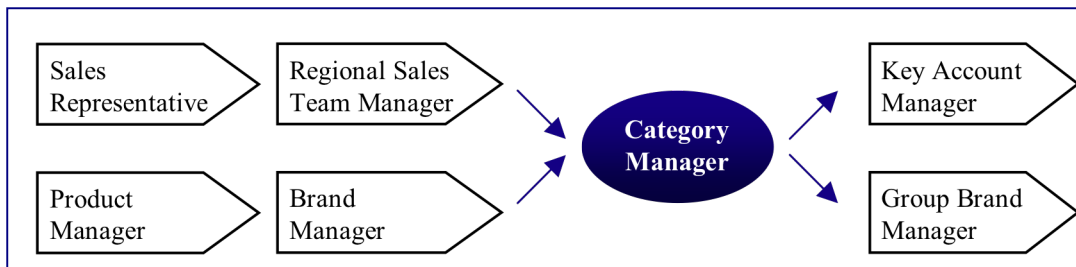
The category management role can be used to promote internal mobility between sales positions and marketing positions. It can facilitate an introduction either to the "external world", eg customers, when coming from marketing, or to the "internal world", eg brands and research, when coming from sales.

For example, Reckitt Benckiser's Corporate Work and Learn section states:

> We're a consumer-centric company and our marketing teams form the core of our future leadership. However, 'pure marketing' doesn't fit with our strong commercial focus - a typical career path in marketing will always include some overlap with sales. This ensures that our country general managers of the future are well prepared to deliver on the full range of commercial challenges they will face.
>
> A marketing-focused career will typically see you join as a commercial graduate trainee in a local business, moving between sales and marketing, from a brand manager up to category manager, then perhaps into a to global strategic role in head office (a global brand market manager for example). From one or two roles at this level, we would expect you to be able to prove yourself able to step up again to be a marketing director of a country and eventually, we'd hope, a country general manager.[10]

Exposure to new responsibilities through a category management role would be managed at the right speed and job performance would need to be accurately assessed. *"People who ticket-punch their way through jobs don't absorb the necessary values and skills".[11]*

1.3. - PREDICAMENT #3
Where should Category Management be placed in CPG companies? Should it be part of the Marketing Department or of the Sales Department?

There is no one, single recommendation to these questions.

According to Dr Brian Harris and Michael McPartland in their article "Category management defined: What it is and why it works" (*Progressive Grocer*, 1993):

> First, the organizational structure should follow the process. There is no single organizational structure that is required for category management. The requirements of implementing the process – in particular the development and execution of category plans – within the organization will determine the required organizational design.
>
> A fairly common organizational structure, however, seems to be emerging. It is built around integrated category management teams of two or three people, headed by a category manager. This team has responsibility for category planning and all buying and selling decisions to its assigned categories. These teams are generally supported by an information / analysis group and an overall coordinator.[12]

In the last fifteen years, under pressure from cost reductions and lower market growth, CPG companies had to reduce senior roles, downsize their local teams, move some positions to

regional level, and outsource activities. These measures, however, did not help to improve the succession plan per role nor did they to stabilize category management resources. A new survey on reporting structures has been recently initiated by the Category Management Association Organizational Development Share Group. It will provide an updated status of the situation. *Donna Frazier, (Founding Director Category Management Association), will present a review of the questions included in the Reporting Structure Survey. The survey results will be available at the end of 1st quarter 2011.*[13]

Category management and key account management

Mr Serge Cogitore highlighted the advantage of having a close relationship between key account managers (KAM) and category managers in *Comment optimiser sa stratégie commerciale en gérant des catégories de produits* (*Dunod*, 2003). In Chapter 14, he described growing interest from CPG companies in better understanding of retailers. In addition to sophisticated negotiations skills and solid finance acumen, this implies that key account managers would need to clearly understand the mode of organization and business strategy of their clients.[14]

From observed situations, corporate reorganizations have often caused temporary disruption, eg when M&A required new operational devices to ensure the reliability of their services and the quality of their account finance monitoring. Key Account Managers (KAMs) were required to fulfill support roles and restore customer management to appropriate levels on basics. Once it was secured, KAMs were able to expand their skills and benefit from their collaboration with category managers. In some CPG businesses, key accounts manage a "Customer Team", bringing together their company's functional experts to deliver a customer specific plan. At a higher level, new roles called "customer marketing", "customer" in this instance referring to "retailer", were created to develop a holistic marketing management of large customer accounts.

We have identified three reporting scenarios for the placement of Category Management within business units.

Category Management placement scenarios

	Category Management integrated into the <u>Sales Department</u>	Category Management reporting to <u>General Management</u> although not necessarily sitting at company board	Category Management integrated into the <u>Marketing Department</u>
S T R E N G T H S	External activities are turned **into action** through frequent contacts with KAMs and with retailer's category management teams.	Category management's independence leading to business development recognized as **factual** targets. Continuity leading to a higher level of capabilities, and especially new **shopper marketing** skills, that are complementary to category management and trade category management know-how.	Category and shopper contributions are more fluently taken at a **strategic** level, through direct contacts with brand teams in the marketing department.
W E A K N E S S E S	Less **data** available (shopper data and / or POS data, lower investment in human resources and opportunities to use acquired data), because the research and budgeting departments tend to report to Marketing, where lots of consumer data are already in place	Risk of becoming isolated **experts,** who may be disconnected from the company's operational priorities and thus have limited impact on external operations. Risk of losing the daily reality at the POS level.	Initiatives at the category level can be less relevant for the strategy of the **retailer**. Possibility of inappropriate timing and insufficient contribution to customer profitability.

Source: EyeLevel-NA

The category management challenge lies in maintaining the integrity of the discipline while increasing the capacity of strategic influence not only externally, with the retailer, but also internally, within the company itself.

Key competencies:
Category managers require <u>organizational understanding</u> to practice their collaborative mindset. The importance of working in a collaborative mode is highlighted by all the experts and reference communities involved in category management.

Example of cross-organization collaboration, between manufacturer and retailer:

> Collaborative Consumer Relationship Management is an integrated ECR demand-management strategy, which helps manufacturers and retailers to jointly recognize and value consumers'/shoppers' individual needs and tailor their offers to them. (*ECR Europe's Guide to Collaborative Consumer Relationship Management*, 2003).

Example of cross-functional collaboration, within a business:

> Category management promotes cross functional working between companies and will generally involve people from buying, finance, supply chain, trade marketing, space planning, store operations, sales, product development and marketing & of course category management. (*IGD Publication*, 2010-2011)

Collaborative mindset at the individual level is not enough to be implemented effectively. In a company, it must be accompanied by appropriate organization, by processes of decision-making across roles, and by the coordination of performance measurement. This is the responsibility of general management.

Until now, the emphasize has been on collaborative behavior to establish externally, between retailer and manufacturer. This dimension is well understood in category management theory, and often takes the form of empowering the retailer with a "go/no go" decision: the company would invest resources in particular category management initiatives only if the retailer's team shows its commitment. However, the situation is less clear when it comes to working internally and balancing dynamics between professionals within a business. Companies that develop mechanisms that ensure good relationships between their departments without reducing the leadership of category management in-house will have a competitive advantage.

Sales professionals also emphasize the need of ability to communication. The role of general management in regard to this is crucial - they must make sure to create space for communication and accessible listening conditions that allow category managers to offer recommendations and to bring up smaller issues before they become larger problems.

An experienced professional shares in discussion:

> Number one will always be understanding accurate information, without that the manager no matter how good they are will be suggesting incorrect action. However in my experience even with good data many managers fall down on the communication.

> Often they are steam rolled by more dominant buying or operations teams, who don't want to take on board valid recommendations. When category managers are not seen as core to senior management they will always play second fiddle to the other commercial teams.

> A successful manager will be able to understand the data, and then communicate the insight in a language that speaks to the broader audience.[15]

One should be careful to avoid confusion between the role of category management and the role of operations manager. When a category manager is made to act as a mediator between sales and marketing, they are unable to perform their core tasks. Category managers lose the opportunity to develop their personal skills and are unable to invest time in Category and in Shopper Marketing initiatives, even though these should be among their highest priorities.

Some advantages of a collaborative framework
The willingness to generate collaboration externally between CPG companies and retailers as well as internally within companies motivated the creation of a proposal called the *Jointly Agreed Growth (JAG) process*, "News ways of working together" (*ECR Toolkit* , 2008).[16]

On the basis on their experience, some European retailers and manufacturers professionals worked together to define potential collaborative processes. They were familiar with the category management concept before starting the project and were convinced of its value. The group began with he underlying premise that working together is productive and would increase chances of a successful market initiative. The JAG recognized that breaking down barriers internally within a company is generally more difficult than accomplishing the same with those between retailer and supplier. They proposed a short and pragmatic process to facilitate this process, which included easy-to-use templates and samples. The goal of their proposal is to allow companies to spend more time looking at ways to increase efficiency in business and to offer innovations to shoppers, rather than reinventing the wheel for every product launch. They also defined an operational framework to welcome more shopper initiatives.

<div style="border:1px solid">

JAG Principles

. Grow the grocery market in a sustainable, profitable and competitive way to ensure the highest shopper and consumer satisfaction

. Deliver commercial benefits for each party

. Be open to every company bringing in knowledge and capabilities, regardless of its category

. Establish a fact-based joint process in which data and insights are shared

. Be committed to a clear execution plan

. Allocate the required resources for the execution of the plan within the agreed time frame

. Review the JAG process on a regular basis

. Respect the Confidentiality Agreement

. Comply with all laws: competition, health, environment, and intellectual property

. Involve senior commitment to ensure the 3-year time frame engagement required by the JAG

</div>

Source: JAG Toolkit, page 11

1.4. The Category Management Competency Framework

The Category Management Competency Framework can be adapted to any company competency reference frame. It consists of three key dimensions:

(1) Business understanding: these competencies define one's ability to grasp and to integrate business metrics, to properly make correlations between data and indicators, to master the complexity of data that comes from different sources and in different amounts, and to develop

strategy. Environment where detailed data are scarce, such as an in emerging markets or in a new business sector, are as demanding as markets where it is abundant, like in the CPG sector.

(2) <u>Interrelationship management</u>: because of the high density of internal and external relationship required in category management, inter-personal skills are important. One must have an awareness and understanding of which type of managerial type would suit to a given situation. Team management is particularly demanding, as category management teams are made of people having different experience and background whereas the other departments would be more homogeneous.

(3) <u>Performance management</u>: these competencies are self-explanatory. The difficulty is in how to assess performance success. Measures for performance in category management are not detailed in this paper, although a thorough exploration of the topic will be required to define reliable performance standards that can provide consistent results. This is especially true with marketing and sales, as success depends on the ability to balance the long term project and the short term day-to-day performance. In 1993, Dr Bryan Harris and Michael McPartland said: *"the work involved ... is significant (it will normally take at least three years to implement category management) ... but the benefits are considerable. "*.[17]

<u>The Category Management Competency Framework:</u>

Competencies	Meaning for Category Management
	<- Operational role Department Leadership role ->
(1) Business understanding	

Business analytics
Management of Complexity
Strategic Thinking
	To understand data flows and tools
	         To manage data and different sources
	                   To drive innovation and differentiation

(2) Inter-relationship	

Organization Understanding
Leadership Influencing
Team Management
	To develop collaborative areas
	         To raise the bar, esp. on shopper insights
	                   To coordinate team and functional networks

(3) Performance delivery:	

Customer orientation
Ability to Communicate
Entrepreneurship
	To design actions up to in-store execution
	         To transform expertise in shared projects
	                   To motivate and keep momentum

Source: EyeLevel-NA

The advanced level of the role in a company can be determined, using a benchmark with peers in other departments, based on "work-levels", "grades" or similar indicators. One will not need to

use each competency at a most advanced level in all category management positions. This will be related to each individual role's responsibilities, the business context, and the maturity of the company regarding category management. The advanced level per competency from operational of managerial responsibilities is shown by a shift from white to colored areas in the chart above but the competency set remains the same regardless of seniority level.

To strengthen the feeling of ownership over competencies evaluation practices, one should create the evaluation standards with the input of those who will be evaluated. The meaning of generic competencies into real situations that are specific to the manager's company creates a commitment to respecting performance criteria and support of evaluation results.

Ultimately, the competency framework should be used as a development tool, highlighting a small number of competencies for an individual to work on during a specific period of time. It should equally help to define the contribution of the category management role to the further development of their career path.

Leadership development with regard to the contribution of each was described in *The Leadership Pipeline model*, first published in 2000 by three coauthors Ram Charan, Steve Drotter and Jim Noel. This work highlighted the value to an individual moving through different types of management roles. It also defines the critical inputs of the intermediate "*passages*" to develop leadership effectiveness.[18]
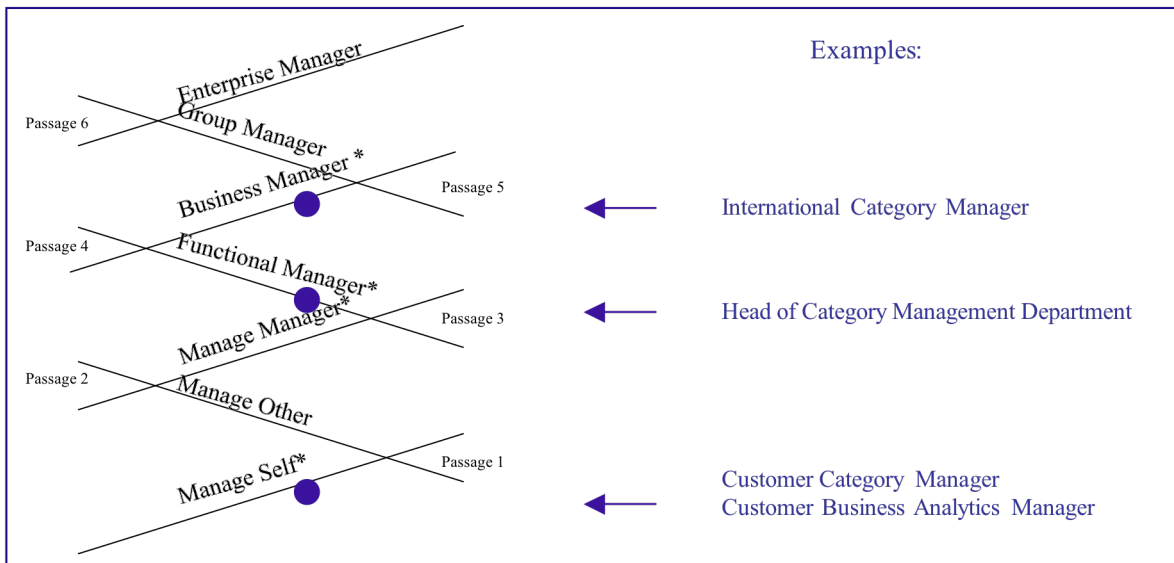
Source: The Leadership Pipeline Model (*Pipeline Model for Small Business)

In *Leadership Passages*, the authors David L. Dotlich, James L. Noel and Norman Walker, explore 13 selected "*passages that can occur in the life of a leader*". They also describe in detail a number of "*intense*" situations and they give *tips and techniques* so that theses stages "*can serve as a career roadmap*".[19]

Category management role should not be considered in isolation. We suggest using the model to better determine the stretching dimensions of a category management assignment for each

individual in a given professional context. On the right side of the chart, we give examples of where category management role would be in the Leadership Pipeline.

Looking at category management from this perspective, we see that the position can offer important potential for development that are rarely exploited by individuals or companies and that could provide performance gains for both. The experience of maintaining a constant orientation toward customer satisfaction combined with exposure to multiple roles is a unique experience that builds future leadership capabilities.

Part II. CONDUCTING SHOPPER MARKETING
TO BOOST PERFORMANCE AND COHESION OF BUSINESS

During discussions with professionals in the category management field and market trend analysis, the topic of Shopper Marketing emerged as an important one. They identify new opportunity for business management to deliver superior performance and also rally teams around a common interest: Shopper.

A large amount of data in business

CPG companies today have access to a large amount of data and many managers find the task of having to utilize and integrate this data daunting. There is an awareness of the significant amount of time required for training with new tools and to extract meaningful results from raw data, to understand their meaning, and to present them in clear, pragmatic ways, often with few standard models to follow.

The level of workload required to accomplish this should not be underestimated. Moving toward increased reliance on shopper data requires significant resources to define which team should receive a given set of data, in which format it should be made available, how the data can be connected to the company goals, and to determine how much time this will require.

Category managers often find that they are requested to assimilate additional datasets to support decisions and to assess their future impact on shopper investments. This is a frequent issue as new technologies and techniques continually provide access to a new set of data. In the past, ad-hoc shopper studies were carried out once every five years and the results were more easily assimilated into a company's existing practice. These studies are still used but they have been augmented with access to frequently updated data at the EAN code level.

Shopper Marketing

A wide array of jargon surrounding Shopper Marketing exists like Consumer and customer solution, Shopper excellence, Shopper Marketing, Consumer & Marketing Insight, Shopper

Development, etc…Meaning can vary from company to company. Below we provide some working definitions of some of those terms.

In this paper, we will use the words <u>Shopper Marketing</u> to identify the category management skills that relate Shopper understanding to Shopper insights and drive initiatives that target shoppers' decisions at both store and brand levels.

We identify three major stages of Shopper Marketing areas for implementation, based on sophistication of data availability, frequency of data collection and data accuracy.

<u>Shopper Marketing Stages</u>:
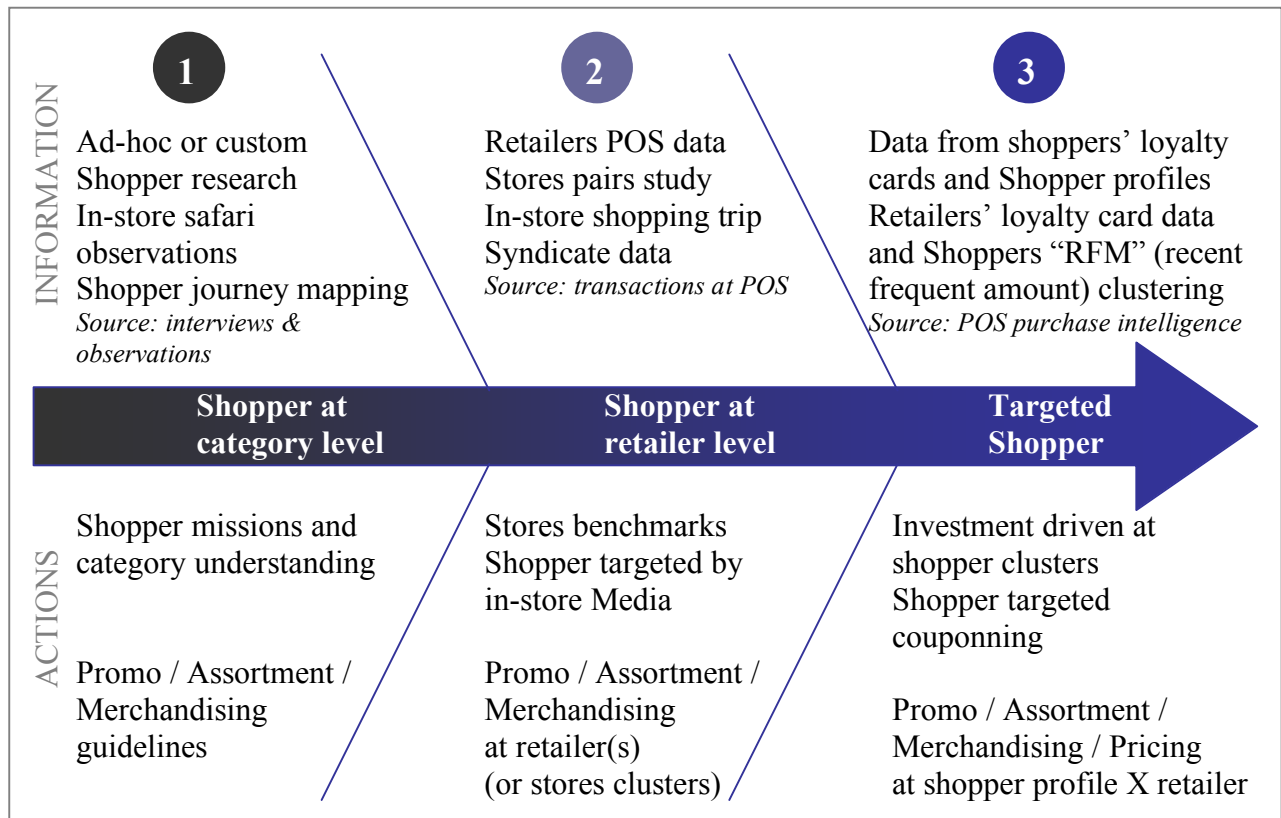
All three stages are used by category managers and can be used to define different actions. The access to shoppers' day-to-day purchasing behavior data brings a new dimension. Many retailers have not yet developed the capability to use the Shopper data in this way despite having tremendous transactional databases at their disposal. This will require strategic commitment and investment on the part of the retailers.

New technologies play a decisive role in understanding Shoppers. They enable massive data collections and an access to updated databases that could see extended use by category managers and marketers. As described in Information Week, (2010 9 August, issue 1 275) *"Success in the*

Big Data era is about more than size. It's about getting insight from these huge data sets more quickly".[20]

Some expert marketing companies collaborating with retailers have been able to master the technologies required to access and effectively use massive databases of shopper purchases. During the last decades, these developed an advanced expertise in shopper marketing solutions, in targeting shoppers segments and in making new initiatives practicable for retailers as well as for manufacturers. Catalina Marketing provides an example demonstrating this capability, their website states that: *"utilizing a sophisticated database, Catalina Marketing enables you to reach shoppers and patients with precision reach on a truly mass scale."[21]*

There is huge opportunity for understanding the impact of business decisions on your product performance and for assessing category health. Through the new access to data, integration of shopper insights can occur in the making of customer and marketing plans.

As summarized by Clive Humby, co-founder of dunnhumby:

> If companies put their customers first in every decision, they become their customer's first choice – which improves their brand value and business performance. It's a simple idea, but it has proven incredibly successful as the companies we've worked with see the value of being relevant to their customers. It is the understanding that customers – the people in the store taking the product off the shelf – make decisions every day based on a changing set of wants, needs and desires. That means the 'customer loyalty' expected and sought by some companies is actually a rarity.[22]

This requires significant changes including new guidelines for general management and decisions about how to move to a "customer-centric" mindset. Some leaders seize this opportunity to refocus their business on their core objective: their clients. When they embrace this strategy, it can prove to be highly motivating for their teams.[23] Furthermore retailers have experienced measurable shopper commitment due to shopper-centric initiatives. Because of this, there is some acceleration in the amount that new retailers are investing on shoppers

Some examples of this are seen within Tesco and Kroger's market communications:

> Tesco consistently invested at better understanding and rewarding its shoppers thanks to its Clubcard program over the last 15 years. In April 2010, the Chief Executive Officer of Tesco, Sir Terry Leahy, said: *"The industry has seen a significant slowdown in sales and we have not. And our relative performance has improved markedly and we are now the leading industry in growing market share"*. (*The Independent*)[24]

> The retailer turned to the most powerful weapon in its armoury - one that (Sir Terry) Leahy himself developed as marketing director - the Clubcard, to lock in its shoppers. The double Clubcard points scheme has been a key driver in helping Tesco to claw back market share from its rivals. (*Retail Link* -June 18, 2010)

Kroger has consistently taken a leadership role in projecting its value image this year (2010), with programs…. The initiatives fit into Kroger's broad "customer first" philosophy, in which the needs of the shopper guide the company's strategic direction. "These types of programs and our associates' exceptional ability to execute them well are just some of the reasons Kroger's business is growing," Dave Dillon said, CEO of Kroger supermarkets.[25]

Consumers love the program. While most gas retailers across the county are experiencing a decline in sales, Kroger is registering solid same-store growth. (*Fortune 500*, May 3 2010)

As a consequence for CPG manufacturers, when retailers give access to their shopper data, the manufacturers need to commit to remaining customer-focus as it becomes the priority for their retail client but also to further develop their brands' market share performance.

In an interview led by us in January 2011, Steve de Rose, VP Global Customer Development Excellence of Unilever, drew a clear picture of what the Shopper means to a Manufacturer and is worth quoting at length:

Everything we do must begin with the shopper... Who is she? Why did she choose a specific format or retailer today? What influences and captivates her when she enters the store? How many interruptions did she encounter on her way to the item she wants to buy... signage, kiosks, displays, demos? Is she pleased that she can find what she is looking for? Is there good communication that helps to educate her or help her understand new products, meal ideas, environmental benefits, health and wellness benefits? Does the in-store shopping experience entice her to look at other products and categories, or do the conditions frustrate her...poor layout, cluttered shelves or aisle, out of stocks? The questions are almost endless and they can vary based on the original intent of the shopping trip, the format, whether she has children with her, the amount of time she has available and the amount of money she has to spend.

Our retailers know the shopper best both from loyalty card data and from the daily hands-on face-to-face connection. As a supplier, we know less about the shopper. Our visits to retail shops is infrequent, we don't visit every format, we are likely at different income and education levels, and our daily regimen is likely quite different.

Learning who the shopper is should not only come from data sources but should come from in-store experience. Spend time in a store as a shopper, not as an employee of a manufacturer. Don't only look at the products and categories you represent. Shop different categories, shop the competitor's brands, shop the store brands, understand what the retailer is doing to attract her and keep her loyal. See the store the way she sees it. Shop meats, fish and produce and see and feel how a shopper sees and understands quality. When in the check-out line, watch how the shopper handles the transaction. Does she pay with cash, check or credit card? Did she have enough money? Did she have to put something back due to going over her

budget? These may seem like small things, however if you understand the full in-store experience for her, it will provide valuable input into the products, the price and proposition we deliver.

Mr De Rose reaffirms the fundamental connection to everyday reality for the shopper in store, which was precisely why category management as a field in the first place:

> To win in today's value oriented, time starved, health conscious and environmentally friendly market, understanding the shopper is the only way to build a sustainable platform for growth. We must dig deeper into her world, experience her world and make her everyday routine easier, fresher, smarter and at a value that makes good sense to her.

Eventually, the value gained from truly integrating Shopper Insights is clearly seen by manufacturers:

> I believe success in the consumer products arena will be determined by those suppliers who are winning shoppers across channels and customers everywhere, everyday. Embed this into the culture and ways of working of the supplier / retailer and all parties will prosper...growing brands and categories while enhancing the in-store shopping experience.

Moving forward …

Shopper Marketing and Shopper Intelligence could inspire the next generation of organizational structure in CPG companies. In the last decade, some international manufacturers moved to coordinated reporting, enabling them to see their business in a consolidated way. They saw that at a global level, a single retailer's invoiced account could be larger than their local business units, all included, or even larger than their brand or category aggregated across countries. It has speed up the clustering of business units and this has triggered the creation of global brand teams, followed by global customer teams, then declined at regional and at country level.

CPG organizations choices:
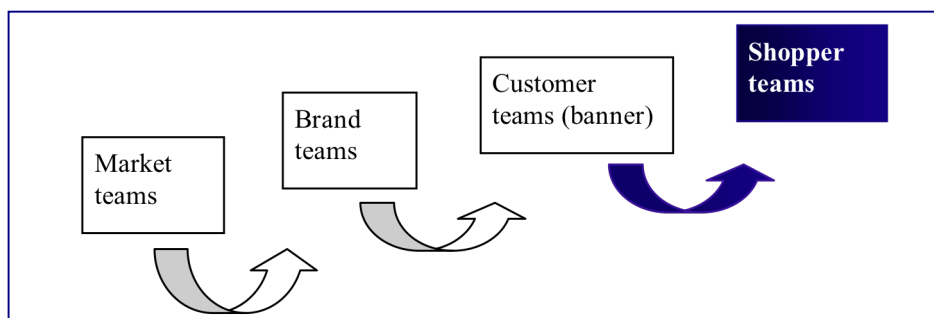
Some businesses lead actions at shoppers "clusters", also referred as "segments" or "profiles". Some retailers develop expert roles, each being specialized on a shopper profile, across categories, across banners, across stores, across boundaries.

This move becomes possible with shoppers clusters. These gather shoppers having similar purchase behavior pattern, they follow the same logic as the definition of categories, that gather products meeting similar needs. Each retailer defines its own shopper profiles as detailed in *Retailer-Developed Shopper Segmentation Classifications* in US grocery (source Hoyt and Company, Consumer segments TBA April' 07).[26] Although they were formulated in a different way, you would notice similar logics and converging shopper clusters across retailers.

Moving to implementation

CPG companies need to establish the correlation between their brands / products portfolio and shopper segments, preferably using published shopper profiles standards to keep it manageable. A differentiated advantage does no longer come from the access to shopper knowledge, but from the ability of manufacturers to integrate Shoppers insight into their business planning across functions. This evolution can be a significant opportunity for general management to re-engineer their business, defining performance driven innovative projects with their teams while being intimately connected to their market.

CONCLUSION

The highest barriers to category management remain internal, within companies. The use of competency standards and a better understanding of the value of a category management position in career path should help to anchor category management in business.

Those who have the ambition to reach general management positions should be seduced by the challenges offered by category management. At an early stage in their career, it gives the opportunity to early practice their competencies and address challenges such as:
- grasp complexity to offer simple breakthrough solutions
- experience the intimate competition between functions, esp. sales and marketing
- manage the unmanageable, high numbers of contacts and contradictory timings
- think customer, understand customer, act with customer
- enjoy delighting shoppers who will choose your brand, even if the value of shopper loyalty is not yet fully integrated in your business, all that without relying yet on company general management status.

The future opportunity for category management role lies in the huge development of Shopper Marketing. As Shopper Marketing is a strategy rather than a functional area on its own, it naturally opens wider the door for category managers to take a deeper role in the companies overall strategy.

REFERENCES

[1] Dr Brian Harris, Founder and Co-Chairman of The Partnering Group

[2] Category Management Association, professional association formed in 2004, http ://www.cpgcatnet.org

[3] ECR, http://www.ecrnet.org

[4] Drotter, Steve – Charan, Ram – Noel, Jim, *The Leadership Pipeline* – How to build the leadership powered companies (2001 *Jossey Bass*, Edition 2 2011).

[5] Lovelock, Christopher (Yale University) and Wirtz, Jochen W (National University of Singapore), *Services Marketing People, Technology, Strategy* (Sixth Edition, 1996) There are an increasing number of opportunities for applying category management approaches in sectors other than CPG, such as new technology companies, home businesses, and communication services.

[6] Drotter, Steve – Charan, Ram – Noel, Jim, *The Leadership Pipeline* - Chapter 9 (2001), Performance Improvement P160.

[7] Category Management Association, http://www.cpgcatnet.org/page/62774/

[8] Doucette, Jeff, Principal of Sales Is Not Simple International, LinkedIn Category Management Learning Forum (July 2010)

[9] Drotter, Steve – Charan, Ram – Noel, Jim, *The Leadership Pipeline* - Chapter 5 (2001)

[10] Reckitt Benckiser's Corporate Website, http://www.rb.com/careers/paths

[11] Drotter, Steve – Charan, Ram – Noel, Jim, *The Leadership Pipeline* (2001)

[12] Harris, Brian – McPartland, Michael, *Category management defined: What it is and why it works* (*Progressive Grocer,* 1993)

[13] Category Management Association
http://www.cpgcatnet.org/page/OrgDev/index.v3page;jsessionid=4sc2ehu3qqkdi

[14] Cogitore, Serge, *Le Category Management, Comment utiliser sa stratégie commerciale en gérant des catégories de produits* (*Dunod,* Paris 2003).

[15] Fisher, Chris - Retail Director at Athean Analytics Ltd, LinkedIn The Category Management and Shopper Insights Executive Group (November 2010)

[16] ECR Europe, Jointly Agreed Growth

[17] Harris, Brian – McPartland, Michael, *Category management defined: What it is and why it works*

[18] Drotter, Steve – Charan, Ram – Noel, Jim, *The Leadership Pipeline* - page 7, 29

[19] Dotlich, David L., Noel, James L., Walker, Norman, *Leadership Passages , The Personal and Professional Transitions that Make or Break a Leader* (*Jossey Bass,* 2004).

[20] Catalina Marketing web site, Henschen Doug, article *Big fast and if you want to understand the challenges of the Big Data era,...* (*Information Week*, The Business Value of Technology, August 9, 2010)

[21] http://www.catalinamarketing.com/company/

[22] Humby, Clive – dunnhumby Co-Founder and Chairman, *Introduction to « dunnhumby on loyalty »* White Paper (2010)

[23] Ellesworth, Richard R., *The Drucker Difference* (*Mc Graw Hill,* 2010) – Chapter 6, Corporate Purpose »

[24] Leahy, Terry, CEO of Tesco, (*The Independent* article, April 21, 2010)

[25] Dillon, David, Chairman and CEO of Kroger, (*Supermarket News article,* Dec 20, 2010)

[26] Hoyt, Chrys – *What's that sounds?* White Paper, page 28, Category segmentation. Consumer segments (Hoyt&Company, April 07), http://hubmagazine.com/archives/the_hub/2007/mar_apr/the_hub17_hoyt.pdf

BIBLIOGRAPHY

Bain & Company - Jean-Marie Pean and Cyrille Fabre, *Slower growth exposes GDD retailers' weakness* (June 10, 2009)

Benoun, Marcand Héliès-Hassid Marie-Louise, *Catégory management, mythes et réalités,* research publication.

Deloitte - Pat Conroy, Anuparm Narula, Siddharth Ramalingam, *A walk down the grocery aisle, Executive survey results exploring private label and national brands* (*Deloitte development LLC*, 2010)

Harris, Brian, *Making category management happen* article (*The Drug Store News*, Sept 18, 1995)

Newell, Frederick – *Why CRM doesn't work. How to win by letting Customers Manage the Relationship* (*Bloomberg Press,* 2003)

significantly different. Specifically, 49 percent of the workers are "single" compared to only 15 percent for the staff employees and 43 percent of the workers are "married" compared to 79 percent of the staff employees.

The ethnicity of the employee population is given in Panel E. The manufacturing plant is located in an area where the majority of the population is "white". Consequently, it is not surprising to see that about 92 percent of the staff and 67 percent of the workers are "white".

Finally, Panel F presents the general breakdown of the employees. In the case of the workers, the "shift" information shows that 44 percent, 23 percent and 33 percent of the workers are assigned to shift 1, 2 and 3, respectively. For the staff, Panel F presents the breakdown along job titles: nine of them are "managers", fourteen are "supervisors", 7 are "continuous improvement specialists", 7 are "others" and the reminding 7 had "missing" information.

Insert Table 1 Here

Table 2 shows the descriptive statistics and reliability tests for the employees in total and by groups. Except for *Autonomy* (Cronbach's Alpha is 0.62) and *Job Satisfaction* (Cronbach's alpha is 0.53), standardized Cronbach's alphas range from a low of 0.77 to a high of 0.91. Panel A provides the results for those constructs relevant to all employees: continuous improvement, job security, organizational support, job satisfaction and effort-reward fairness. Panel B provides the information on the questions (training and autonomy) that were responded to by the plant workers. Finally, Panel C provides the results from the questions answered by the staff employees. Those questions pertained to career satisfaction, supervisory support, and top management support of the continuous improvement strategy.

Insert Table 2 Here

Statistical Analysis

The analysis of variance results comparing plant workers' and staff employees' perception on various elements are given in Table 3. As discussed earlier, all other variables are measured using the 5-point Likert scale (where 5 is Strongly Agree), except for effort-reward fairness, which was measured on a 7-point Likert scale. The overall results from the employee's responses support the position that the employees feel positive about the commitment that the company made toward the adoption of the continuous improvement programs, with an average score of 4.4. The staff, however, viewed the commitment significantly higher than the workers (*p* = .002).

The top management support was likewise rated high with a mean score of 4.1, which is quite encouraging to management. Given the facts that lean production was implemented for less than 3 years and the company has just gone through a management buyout, the response for job satisfaction score was quite high (3.70) and as such, management was not overly concerned.[3] Finally, we also assessed "career satisfaction" of the staff employees. The mean score for this construct was 3.73, indicating that, on average; the staff "agrees to some extent" that they are making progress toward meeting their career goals and career advancements. Again, this is seen as a positive sign in light of the new owners and process adoption.

---

[3] A cautionary note is that job satisfaction has only 3 items and its Cronbach's alpha is 0.53.

JOB SECURITY, JOB SATISFACTION, EFFORT-REWARD EQUITY AND LEAN MANUFACTURING: A FIELD STUDY

Khim L. Sim, Ph.D.,CBE, Western Washington University,
Tel: (360) 650-6281, Khim.Sim@wwu.edu
Anthony P. Curatola, Ph. D., LeBow College of Business, Drexel University,
Tel: (215) 895-1453, Email:Curatola@Drexel.edu
John W. Rogers, Ph. D., American International College,
Tel: (413) 205-3378,Email: jrogers@acad.aic.edu

ABSTRACT

Focusing on a well respected manufacturing company located in the Eastern United States, this filed study examines some plausible reasons for a successful implementation of lean manufacturing. This is also a story of how the management and the employees working together to save the plant from being shut down. Results show that giving workers a voice does more than boosting their ego – it seems to improve their perception of increased job security. Likewise, quality training provides the necessary tools for the workers to remain productive which may have contributed to job satisfaction. Findings also show that while supervisory support enhanced career satisfaction and effort-reward fairness, it does not have an impact on "job security" for the staff employees. On the other hand, although top management support is not related to career satisfaction and effort-reward fairness, it does enhance the perception of increased job security for the staff employees. Finally, when management nurtures (measured as perceived organizational support), they provided the support which may have enhanced employees' perception of increased job security, more effort-reward fairness, and lastly job satisfaction. Thus, findings from this study reinforced the importance of nurturing (i.e., support), training and empowering which may have contributed to the successful implementation of lean manufacturing.

INTRODUCTION

The future of manufacturing in the United States and other advanced industrial countries depends on the ability to achieve dramatic improvements in productivity – output per employee – while continuously improving quality to meet rising customer expectations. In other words, survival in the competitive global economy requires successful implementation of lean production. Achieving this objective is challenging under any circumstances, but doing so in an organization whose employees and culture have been shaped by traditional work habits is doubly difficult. Overcoming resistance to change requires effective application of human relations and organizational development know-how along with skill in the technical aspects of operations management and process improvement. The track record of firms that have made this transition from a traditional to a new method of working can be instructive to any organization that seeks to embrace competition with the full engagement of its employees rather than retreat from it.

This study provides evidence on the employee's response to the implementation of a lean production system to a manufacturing plant.

Background

The firm in this field study began the lean journey from a long history as a well respected manufacturer that had created a pre-eminent consumer brand. Acquisition of the business – located in a highly unionized region of the Eastern United States – by a major multinational put the firm's strategy into bleak perspective. Unless the new management team brought in by the new owners could achieve dramatic improvements in both productivity and quality, the parent company would have no choice but to shut down the plant, relocating to a lower cost region and causing the loss of over 700 jobs in an area already suffering from the effects of extensive de-industrialization. This bleak choice energized the management team with the mission to "save the plant" and demonstrate that a high seniority, unionized workforce could be taught the techniques of lean production and motivated to get behind a comprehensive program of organizational change.

The initiative to transform the organization officially began in May 2005 with the kick off of a Total Employee Involvement meeting. After two short years, the entire team – employees and managers – received an Achievement Award, in recognition for their dedication to the continuous improvement through collaboration and mutual support. By the end of 2008, the firm was on its way to becoming a World Class "Best Practices" Manufacturer. Among the key metrics standing behind this achievement were:

- Customer returns decreased by 84%
- Production cycles reduced from 21 to 5 days
- Finished goods inventory reduced by $10,000,000
- Order lead times for custom products reduced by 44%
- Employee grievances reduced by 42%
- Accidents claims down 23.5% - with a $400,000 savings in Workers Compensation
- Reduction in greenhouse gas emissions and solid waste - with savings rising from $3,000,000 in 2006 to $9,300,000 in 2007.

Because of the above achievements, in less than 3 years, the quality manager was promoted to the plant manager. Being a certified six sigma expert, the new plant manager understood that lean is "all about PEOPLE being engaged and People are a company's greatest asset". In the quest for continuous improvement, the authors of this paper were invited to conduct a study to further understand the perception of the employees as well as areas of interest to the management. The president of the Union was informed and gave his support to this study. Before the survey was designed, the authors visited the manufacturing plant and met with the quality manager, the plant manager and the Vice President of manufacturing (hereafter, our research collaborators). During this meeting, our research collaborators (or the management team) provided a list of items which they are interested in learning about. These items became the basis of the study and are listed below:

1. Given that lean has been implemented about 3 years ago, where are they now in this journey? Is the company on the right track? What are the things that the company needs to focus on for continuing success?

2. Specifically, the management is very interested in knowing the employees' perception on job security and the perception of their future?
3. According to the plant manager, "Lean is all about people." Thus, the management is particularly interested in learning the perception change of the employees on job security, job satisfaction, career satisfaction, effort-reward fairness, organizational support, adequacy of training, supervisory support, and top management support of continuous improvement strategy.

LITERATURE REVIEW

Lean production or lean manufacturing is rooted in the Toyota Production System and aimed at reducing costs and improving quality simultaneously. For example, the International Motor Vehicle Project showed that, on average, cars built using lean techniques required one-third fewer hours and had one-third fewer defects than those produced from mass production (Womack, Daniel & Roos, 1990). Today, in order to remain competitive, many manufacturing companies have implemented lean manufacturing systems.

Although there are different definitions for lean production, it is often described as a relationship between the technical and the social organization of work. That is, the presence or the absence of any single technological or social system is insufficient to characterize it as a lean production (Babson, 1993). The technological system often includes such items like standardized work, visual control, planned maintenance and just-in-time inventory system. On the other hand, the social organization system has a direct impact on the quality of work life. Typically, it includes screening and selection in human resource (HR) practices, quality training, work teams, suggestions and ideas, employee discretionary authority to control the work, management support and management commitment. Because of its association with work life quality, and subsequently the motivational aspect as well as employees' job satisfaction, anecdotal evidence indicate that it is often the social organization system that dictates the success or the failure of the lean implementation in most organizations (Puvanasvaran, Megat, Tang & Razali, 2009; Lewchuck, Stewart & Yates, 2001).

Job Security and Lean Manufacturing

Although some companies have achieved success in adopting the lean production system, many were not able to sustain the potential benefits from its adoption (Rinehart, Huxley & Robertson, 1997; Womack et al., 1990; Womack and Jones, 1996; Bhasin & Burcher, 2006). The intention of introducing lean is often cited as reducing costs, improving efficiency, and increasing profitability. This tends to create suspicion among employees for the potential of job losses and outsourcing (Sim & Rogers, 2009; Bruno & Jordon, 1999). As a result, lean is often viewed as pro-company and not pro-employee. Research studies have also shown that employees often feel a sense of insecurity and perceive lean as 'redundancy threat' with adverse effect on morale, leading to worker unhappiness, withdrawal, and ultimately operational failure (Hines, Holweg & Rich, 2004; Bruno & Jordon, 1999; Bruno & Jordon, 2002; Sim & Rogers, 2009). For example, when Bacardi and Martini of the UK switched to a new work system, the management promised that no jobs would be lost as a result of the change process (Sung & Ashton, 2005). Similarly, Eastman Chemical, a Baldridge Award winner, promised that they will never lay off anyone because of quality improvement. The company attributed its success to recognizing that "people create quality" (Milliken, 1996). Likewise, in NUMMI, a facility cited as "spectacular success"

in lean production, management has been credited for offering job security in exchange for the risk associated with job restructuring when implementing lean manufacturing (Alder, 1995).

Management Commitment, Management Support and Lean Manufacturing
Critics claimed that lean manufacturing is associated with high stress (Lewchuck and Robertson, 1996; 1997; Lewchuck et al., 2001; Brenner, Fairris & Ruser, 2004). Yet, a large scale research study by Conti, Angelis, Cooper, Faragher & Gill (2006) suggested that high performance work practices such as lean manufacturing are not inherently stressful or pro-company. That is, findings based on the responses of 1,391 workers at 21 sites in four UK industries suggest that workers' stress levels are significantly related to management decisions in designing and operating the lean manufacturing systems. More succinctly, the authors of this study assert that stressful practices do not appear to be necessary conditions for attaining the benefits of a lean system. This means that management can mitigate workers' stress with better alignment of its day-to-day operations. Other case study, spanning across eight-years, the Mitsubishi's plant in Illinois, shows that deterioration occurs mainly due to unequal power sharing and management mistrust (Bruno & Jordon, 1999). Finally, Kumar (2000) asserts that although high-performance work systems (HPWS), such as lean manufacturing, have a positive effect on the economic and financial performance of an organization, the lack of supportive and committed organizational environment continues to be the most important barrier to a greater diffusion of such practices. Research studies also documented that employees who develop skills, competencies, and have autonomy over their jobs are more cooperative, creative, and effective in their problem solving efforts (Lawler, 1992; Tomer, 2001). In summary, social organization theory together with the above literature suggest that management plays an important role in the success or failure of the lean manufacturing system. Therefore, providing support, training and empowering workers, such as giving autonomy to them, are things that management can and must do to improve the quality of work life for the employees in an era when global competition is intense and the quest for continuous improvement is a never ending process. In doing so, management may help to mitigate the image of "lean" being equal to "job lost". Consequently, our hypotheses for job security in the null form are:

> H1a: Perceived organizational support, quality training, and having more autonomy are positively related to perceiving more job security for the plant workers.

> H1b: Perceived organizational support, supervisory support, and top management support in continuous improvement strategies are positively related to perceiving more job security for the non-unionized employees

Effort- Reward Fairness and Perceived Support
According to equity theory, employees evaluate the exchange relationships in organizations where they work in terms of a ratio between their efforts spent and rewards received. In employee-employer exchange relationships, employees are more likely to perceive under-reward unfairness than over-reward unfairness. Consequently, perceived inequity resulting from job effort-reward discrepancies provides an unpleasant emotional state, and cognitive dissonance, which requires the worker to reduce this tension. To diminish perceived inequity, a worker may actually alter efforts or may withdraw from the job resulting in substandard work behavior (Adams, 1965; Mowday, 1991). Therefore, actions by the organization or its agents (e.g.,

supervisors or top management) that are perceived as negative could lead to low levels of perceived support. Organizational support theory (OST), on the other hand, assumes that positive actions of the organization or its agents toward its members could lead to perceptions of being valued by the organization, thus having high levels of perceived support. Research studies have shown that perceived support affects job satisfaction, absenteeism, turnover, and retention (Eisenberger, Huntington, Hutchison, & Sowa, 1986; Eisenberger, Stinglhamber, Vandenberghe, Sucharski, & Rhoades, 2002; Randall, Cropanzano, Bormann, & Birjulin, 1999; Shore & Shore, 1995). In lean manufacturing, because of work pace intensification or the claim of "lean becomes mean", effort-spent *vis-a-vis* rewards-received may have a greater implication. Consequently, the following two hypotheses are tested to determine the association between perceived support and effort-reward fairness.

H2a: Perceived organizational support is negatively related to effort-reward fairness[1] for the plant workers.

H2b: Perceived organizational support, supervisory support, and top management support in continuous improvement strategies are negatively related to effort-reward fairness for the non-unionized employees.

Intrinsic Satisfaction - Job Satisfaction and Career Satisfaction

Much of the research on employee job-satisfaction shows that job satisfaction is a potential determinant of absenteeism, performance, turnover, and retention. Research studies show that job attitudes, to a greater extend, are within the ability of management to influence (Eisenberger, Cummings, Armeli, & Lynch, 1997; Griffeth, Hom, & Gaertner, 2000; Lance, 1991). In addition, Witt & Nye (1992) and Brockner & Adsit (1986) show that fairness is an important component of job satisfaction. These results imply management can play an important role in nurturing the employee's affective feeling by providing support (i.e., organizational, supervisory, and top management support), training, autonomy, and rewarding fairness to employees. Likewise, workers and professional staff may perceive job-career satisfaction differently. For example, workers on the assembly line may perceive family-friendly policies and improved facilities of greater value to them while professional staff may see personal development opportunities and career advancement as more important (see, Sparham & Sung, 2007, p. 8). Accordingly, job satisfaction is used to measure workers' intrinsic satisfaction with their work while career satisfaction, an internally defined career outcome, is used to assess the professional staff's intrinsic satisfaction with their current job. This leads us to the following two hypotheses:

H3a: Perceived organizational support, quality training, and having more autonomy are positively related job satisfaction; while effort-reward fairness is negatively related to job satisfaction for the plant workers.

H3b: Perceived organizational support, supervisory support, and top management support in continuous improvement strategies are positively related to career satisfaction; while effort-reward fairness is negatively related to career satisfaction for the non-unionized employees.

---

[1] This construct was reversed coded, which means a high score represents a higher level of "effort-reward inequity".

METHODOLOGY

Survey

A random sample of 35 percent of the shop floor employees was selected to participate in our survey. This level of participation was selected because strong support for the study was present from both management and the union. Our research collaborator distributed the questionnaire, as evenly as possible, to the unionized hourly workers across the three shifts, spanning across multiple production departments. The workers were asked to fill out the survey in the company's meeting room and return the survey the same day. The data collection process stretched across a three week period. A total of 151 surveys were received, which resulted in a 100 percent response rate. Due to incompleteness or failing to pass some "validity check", 15 surveys were dropped from the study, resulting in only 136 useable surveys.

The non-unionized employees (hereafter, staff), on the other hand, were given a self addressed envelope to return their completed survey. A total of 60 surveys were distributed to this group of which 48 surveys were completed and returned. This response translates into an 80 percent response rate or about 45 percent of the total staff population. The staff employees included office support personnel, engineering, administration, management, and continuous improvement specialists. All 48 surveys were useable.

Constructs Measures:

The effort-reward fairness was measured using a 7-point Likert scale (1= Totally Disagree, 7 = Totally Agree). All other constructs were measured using a 5-point Likert scale (1= Strongly Disagree, 5= Strongly Agree).   Appendix A provides detailed information related to the questionnaires.

There are three groups of questions in the questionnaire. The first group applies to all employees (workers and staff). The second group is applicable only to the plant workers and the third group is applicable only to staff personnel. As a result, the specifics for each set of questions is discuss independently for each group.

Constructs Applicable For All Employees:

Continuous Improvement Effort

To appreciate the employees' perception of the continuous improvement programs, four questions were used to assess the climate within this plant. It should be noted that "Continuous Improvement Effort" is not used as an independent variable or dependent variable in this study. Nevertheless, for successful implementation of a 'lean production' it is important to understand how receptive the employees are regarding the 'quest for continuous improvement'. Thus, the focus of the questions is, the company is working hard (1) toward the goal of total customer satisfaction, (2) to eliminate waste in the processes, (3) to meet and exceed expectations in product quality, and (4) to reduce product cost. These questions are derived from Womack et al., 1990. The Cronbach's alpha score for this construct is 0.79.

Perceived Job Security

Perceived Job Security is considered a critical issue to all employees. We use a 4-item scale adapted from Sim & Roger (2009) for this construct. One item, however, was replaced with the

following item: "Overall, my future in this company appears to be more promising compared to 2-3 years ago". This item was included to provide relevant information to management about the overall success of the system from the employee's point of view. The Cronbach's alpha for this construct is 0.80.

Effort-Reward Fairness

Perceived Effort-Reward Fairness is adapted from Van Yperen (1996). The 6-item scale has a high internal consistency and has been used in numerous scholarly research papers (see, Janssen, 2000, 2001; Van Yperen, 1998). All 6-item were reversed coded, which means a high score represents a higher level of "effort-reward inequity". Cronbach's alpha for this construct is 0.88.

Perceived Organizational Support

The final construct applying to all participants is the perceived organizational support (POS). This construct is assessed by means of a modified scale from Eisenberger et al., 1986. The original 1986's scale has 16 items while the shorter 1997's scale has eight items. The shorter 1997 scale was modified for our study by deleting the item: "If given the opportunity, the company would take advantage of me". This item was deleted because we felt that this statement is too negative. The Cronbach's alpha for POS is 0.84.

Constructs Applicable For Plant Workers Only
Training

Training is seen by management as one of the critical element to lean and people. Hence, we measured it using a four item scale from Sim & Rogers (2009). Since the scale is relatively new, it was tested in another pilot study where 83 employees filled out the questionnaire. Cronbach's alpha for the 4 items in the pilot study was 0.72; no changes were made. With respect to this construct, the Cronbach's alpha is 0.78.

Autonomy

The relationship between autonomy and job security are of particular importance in this field study. To measure the employee empowerment construct, we selected the 4 item of Power (1995). The four items measure employees' involvement and employees' suggestions. We modified Power's scale by replacing the item which reads "A more active employee suggestion system" with two items for a total of 5 items. The two items have been tested in a pilot study using 83 plant workers and result in no changes. The newly added items are "The Company does not value my idea/suggestions relative to continuous improvement"[2] and "The Company allows me through programs or forums to express my ideas and opinion about continuous improvement." The Cronbach's alpha on the five items is 0.53, which means or suggests that the internal consistency of the scale is very low. Thus, items 1 and 2 from the Power (1995) were dropped. As a result of dropping the items, the remaining 3 items' scale has a Cronbach's alpha of 0.62.

Job Satisfaction

Job satisfaction as a result of the work philosophy change is another important element. To determine the resulting job satisfaction, a three item scale, is adapted from Hackman & Oldham

---

[2] This item is reversed coded.

(1975). Hackman and Oldham (1975) originally reported an internal consistency reliability of 0.76. Our internal consistency reliability, however, is low, with Cronbach's alpha of 0.53.

Constructs Applicable For Staff Employees
Career Satisfaction
There are three items of particular interest with respect to staff employees. The first item is career satisfaction. For this construct, we selected a five item construct from Greenhaus, Parasuraman, & Wormley (1990). The resulting Cronbach's alpha for this construct is 0.86.

Supervisory Support
Another item of concern for management is the perception by the staff employees of supervisory support. To test for this element, we selected nine item scale from Greenhaus et al., (1990). For supervisory support, the Cronbach's alpha is 0.91.

Top Management Support of Continuous Improvement Strategy
The final item of interest for this subpopulation of employees is top management support of continuous improvement strategy. For this measure, we adopted three items from Fullerton, McWatters, & Fawson (2003). The Cronbach's alpha for this element is 0.77.

RESULTS
The results for this study are presented in two parts. The first part presents the descriptive statistics based on the demographic information for the participants. The second part presents the statistical analysis based on the participant's responses.

Descriptive Statistics
The general demographics for the participants are given in Table 1. Panel A of the Table compares year of services for the plant workers and staff employees. As shown in this Panel, about 35 percent of the workers and 21 percent of staff have worked in the plant for 5 years or less. In addition, about 45 percent of the workers and 53 percent of the staff had been employed by the Company for 6 to 20 years, with the remaining 18 percent of the workers and 27 percent of the staff having been employed for more than 20 years.

Two observations can be made from this information. First, it is interesting to note that the percentage of workers is relatively consistent across years of service, except for the category of 21 to 30 where there is only 2 employees (or 1.5 percent). Second, the turnover rate for the plant workers is slightly higher than the staff for the newest hires (i.e., 5 years or less). This tends to be a normal trend for most manufacturing companies (see, Sparham and Sung, 2007, p. 8).

Panel B shows the educational level of the plant workers and the staff. As shown in the Panel, the majority of the workers (78%) have a high school degree; while the other workers have a college degree. In contrast, the majority of the staff (nearly 69 percent) has a college degree.

Panel C presents the breakdown of the employees along gender lines. Seventy nine percent of the workers are male with the remaining 21 percent are female. Somewhat surprising, these same percentages exist for the staff employee.

Panel D provides marital status of the employees. Unlike the demographic found for gender, where the percentages were nearly the same for both employee groups, the marital status is

On a less positive note, the perceived job security and perceived organizational support have mean scores of 3.1 and 3.3, respectively. In addition, the plant workers provided significantly lower rating for both constructs (i.e., 2.95 and 3.12, respectively, with $p = 0.000$). Before the study was conducted, the plant manager had told us that "The workers are the life blood of this company. I told them that we keep our job only when they are able to keep theirs." Since this company has just gone through a management buyout and a mission to "save the plant", perhaps, it is not surprising to see a sense of "insecurity" among the employees, more so for the plant workers. Nevertheless, the lower score in perceived organizational support suggests that the management may want to examine or understand plausible reasons for the lower rating. Finally, effort-reward fairness (assessed on a 7-point Likert scale, with 4 being neutral, and reversed coded) has a mean score of 4.0 with no significant differences between hourly workers and salary employees.

<center>Insert Table 3 Here</center>

The correlation matrix for the constructs within the workers (Panel 4A) and the staff (Panel 4B) are given in Table 4. Panel 4A is limited to the constructs applicable to the workers. As shown in that Panel, perceived organizational support, training, autonomy, job security, effort-reward fairness, and job satisfaction are correlated at the 1 percent level, which is consistent with the fact that these are common aspects of lean manufacturing. Tenure is correlated with gender while gender is correlated only with job satisfaction.

Panel 4B shows the correlation matrix for the constructs pertaining to the staff's responses. Naturally, the data set is much smaller, with $n = 48$ (in contrast to 136 workers). Job security, job satisfaction, and career satisfaction are not correlated. Interestingly, the low correlation ($r = 0.16$) between job satisfaction and career satisfaction suggests that job satisfaction is different from career satisfaction. Finally, while top management support is positively related to POS ($r = 0.42$), job security ($r = 0.53$) and negatively related to effort reward fairness ($r = -.34$), it is not correlated with job satisfaction ($r = .13$). The overall results suggest that top management support in continuous improvement strategy can have an impact on organizational policies but it may have little influence on job satisfaction. Perhaps, one of the more interesting results relates to "gender". For example, while female workers reported a higher job satisfaction ($r = 0.18$), the male staff reported a higher perceived job security ($r = 0.17$) and a lower effort-reward inequity ($r = 0.22$) than their female colleagues. These differences may be driven by gender and its association with "blue collar" vis-à-vis "white collar" jobs.[4]

<center>Insert Table 4 Here</center>

Tables 5 and 6 report the regression results of Job Security, Job Satisfaction, and Effort-Reward Fairness, for the workers and staff, respectively. As shown in Panel 5a, perceived organizational support ($p = 0.001$) and having more autonomy ($p = 0.009$) are significant and positively related to feeling more job security for the plant workers. In contrast, organizational support ($p = 0.036$) and top management support ($p = 0.003$) are significant and as such, provide a sense of job

---

[4] Research studies, however, in the past few decades have produced mixed results. That is, despite differential opportunities and pay inequity, relationships between gender and job satisfaction have been inconsistent (see, Harris & Earle, 1986; Witt, 1988).

security for staff employees. As a result of these findings, hypotheses 1A and 1B are partially supported.

As expected, Perceived Organizational Support ($p = 0.000$) is negatively related to effort-reward fairness for the plant workers; while year of services and gender are only marginally significant ($p = 0.08$ and p = 0.09). This result suggests that the female workers seem to feel less effort-reward inequity compared to the male workers (only marginally significant). Results for the staff employees are reported in Table 6, Panel 6B. Both supervisory support ($p = 0.009$) and Perceived Organizational Support ($p = 0.04$) are negatively related to Effort-Reward Fairness for the staff employees. Unlike the perceptions of the plant workers, male staff ($p = 0.08$) seems to feel less effort-reward inequity compared to their female colleagues. As a result, hypotheses 2A and 2B also are partially supported.

Finally, Panel 5C and 6C show the results of job satisfaction (for the workers only) and career satisfaction (for the staff only), respectively. Organizational support ($p = 0.000$) and training ($p = 0.023$) are positively related to job satisfaction; while Effort-Reward Fairness ($p = 0.005$) is negatively related to job satisfaction for the workers. As explained earlier, motivated employees generally care about meeting their career goals and career advancements. Consequently, career satisfaction was assessed. Interestingly, supervisory support is the only factor which explains career satisfaction. This means that top management support, organizational support and effort-reward fairness are not statistically significant. As a result, hypotheses 3A and 3B are partially supported.

Insert Tables 5 and 6 Here

Discussion and Conclusion
Focusing on the social organization theories, results from this field study show that perceived organization support is statistically significant in 5 out of the 6 hypotheses tested. Consequently, when management nurtures, they are providing the support which may have enhanced the workers' feeling of more job security, effort-reward fairness and lastly job satisfaction. Results also show that giving workers a voice does more than boost ego – it seems to improve their perception of increased job security. Likewise, quality training provides the necessary tools for the workers to remain productive which may be contributed to job satisfaction. Results show that while supervisory support enhanced career satisfaction and effort-reward fairness, it does not have an impact on "job security" for the staff employees. On the other hand, although top management support is not related to career satisfaction and effort-reward fairness, it does enhance the perception of increased job security for the staff employees. Thus, findings from this study reinforce the importance of nurturing, training and empowering which may have contributed to the successful implementation of lean manufacturing.

Lack of trust and resistance to change are often reasons cited for implementation failure in high performance work systems, or lean production. In a study of Mitsubishi plant in Illinois, Bruno & Jordan (1999) asserted that "… in lean production facilities, workers must be able to trust not only management's intentions, but also the integrity of the workplace transformation process." Other studies have claimed that the heavy emphasis on downsizing and organizational restructuring has created unsecured and stressful environments for their employees. That is, rather than empowering the workers, the new work systems have led to a loss of control and autonomy and have placed a wide range of increased demands on workers (Kumar, 2000). On

the other hand, Eastman Chemical Company, a winner of Malcolm Baldrige quality award, has attributed their success to "the Eastman way" – a culture where quality is very much a people issue. Thus, honesty, fairness, trust, employee wellbeing, citizenship and a winning attitude have helped them to maintain the competitiveness in the market place (Milliken 1996).

The principles of organizational change employed by the subject firm represent a creative mixture of common sense management and technical re-engineering. Everything begins with listening to people, convincing them that their concerns will be acknowledged, and asking for their help in solving operational problems. Leadership in this sense is about listening to and empowering people. But it is also about bringing into play the latest techniques of process improvement. As noted earlier, the subject firm has enjoyed many successes including the overarching achievement of saving the plant and preserving jobs. Nevertheless, challenges lie ahead for the management to sustain this rate of improvement. In an era with increasing global competition, the threat for job lost remains a concern to most hourly employees, as evidenced in the survey. Other areas, where the ratings are somewhat lower, may need additional attention. These include workers' autonomy and organizational support. Otherwise, it appears that the achievements of lean transformation have positioned the firm to meet future challenges.

The broader lesson of this study is that saving the plant and preserving jobs was achieved, not through special concessions but by becoming a competitor on the stage of the global economy and a supplier of choice for existing and new customers. The transformation wrought by lean has made an established industrial firm with a unionized labor force competitive on the world stage and has repositioned it for the future.

<div align="center">

Appendix A

</div>

Continuous Improvement Effort
Which of the following describes your feelings towards the continuous improvement programs?
1. The company is working hard towards the goal of total customer satisfaction.
2. The company is working hard to eliminate waste in processes.
3. The company is working hard to meet and exceed expectation in product quality.
4. The company is working hard to reduce product cost.

Perceived Job Security
Which of the following describes your feelings towards job security?
1. Continuous improvement initiatives have increased our job security.
2. Utilizing continuous improvement tools, the company will focus on keeping local jobs.
3. The company will try its best to reduce and/or eliminate layoffs.
4. Overall, my future in this company appears to be more promising compared to 2-3 years ago.

Perceived Organizational Support
Indicate to the extent to which you agree or disagree with the following statements:
1. Help is available from the company when I have a problem.
2. The company is willing to extend itself in order to help me perform my job to the best of   my ability.
3. Even if I did the best job possible, the company would fail to notice. (Reversed Coding)
4. The company takes pride in my accomplishments at work.
5. The company really cares about my well-being.
6. The company cares about my general satisfaction at work.
7. The company show very little concern for me. (Reversed Coding)

Effort-Reward Fairness
Indicate to the extent to which you agree or disagree with the following statements (Totally Disagree=1, Neutral=4, Totally Agree=7):
1. I work too hard considering my outcome.
2. I give a great deal of time and attention to the organization, but do not feel appreciated.
3. I invest more in my job than I receive in return.
4. The rewards I receive are not proportional to my investments.
5. I put more energy into my job than it is worth.
6. I feel unfairly treated in my job.

Quality Training
Indicate to the extent to which you agree or disagree with the following statements:
1. People leading improvement initiatives have the proper amount of training to effectively produce desired results.
2. My knowledge of continuous improvement allows me to apply them at work.
3. The company provides me adequate training to be productive during improvement events.
4. Continuous Improvement training is provided in a clear concise manner with many   practical examples on how to best use the tools.

Autonomy

Indicate to the extent to which you agree or disagree with the following statements:
1. Over the years, interactions between employees with customers and suppliers have increased. (Dropped)
2. I am able to act independently of my supervisor in performing my job function. (Dropped)
3. The company encourages employees to involve in design, planning, and problems solving.
4. The company does not value my ideas/suggestions relative to Continuous Improvement. (Reversed Coding)
5. The company allows me through programs or forums to express my ideas and opinion about continuous improvement

Job Satisfaction

Indicate to the extent to which you agree or disagree with the following statements:
1. Generally speaking, I am very satisfied with this job.
2. I frequently think of quitting this job. (Reversed Coding)
3. I am generally satisfied with the kind of work I do in this job.

Career Satisfaction

Indicate to the extent to which you agree or disagree with the following statements:
1. I am satisfied with the success I have achieved in my career.
2. I am satisfied with the progress I have made towards meeting my overall career goals.
3. I am satisfied with the progress I have made towards meeting my goals for my income.
4. I am satisfied with the progress I have made towards meeting my goals for advancement.
5. I am satisfied with the progress I have made towards meeting my goals for development of new skills.

Perceived Supervisory Support

Indicate to the extent to which you agree or disagree with the following statements:
1. My supervisor takes the time to learn about my career goals and aspirations.
2. My supervisor cares about whether or not I achieve my career goals.
3. My supervisor keeps me informed about different career opportunities for me in the organization.
4. My supervisor makes sure I get the credit when I accomplish something substantial on the job.
5. My supervisor gives me helpful feedback about my performance.
6. My supervisor gives me helpful advice about improving my performance when I need it.
7. My supervisor supports my attempts to acquire additional training or education to further my career.
8. My supervisor provides assignment that gives me the opportunity to develop and strengthen new skills.
9. My supervisor assigns me special projects that increase my visibility in the organization.

<u>Top Management Support of Continuous Improvement Strategy</u>
Which of the following best describes your feelings towards top management support
   (Indifferent=1, Moderately Support=3, Very Supportive=5)?
1. In initiating change programs.
2. In implementing lean manufacturing practices.
3. In providing training for new production strategies

References

Adams, J. S. (1963). Toward an understanding of equity. *Journal of Abnormal and Social Psychology, 67,* 422-436.

Adler, P. (1995). ''Democratic Taylorism'': The Toyota Production System at NUMMI. In S. Babson (Ed.), Lean work: Empowerment and exploitation in the global auto industry (pp. 207-219). Detroit, MI: Wayne State University Press.

Babson, Steve. (1993). Lean or Mean: The MIT Model and Lean Production at Mazda. *Labor Studies Journal*, Vol. 18. No. 2. (Summer), 3-24.

Bhasin, S., & Burcher, P. (2006). Lean Viewed as a Philosophy. *International Journal of Manufacturing Technology Management*, 17(1):56 -72.

Brenner, M., Fairris, D., & Ruser, J. (2004). Flexible work practices and occupational safety and health: Exploring the relationship between cumulative trauma disorders and workplace transformation. *Industrial Relations*, Vol. 43 pp. 242-66.

Brockner, J., & Adsit, L. (1986). The moderating effect of sex on the equity-satisfaction relationship. *Journal of Applied Psychology, 71,* 585-590.

Bruno, R., & Jordan, L. (1999). From High Hopes to Disillusionment: The Evolution of Worker Attitudes at Mitsubishi Motors, in David Lewin and Bruce E. Kuafman (Eds.), *Advances in Industrial and Labor Relations*, 9, 153-82. Stanford, Connecticut: JAI Press

_____. (2002) "Lean production and the politics of dissent: radicalising the shop floor at Mitsubishi motors." Working USA 6:108-134

Conti, R., Angelis, J., Cooper, C., Faragher, B., & Gill, C. (2006). The effects of lean production on worker job stress. *International Journal of Operations & Production Management*, Vol. 26 No. 9, 2006, pp. 1013-1038.

Eisenberger, R., Cummings, J., Armeli, S., & Lynch, P. (1997). Perceived organizational support, discretionary treatment, and job satisfaction. *Journal of Applied Psychology, 82,* 812–820.

Eisenberger, R., Huntington, R., Hutchison, S., & Sowa, D. (1986). Perceived organizational support. *Journal of Applied Psychology, 71,* 500–507.

Eisenberger, R., Stinglhamber, F., Vandenberghe, C., Sucharski, I., & Rhoades, L. (2002). Perceived Supervisor Support: Contributions to perceived organizational support and employee retention. *Journal of Applied Psychology,* Vol. 87, No. 3, 565–573.

Fullerton, R., McWatters, C., & Fawson, C. (2003). An examination of the relationships between JIT and financial performance. *Journal of Operations Management*, Vol. 21 pp.383-404.

Greenhaus, J. H., Parasuraman, S., & Wormley, W. M. (1990). Effects of race on organizational experience, job performance evaluation, and career outcomes. *Academy of Management Journal*, vol. 33 No. 1, pp. 64-86.

Griffeth, R. W., Hom, P. W., & Gaertner, S. 2000. A meta-analysis of antecedents and correlates of employee turnover: Update moderator tests, and research implications for the next millennium. *Journal of Management*, 26: 463–488.

Hackman, J.R., & Oldham, G.R. 1975. Development of the job diagnostic survey. *Journal of Applied Psychology*, Vol. 60, pp. 159-70.

Harris, C. T., & Earle, J. R. (1986). Gender and work values: Survey findings from a working-class sample. *Sex Roles,* 15*,* 487-494.

Hines, P., Holweg, M., & Rich, N. (2004). Learning to evolve: A review of contemporary lean thinking, *International Journal of Operations & Production Management*, Vol. 24:10, 994 -1011

Janssen, O. (2000). Job demands, perceptions of effort–reward fairness, and innovative work behavior. *Journal of Occupational and Organizational Psychology*, 73, 287–302.

Janssen, O. (2001). Fairness perceptions as a moderator in the curvilinear relationships between job demands, and job performance and job satisfaction. *Academy of Management Journal*, 44, 1039–1050.

Kumar, P. (2000). *Rethinking High-Performance Work Systems.* Industrial Relations Centre Queen's University Kingston, Ontario, Canada K7L 3N6.

Lance, C. E. (1991). Evaluation of a structural model relating job satisfaction, organizational commitment, and precursors to voluntary turnover. *Multivariate Behavioral Research*, 26: 137–162.

Lawler, E. E. III. (1992). *The Ultimate Advantage: Creating high involvement organization*. San Francisco, Ca: Jossey-Bass.

Lewchuck, W., Stewart, P., Yates, C. (2001), Quality of working life in the automobile industry: a Canada – UK comparative study. *New Technology, Work, and Employment*, Vol. 16 No.2, 72-87.

Lewchuck, W., & Robertson, D. (1996). Working Conditions Under Lean Production: A Worker-based Benchmarking Study. *Asia Pacific Business Review*, Vol. 2, Summer, 60-81.

_____ (1997). Production without Empowerment: Work reorganization from the Perspective of Motor Vehicle Workers *Capital and Class*. Vol. 63 (Autumn), pp. 37-65.

Milliken, W. F. (1996). The EastmanWay. *Quality Progress*, October, 57-62.

Mowday, R. T. 1991. Equity theory perceptions of behavior in organizations. In R. M. Steers & L. W. Porter (Eds*.), Motivation and work behavior* (pp. 111-131). New York: McGraw-Hill.

Powell, T. (1995). Total quality management as competitive advantage: a review and empirical study. *Strategic Management Journal*, Vol. 16, 15-37.

Puvanasvaran, P., Megat, M., Tang, S. H., & Razali, M. M. (2009). The roles of communication process for an effective lean manufacturing implementation. *Journal of Industrial Engineering and Management*, 2(1), 128-152.

Randall, M. L., Cropanzano, R., Bormann, C. A., & Birjulin, A. (1999). Organizational politics and organizational support as predictors of work attitudes, job performance, and organizational citizenship behavior. *Journal of Organizational Behavior, 20,* 159–174.

Rinehart, J., Huxley, C., & Robertson D. (1997). *Just Another Car Factory? Lean Production and Its Discontents*, ILR Press, Ithaca, NY,

Shore, L. M., & Shore, T. H. (1995). Perceived organizational support and organizational justice. In R. Cropanzano & K. M. Kacmar (Eds.), *Organizational politics, justice, and support: Managing social climate at work* (pp. 149-164). Westport, CT Quorum Press.

Sim, K. L., & Rogers, J. (2009). Implementing lean production systems: barriers to change, *Management Research News,* Vol. 32, 1, 37-49.

Sparham, E., & Sung, J. (2007). High Performance Work Practices: Work Intensification or 'win-win'? Center for Labor Market Studies, University of Leicester, CLMS Working Paper No. 50.

Sung, J., & Ashton, D. (2005). *High Performance Work Practices: Linking Strategy and Skills to Performance Outcomes.* London: DTI.

Tomer, J. (2001). Understanding high-performance work systems: the joint contribution of economics and human resource management. *Journal of Socio-Economics*, Vol. 30, 1, 63-73.

Van Yperen, N. W. (1996). Communal orientation and the burnout syndrome among nurses: A replication and extension. *Journal of Applied Social Psychology*, 26: 338-354.

Van Yperen, N. W. (1998). Informational support, equity and burnout: The moderating effect of self-efficacy. *Journal of Occupational and Organizational Psychology*, 71: 29-33.

Witt, L. A., & Nye, L. (1992). Gender and the relationship between perceived fairness of pay or promotion and job satisfaction. *Journal of Applied Psychology*, Vol. 77, No. 6, 910-917.

Witt, L. A. (1988). Breadwinner vs. non-breadwinner differences in married women's job satisfaction and perceptions of organizational climate. *Human Relations, 41,* 483-491.

Womack, James P., Daniel, T. J., & Roos, D. (1990). *The Machine that Changed the World*. New York: Harper Perennial.

Womack, J., & Jones, D. T. (1996). *Lean Thinking: Banish Waste and Create Wealth in Your Corporation,* Simon and Schuster, London.

Table 1- Demographic Information of Workers and Staff

Panel 1A -Years of Service

Years of Service	Workers		Non-Unionized Staff	
	# of Respondents	Percent	# of Respondents	Percent
0-5 yrs	47	34.6%	10	20.8%
6-10 yrs	33	24.3%	12	25.0%
11-20 yrs	29	21.3%	13	27.1%
21-30 yrs	2	1.5%	6	12.5%
>30 yrs	23	16.9%	7	14.6%
Missing	2	1.4%	0	0%
Total	136	100%	48	100%

Panel 1B - Educational level

Education	Workers		Non-Unionized Staff	
	# of Respondents	Percent	# of Respondents	Percent
High School	106	77.9%	15	31.2%
Associate Degree	22	16.2%	12	25.0%
Bachelor Degree	5	3.7%	15	31.2%
Master Degree	1	0.7%	6	12.5%
Missing	2	1.5%	0	0%
Total	136	100%	48	100%

Panel 1C - Gender

Gender	Workers		Non-Unionized Staff	
	# of Respondents	Percent	# of Respondents	Percent
Male	108	79.4%	38	79.2%
Female	28	20.6%	10	20.8%
Total	136	100%	48	100%

Panel 1D - Marital Status

Marital Status	Workers		Non-Unionized Staff	
	# of Respondents	Percent	# of Respondents	Percent
Single	66	48.5%	7	14.6%
Married	58	42.6%	38	79.2%
Other	11	8.1%	2	4.2%
Missing	1	0.8%	1	2%
Total	136	100%	48	100%

Panel 1E - Ethnicity

Ethnicity	Workers		Non-Unionized Staff	
	# of Respondents	Percent	# of Respondents	Percent
African American	10	7.4%	1	2.1%
Asian	5	3.7%	1	2.1%
Hispanic	19	14.0%	0	0.0%
White	91	66.9%	44	91.7%
Others	11	8.1%	1	2.1%
Missing			1	2.0%
Total	136	100%	48	100%

Panel 1F

Shift	Workers		Job Title	Non-Unionized Staff	
	# of Respondents	Percent		# of Respondents	Percent
No. 1	58	42.6%	Director	3	6.2%
No. 2	31	22.8%	Manager	9	18.8%
No. 3	44	32.4%	Supervisor	14	29.2%
			CI Specialist	7	14.6%
			Others	7	14.6%
			Missing	8	16.6%
Total	136	100%		48	100%

Table 2: Descriptive Statistics

Panel 2A: All Employees

	N	Min	Max	Mean	Std. Dev	Cronbach's Alpha
Job Security	186	1.00	5.00	3.13	1.021	0.80
Perceived Org Support	186	1.00	5.00	3.25	.8263	0.84
Job Satisfaction	185	1.67	5.00	3.83	.8099	0.53
Continuous Improvement	186	1.75	5.00	4.40	.6682	0.79
Effort- Reward Fairness	182	1.00	7.00	4.00	1.470	0.88

Panel 2B: Workers Only

	N	Minimum	Maximum	Mean	Std. Deviation	Cronbach's Alpha
Training	135	1.00	5.00	3.4688	.8026	0.78
Autonomy	135	1.00	5.00	2.9343	.9446	0.62

Panel 2C: Staff Only

	N	Min	Max	Mean	Std Dev	Cronbach's Alpha
Career Satisfaction	48	2.00	4.80	3.7292	.76240	0.86
Supervisory Support	48	1.33	4.89	3.4838	.83132	0.91
Top Mgmt Support	48	1.00	5.00	4.1389	.83887	0.77

Table 3: *ANOVA* by Job Type

		N	Mean	Std. Dev	Std. Error	F	Sig
Job Security	Workers	135	2.9500	.96959	.08345	19.720	.000
	Staff	48	3.6476	.82733	.11941		
	Total	183	3.1330	.98163	.07256		
Perceived Org Support	Workers	136	3.1171	.82897	.07108	16.580	.000***
	Staff	48	3.6607	.68899	.09945		
	Total	184	3.2589	.82834	.06107		
Job Satisfaction	Workers	135	3.8296	.81839	.07044	.111	.740
	Staff	48	3.8750	.79188	.11430		
	Total	183	3.8415	.80960	.05985		
Continuous Improvement	Workers	136	4.3199	.71860	.06162	10.310	.002**
	Staff	48	4.6719	.40984	.05916		
	Total	184	4.4117	.66940	.04935		
Effort-Reward Fairness	Workers	133	3.9898	1.51918	.13173	.067	.796
	Staff	48	4.0542	1.35505	.19558		
	Total	181	4.0069	1.47402	.10956		

** sig at .01
*** sig at .001

Panel A: Workers

	Tenure	Gender	POS	Training	Autonomy	Job Security	Effort-Reward Fairness
Gender	.183*						
POS	.091	-.004					
Training	-.094	.141	.47**				
Autonomy	.024	.038	.583**	.608**			
Job Security	.084	.145	.510**	.412**	.494**		
Eff. Reward Fairness	.053	-.09	-.581**	-.356**	.-423**	-.324**	
Job Satisfaction	.053	.182*	.565**	.428**	.351**	.378**	-.498**

 N = 136; * significant at 5%; ** significant at 1%;

Panel B: Staff

	Tenure	Gender	POS	Supervisory Support	Top Mgmt Support	Job Security	Effort-Reward Fairness	Job Satisfaction
Gender	-.215							
POS	-.151	-.078						
Supervisory support	-.072	.176	.485**					
Top Mgmt Support	-.051	-.127	.420**	.249				
Job Security	-.21	-.17	.532**	.361*	.542**			
E_Reward Fairness	-.06	.224	-.49**	-.465**	-.344*	-469**		
Job Satisfaction	.08	-.07	.415**	.257	.133	.262	-.277*	
Career Satisfaction	.045	-.061	.280	.403**	.202	.056	-.198	.166

N= 48; * significant at 5%; ** significant at 1%

Table 5: Results for the Plant Workers

		Panel 5A Job Security				Panel 5B Effort -Reward Fairness [a]				Panel 5C Job Satisfaction			
Variable	Pred Sign	Coeff	Std Err	*t*-stat	*p*-value	Coeff	Std Err	*t* stats	*p*-value	Coeff	Std Err	t-stats	*p*-value
Constant		.265	391	.677	.500	7.554	.536	14.09	.000***	6.928	1.44	4.80	.000***
Tenure	?	.028	.051	.541	.589	.138	.078	1.766	.08#	.046	.127	.360	.720
Gender	?	.255	.179	1.42	.158	-0.468	.276	-1.589	.09#	.710	.447	1.59	.114
POS	+	.376	.106	3.56	.001***	-1.069	.13	-8.240	.000***	1.183	.288	4.11	.000***
Quality Training	+	.114	.111	1.02	.306	ns	ns	ns	ns	.596	.263	2.27	.023*
Autonomy	+	.250	.103	2.42	.009**	ns	ns	ns	ns	-.224	.227	-.989	.324
E_Reward Fairness	-	ns	ns	ns	ns	ns	ns	ns	ns	-.372	.141	-2.64	.005***

Adj. *R*-sq=.31 F=12.90 Sig=.000***	Adj. *R*-sq=.34 F=14.3 Sig=.000***	Adj. *R*-sq=.39 F=14.7 Sig=.000***
n=133	n=133	n=133

significant at 10%; * significant at 5%; ** significant at 1%; *** significant at < 1%

[a] Effort-Reward Fairness was reversed coded. This means a higher score represents "a higher inequity", or predicted signs of '-ve' with the independent variables.

Table 6: Results for Staff Employees

Panel 6A
Job Security

Variable	Pred Sign	Coeff	Std Err	t-stat	p-value
Constant		1.184	.765	1.547	.129
Tenure	?	-0.111	.074	-1.508	.139
Gender	?	-0.345	.245	-1.408	.166
POS	+	0.314	.170	1.843	.036*
Supervisor Supp	+	0.161	.134	1.205	.235
Top Mgmt Supp	+	0.356	.124	2.866	.003**
E_Reward Fairness	ns	ns	ns	ns	ns

Adj R-sq=.396   F= 7.16   Sig=.000**
n=48

Panel 6C
Career Satisfaction [a]

Variable	Pred Sign	Coeff	Std Err	t stats	p-value
Constant		1.788	1.063	1.682	.100
Tenure	ns	ns	ns	ns	ns
Gender	ns	ns	ns	ns	ns
POS	+	0.099	.194	.511	.612
Supervisor Supp	+	0.327	.152	2.156	.018**
Top Mgmt Supp	+	0.082	.141	.585	.561
E_Reward Fairness	-	-0.025	.097	-.258	.798

Adj R-sq=.102   F= 2.33   Sig=.070 #
n=48

Panel 6B
Effort -Reward Fairness

Variable	Pred Sign	Coeff	Std Err	t-stat	p-value
Constant		7.806	1.304	5.985	.000***
Tenure	?	-.080	.125	-.645	.514
Gender	?	.755	.418	1.806	.08#
POS	-	-.504	.290	-1.737	.045*
Supervisor Supp	-	-.586	.228	-2.491	.009***
Top Mgmt Supp	-	-.191	.212	- .901	.373

Adj R-sq=.32   F=5.42   Sig=.001***
n=48

significant at 10%; * significant at 5%; ** significant at 1%; *** significant at < 1%

[a] Due to a small sample size, when tenure and gender were included, the model was significant at 0.14 and the only significant variable was supervisory support. Thus, both tenure and gender were excluded and results reported above. Using the same explanatory variables, job satisfaction was also analyzed; organizational support is the only significant variable ($p=0.01$).

LEADERSHIP SUCCESS, SPIRITUALITY, AND CONSCIENTIOUSNESS A CROSS-CULTURAL STUDY OF THE UNITED STATES AND FRANCE: AN EMPIRICAL STUDY

Nicole Jean Christian, MPA
Dowling College
150 Idle Hour Blvd.
Oakdale, NY 11769
(631) 879-1588 phone
Email: christian@dowling.edu , nicolejchristian@gmail.com

Yvette Essounga-Njan, Ph.D
Fayetteville State University
1200 Murchison Road, SBE 332, Fayetteville, NC 28301
(910) 672-1595 (phone) – (910) 672-2046 (fax)
Email: yessoung@uncfsu.edu

Maxine Morgan-Thomas, JD, CPA
Long Island University - Brooklyn Campus
1 University Plaza, H700, Brooklyn, NY 11201
718 246-6464 (phone) - 718 488-1125 (fax)
Maxine.morgan-thomas@liu.edu

ABSTRACT

This paper is a follow-up discussion of the role of leadership style on organizational and economic performance, especially in times of economic upheaval the kind the Unites States and just about every country in the world has been going through. The discussion proposes leadership style success depends on behavior on the leader's part of a nature to convince followers he or she is working for the long term well-being of followers. This should fuel economic recovery, helping to quash the current enduring economic slumber. Data collected from the United States and France will be used to test hypotheses offered, testing the construct of spirituality, a paradigm at the core of leader's conscientiousness. The spirituality paradigm in this study will be operationalized through four dimensions: accountability, leader's concern for followers, rationality, and transparency.

Keywords: Leadership success, Spirituality, Conscientiousness, Economic Performance.

INTRODUCTION

Everywhere one turns, signs of difficult economic times abound in the United States and all over the world. Such circumstances call for a renewed interest not only in leadership and its impact [45] on organizational and economic outcomes [34], but also on characteristics most likely to make the leadership style exercised effective [1] [42]. On previous versions of this discussion, conscientious leadership was proposed as a likely successful tool in refraining thoughtless behavior on the leader's part, leading to positive outcomes in the community and society at large [33].

The discussion was expounding on conscientiousness as being that higher level of leader's commitment to the enduring good of the community [32] instead of the prevailing meism leaders today seem to portray through their actions [6]. The same discourse was suggesting such egotistical behavior, short term in nature and in the results it produced [38] was indicative of an attitude of *après moi , le deluge* [Let the flood start, after me!] [33], rather than emanating from a leader's inner [2] and transcendent [6] desire for enduring and sustainable positive outcomes [10] for society, and communities around the world [4] [6].

At that juncture, it became apparent including leadership spirituality in the discourse was a way to capture a missing albeit critical element in helping to schematize that higher level of leadership involvement in attempting to positively influence organizational and societal outcomes [2] [4] [6] [10] [12] [27] [30] [31] [32] [38] [40].

RESEARCH QUESTION

This study will discuss the still nascent notion of leadership and spirituality [4] [38] and how it is a factor in capturing conscientious leadership, defined in this review as a paradigm encompassing the following four dimensions: (1) accountability; (2) leader's concern for followers; (3) transparency; and (4) rationality.

RATIONALE FOR STUDY

The reason for such a discussion is there is such a thing as an environment either ripe for innovative ways of schematizing leadership styles most likely to be effective [6] or calling for searching such paradigms [27]. The current enduring economic crisis presents just such a setting [10] [31] [32] offering a backdrop to join in the exploration of leadership and spirituality [12] [38].

PURPOSE OF THE STUDY

Said differently, investigating leadership and spirituality will assist in assessing the sustainable positive impact [27] of leader's conscientiousness on individuals, organizations, and society at large [2] [32], daring to hope [37] this is the kind of leadership style most likely to result in the expected outcomes [6]. This proposal will introduce religion and spirituality, while the full version of the paper will expand on the current discussion, looking at a comprehensive model

of leadership spirituality and conscientiousness as the engine for sustainable positive economic performance, tested in a cross-cultural environment, that of the United States and France.

RELIGION AND SPIRITUALITY IN THE WORKPLACE

Religion and spirituality can be seen as two dimensions in the human experience [27]. [27] defines spirituality broadly as "being attentive to what inspires and drives us in our lives" (p. 124) and also as any activity that focuses on shared values that include the "whole self" (p. 124). Spirituality can also be seen as acting in alignment with individual core values where a deeper sense of meaning and purpose in life is conceptualized [6] [27]. Thus understood, it has less to do with religion [2] [12] [31] than it does with a heightened consciousness driving one towards the pursuit of excellence [30]. It encompasses the practice of ethical behavior and accountability exhibited in the leader's integrity and trust [38], as well as in a leader showing concern for followers and lastly, displaying transparency and rationality [10].

Spiritual matters, although personal to many people, are now a matter of public discourse, particularly as it pertains to the workplace. Upon closer examination it would appear spirituality is in fact, to some, a public or a less private matter since so much of what characterizes the workplace: conflict, toxicity and competiveness, can be addressed by integrating elements of spirituality into teams and by further extent, organizational dynamics [27].

As underscored in a grounded theory study on the role of retreats in a higher educational institution, spirituality can help employees appreciate and understand each other better [27]. Spirituality enhances loyalty and more positive team functionality [27]. [27] examined the role of retreats, specifically how these afforded participants with much needed time for reflection and contemplation. Time for both of these activities – reflection and contemplation – is often lacking in the workplace because these endeavors can take time, a precious resource many organizations are hesitant to commit. Retreats, the focus of the [27] study and other activities help give employees the space and permission to reflect on and contemplate their behavior and their connection to each other. This connection is what supports and encourages enhanced loyalty, interconnection and transparency [27] [31]. Employees who feel more connected are more loyal and as a result act in a more transparent and open manner with one another.

Whether or not spirituality can occur with or without religion is a current debate [31] on which we are not focusing in this discussion. The important notion is that spirituality is perceived as "a higher awareness that drives human beings to do well" [31, p. 885]; it incorporates interconnection, mutual respect and the notion that there is more to life than the pursuit or attainment of material or physical possessions. Especially, spirituality in this context is understood as driving a leader to seek opportunities for personal growth, leading to a purposeful quest on the leader's part, towards manifesting more care and compassion towards followers and by extension, all other stakeholders. Leaders exhibiting such behaviors increase connections with and among followers and reap enhanced organizational performance. Connection, dealing with another and with the individual self, is a common theme in the literature surrounding the issue of spirituality in the workplace [44].

Insecurity and alienation and other negative elements of globalization have created a need for "development on the spiritual side" [10] in the workplace. There has been a move globally to understand and create a more "meaningful work orientation" [10, p. 109] that involves the entire person. This has been accomplished through the establishment of, among business executives and professionals in the US and in London, "faith-based discussion groups" (p. 109) where participants argue the notion of a higher purpose, ethics and commitment to public service. Exploration and reflection are the hallmarks of these discussion groups' activities; self awareness and organizational interconnectedness are also emphasized. Clearly, transparency and accountability (acknowledgement of one's actions) are also at the core of current applications of workplace spirituality [2].

In the upcoming full version of this paper, data collected from the United States and France will be used to test a model of conscientious leadership and spirituality as operationalized in the four aforementioned dimensions of accountability, leader's concern for followers, transparency, and accountability. We will draw and test hypotheses in order to validate the construct under study. We are aiming at identifying the role spiritual leadership may have in yielding long term, positive and sustainable organizational and societal performance.

CONCLUSION

The preceding opens the way to a discussion of leadership success resulting from the spirituality aspect of leader's conscientiousness in which accountability [2] [4], leader's concern for followers [2] [40], transparency [4] and rationality will be assessed to determine if these constitute the inner motivation [12] [32], sense of purpose [38] [31], and altruistic behavior on the leader's part, conducive to long term positive economic outcomes [6] [10] [32]. A discussion of this nature can be of interest to both academicians and practitioners, as wells as policy-makers since many practitioners and academicians believe we have been looking at the wrong factors for too long. [44] indicated we must go back to some of the earlier studies of human nature and look at the most basic elements that were said to support organizational effectiveness: happiness and emotional contentment are two such examples. These can only be achieved when the spiritual – a connection to a source outside of and higher than one self – is supported and encouraged. In this sense spirituality becomes the essence of who the leader is. It elicits behavior conducive to sustainability and happiness, resulting in leaders acting for a purpose transcending themselves to benefit current and future stakeholders far and near, thus adding meaningfully but also positively to the interconnectivity of our society.

Nicole Jean Christian holds her Master of Public Administration with a concentration in Nonprofit Management from George Mason University (2001). Currently, she is a Senior Adjunct Professor of Management and Leadership at Dowling College in Oakdale Long Island. She is pursuing a Ph.D in Applied Management and Decision Sciences from Walden University with a specialization in knowledge and learning management. She is slated to defend in 2011.

Dr. Yvette Essounga-Njan earned her Ph.D in Business Administration and International Business at the University of Texas-Pan American in 2008. Currently, she is an Assistant Professor of Management at Fayetteville State University, in Fayetteville, NC.

Maxine Morgan-Thomas holds a J.D. (1991) from Columbia University School of Law. She also holds a Certified Public Accountancy license (New York). Currently, she is an Assistant Professor of Business Law at the Brooklyn campus of Long Island University in New York.

References

[1] Anderson, D. W., Krajewski, H. T., Goffin, R. D., & Jackson, D. N. (2008). A leadership self-efficacy taxonomy and its relation to effective leadership. The Leadership Quarterly, 19, 595-608.

[2] Avolio, B. J., Walumbwa, F. O., & Weber, T. J. (2009). Leadership: Current theories, research, and future directions. Annual Review of Psychology, 60, 421-449.

[3] Bedell-Avers, K. E., Hunter, S. T., & Mumford, M. D. (2008). Conditions of problem-solving and the performance of charismatic, ideological, and pragmatic leaders: A comparative experimental study. The Leadership Quarterly, 19, 89-106.

[4] Brown-Daniels, D. (2002). Leadership and spirituality: A response. Frontiers of Heath Services Management, 18(4), 39-41.

[5] Bruno, L. F.C., & Lay, E. G. E. (2008). Personal values and leadership effectiveness. Journal of Business Research, 61, 678-683.

[6] Bugenhagen, M. J. (2009). Spirituality and leadership: Engaging both in higher education. Journal of Leadership Studies, 3(3), 69-74.

[7] Byrme, G. J., & Bradley, F. (2007). Culture's influence on leadership efficiency: How personal and national cultures affect leadership style. Journal of Business Research, 60, 168-175.

[8] Cassiday, P. A. (2005). Expatriate leadership: An organizational resource for collaboration. International Journal of Intercultural Relations, 29, 391-408.

[9] Church, T. A., Katigbak, M. S., Reyes, J. A. S., Salanga, M. G. C., Miramontes, L. A., & Adams, N. B. (2008). Prediction and cross-situational consistency of daily behavior across cultures: Testing trait and cultural psychology perspectives. Journal of Research in Personality, 42, 1199-1215.

[10] Collins, P., Kakabadse, N. (2006). Perils of religion: Need for spirituality in the public sphere. Public Administration and Development, 26(2), 109-121.

[11] Crossan, M., Vera, D., & Nanjad, L. (2008). Transcendent leadership: Strategic leadership in dynamic environments. The Leadership Quarterly, 19, 569-581.

[12] Delbecq, A. L. (2009). Spirituality and business: One scholar's perspective. Journal of Management, Spirituality and Religion, 6(1), 3-13.

[13] Diawara, M. (2003). We won't budge. New York, NY: Basic Civitas Books.

[14]   Dickson, M. W., Hartog, D. N. D., & Mitchelson, J. K. (2003). Research on leadership in a cross-cultural context: Making progress, and raising new questions. The Leadership Quarterly, 14, 729-768.

[15]   Essounga, Yvette N. (2009). A review of the effect of national culture on corporate culture: an example of the United States & France. Review of Business Research, 9(5), 70-74.

[16]   Essounga-Njan, Y. & Morgan-Thomas, M. (2009). Leadership style effectiveness and accountability in times of crisis: A cross-cultural study of the USA and France. Int. J. Services & Standards 6(1), 62-78.

[17]   Flamholtz, E., & Kannan-Narasimhan, R. (2005). Differential impact of cultural elements on financial performance. European Management Journal, 23(1), 50-64.

[18]   Glader, P. (2009). Rebound for factories, but economic concerns linger. The Wall Street Journal, CCLV (3), A1 & A6.

[19]   Hall, A. T., Blass, F. R., Ferris, G. R., & Massengale, R. (2004). Leader reputation and accountability in organizations: Implications for dysfunctional leader behavior. The Leadership Quarterly, 15, 515-536.

[20]   Harrison, J. K,, & Clough, W. M. (2006). Characteristics of "state of the art" leaders: Productive narcissism versus emotional intelligence and Level 5 capabilities. The Social Science
[21]   The Leadership Quarterly, 13, 193-215.

[22]   Licht, A. N., Goldschmidt, C., & Schwartz, S. H. (2007). Culture rules: The foundations of the rule of law and other norms of governance. Science Direct, 35, 659-688.

[23]   Ligon, G. S., Hunter, S. T., & Mumford, M. D. (2008). Development of outstanding leadership: A life Journal, 43, 284-292.

[24]   Hofstede, G. (2001). Cultures consequences (2nd ed.). CA: Sage Publications.

[25]   Ketokivi, M., & Mantere, S. (2010). Two strategies for inductive reasoning in organizational research. Academy of Management Review, 35(2), 315-333.

[26]   Koene, B. A. S., Vogelaar, A. L. W., & Soeters, J. L. (2002). Leadership effects on organizational climate and financial performance: Local leadership effect in chain organizations. narrative approach. The Leadership Quarterly, 19, 312-334.

[27]   Lips-Wiersma, M. (2006). Business as an agent for world benefit? The Business Renaissance Quarterly, 1(3), 25-33.

[28]   Lips-Wiersma, M. (2006). The role of spiritual retreats in higher education: The necessity for, and difficulties with, reflection on lasting values, in a new public management context. Public Administration and Development, 26(2), 123-133.

[29]   Lussier, R. N. (2010). Human relations in organizations (8th ed.). NY: McGraw-Hill/Irwin.

[30]   Marques, J. (2006). The spiritual worker: An examination of the ripple effect that enhances quality of life in – and outside the work environment. Journal of Management Development, 25(9), 884-895.

[31]   Marques, J., Dhiman, S., & King, R. (2009). Leader, know thyself: Crisis, values, and spirituality. The Business Renaissance Quarterly, 4(2), 13-19.

[32]   Milliman, J., & Ferguson, J. (2008). In search of the "spiritual" in spiritual leadership: A case study of entrepreneur Steve Bigari. The Business Renaissance Quarterly, 3(1), 19-40.

[33]   Morgan-Thomas, M., & Essounga-Njan. (2010). Leadership success in the United States and France: When conscientiousness leads the way to economic recovery. Paper presented at the meeting of the Northeast Business And Economics Association, Morristown, NJ.

[34]   Matthews, D. J. (February, 2010). Trust me: Credible leadership delivers results. Chief Learning Officer, 28-31.

[35]   Nadeau, J. B., & Barlow, J. (2003). Sixty million Frenchmen can't be wrong. Naperville, IL: Sourcebooks, Inc.

[36]   Rotberg, R. I. (2009). Governance and leadership in Africa: Measures, methods and results. Journal of International Affairs, 62(2), 113-125.

[37]   Obama, B. (2006). The audacity of hope: thoughts on reclaiming the American dream. NY: Crown Publishers.

[38]   Pawar, B. S. (2009). Workplace spirituality facilitation: A comprehensive model. Journal of Business Ethics, 90, 375-386.

[39]   Testa, M. R. (2009). Nationl culture, leadership and citizenship: implications for cross-cultural management. International Journal of Hospitality Management, 28(1), 77-85.

[40]   Usman, A., Danish, R. Q. (2010). Leadership spirituality in banking professionals and its impact on organizational commitment. International Journal of Business and Management, 5(3), 185-193.

[41]   Waldman, D. A., & Galvin, B. M. (2008). Alternative perspectives of responsible leadership. Organizational Dynamics, 37(4), 327-341.

[42]   Wible, M. V. (2010). Enhanced governance and risk oversight disclosure for investment companies. The Investment Lawyer, 17(3), 14-19.

[43] Wilson, K. S., Sin, H.-P., & Conlon, D. E. (2010). What about the leader in leader-member exchange? The impact of resource exchanges and substitutability on the leader. Academy of Management Review, 35(3), 358-372.

[44] Wright, T., Cropanzano, R. (2004). The role of psychological well-being in job performance: A fresh look at an age-old quest. Organizational Dynamics, 33(4), pp. 338-351.

[45] Yukl, G. (2006). Leadership in Organizations (6th Ed.). Upper Saddle River, NJ: Pearson Education, Inc.

IDENTITY CONSTRUCTION OF RECENT INDIAN IMMIGRANTS AND ITS IMPACT ON

COMMUNICATION IN THE ATLANTIC CANADIAN WORKPLACE: A THEORETICAL

APPROACH

Peruvemba S Jaya, PhD
Associate Professor
Department of Communication
University of Ottawa
554 King Edward Avenue
Ottawa
ON
K1N 6N5
Canada
Tel: 613-562-5800 extn 2538
Fax: 613-562-5240
Email:jperuvem@uottawa.ca

ABSTRACT

This paper develops a theoretical approach using the concepts of ethnic identity Eisenbruch, 1984; Phinney, 1990), social identity, (Hogg & Terry, 2001) postcolonial theory (Bhabha, 1990) and intercultural communication specifically co cultural communication theory (Orbe, 1996) to understand the processes of construction and formation of identity of recent Indian immigrants in the Canadian context. Specifically, the goal is to use the approach developed to understand this process in the Atlantic Canadian context, by collecting data in future research. This is a region of unique immigration patterns and the experience is likely to be very different from the rest of Canada. This is particularly important as immigration into Canada has an impact on the context of the workplace, and communication in the workplace setting.

Key words: ethnic identity; social identity; postcolonial theory; co cultural communication theory; workplace communication.

Women's Networks in France and Québec: Motivations, Perceived Benefits and Expectations of Members

Dr. Krista Finstad-Milion*

ICN Business School
krista.finstad-milion@icn-groupe.fr

Sophie Schwartz*

ICN Master Grande Ecole Alumni
sophieschwartz.qc@gmail.com

Angie Celaya

ICN Executive MBA Alumni
angie_celaya@hotmail.com

For Graen (2007), women's web-building knowledge and skills are on the average superior to men's from early childhood due to a number of socialisation practices. Building on earlier studies (Uhl-Bein, Graen and Scandura, 2000), the author and associates observe that women tend to build better mutual-aid webs, or networks. Although organisational psychologists contribute to our understanding of women's predominant role in developing professional networks in the electronic age, economists observe by way of statistics, a more modest trend of women contributing to economic activity (Ferro-Vallé, 2009). Likewise, gender researchers observe the slow advancement women are making in the political sphere and upper echelons of organisations (Burke and Mathis 2005; Ferro-Vallé, 2009). We may consequently ask why women in the working world engage particularly in women's professional networks.

Building on Blau and Scott (1966), and the stakeholder theory (Freeman, 1984), membership is the prime beneficiary of mutual-aid networks. We propose to carry out a qualitative and quantitative analysis of the nature of membership in three different types of women's networks, drawing on the typology of Celaya (2010). The types are the external professional network, the internal firm network and the entrepreneur network. Why do women choose to become a member of one of

these types of women's networks? What are the expected benefits of women belonging to such networks? Furthermore, how and why do members of women's networks become engaged, more or less intensively, in network activities? The chosen methodology will be based on exploratory individual interviews followed by statistical analysis of a questionnaire sent out to members of women's networks in France and Quebec. The purpose of this research paper is to identify the motivations, perceived benefits, and future expectations of members of women's networks. The authors aim to distinguish current trends, on an international level, for enhancing the attractiveness of women's networks which are in the early stages of development.

Bibliography

Blau, P. and Scott, W. (1966), *Formal Organisations*. Routeledge and Kegan Paul: London.

Burke, R. and Mathis, M. (2005), *Supporting Women's Career Advancement: Challenges and Opportunities*, Edward Elgar: Cheltenham.

Celaya, A. (2010), *The Next Generation Of Women's Networks*, MBA Dissertation, under the direction of K. Finstad-Milion, ICN Business School, Nancy, France.

Ferro-Vallé, E. (2009), *Professionnelle entre les femmes et les hommes : Comprendre et Agir*, Afnor éditions, La Plaine.

Freeman, R. (1984), *Strategic Management: A Stakeholder Perspective*. Englewood Cliffs. New Jersey: Prentice Hall.

Graen, G. (2007), *Jessica's Web: Women's Advantages in the Information Age*. Information Age Publishing, North Carolina: Charlotte.

Uhl-Bein, M., Graen G. & Scandura, T. (2000), "Leader-member exchange (LMX) for strategic human resource management systems: Relationships as a social capital for competitive advantage", in *Research in Personnel and Human Resource Management*, 18, pp. 137-185.

APPENDIX A: Interview Guide / ENGLISH

Participant: …………………. Women's Professional Network
(WPN): …………………… Date: …………………….

Age: …………. Length of time as a member of the
WPN: …………..

1-Your motivations

1.1 Why did you choose to become a member of this Women's Professional Network?

1.2 Were you already part of other professional networks before joining this Women's Professional Network? If so which ones?

1.3 How do you contribute to this network? What are your key findings from this involvement?

1.4. What factors condition your current contribution to this network?

1.5. What factors would intensify your contribution to this network?

2-Perceived benefits

2.1 What are the differences that you found in belonging to a Women's Professional Network this network compared to belonging to other professional networks?

2.2 How does your involvement in this network help you in your professional career?

2.3 What are the new professional competencies that you have developed thanks to your participation in this Women's Professional Network?

3-Your future expectations

3.1 Looking ahead, in your future professional life, how might this particular Women's Professional Network have helped you to reach a higher performance?

3.2 Now, looking at the development of this Women's Professional Network itself, how might this network further develop to better answer your individual needs?

3.3 In the future, what would you like to be able to say to new and future members, institutions or investors about the added value of this Women's Professional Network to your organisation/firm, or to your local economic environment?

THANK YOU !

APPENDIX B: Guide d'Entretien / FRANCAIS

Participant : …………………. Réseau Professionnel
Féminin (RPN): ………………… Date :
…………………….

Age: …………. Période de temps en tant que membre du
RPN: …………..

1-Vos motivations

1.1 Pourquoi avez-vous choisi de devenir membre de *ce Réseau Professionnel Féminin?*

1.2 Faisiez-vous déjà partie d'un Réseau Professionnel Féminin avant de joindre ce réseau? Si oui, le(s)quel(s) ?

1.3 Comment participez-vous dans la vie de ce réseau ?

1.4 Quels enseignements tirez-vous de cette implication *(ou participation)?*

1.5 Quelles sont vos motivations pour participer à ce réseau ?

1.6. Qu'est-ce qui pourrait-vous inciter à participer davantage à ce réseau ?

2-Bénéfices perçus

2.1 Quelles différences avez-vous perçues de votre appartenance à ce Réseau Professionel Féminin comparé à d'autres réseaux professionnels que vous avez connu?

2.2 Comment cette implication dans ce Réseau aide-t-elle votre carrière professionnelle?

2.3 Quelles nouvelles compétences professionnelles avez-vous développées grâce à votre participation à ce Réseau Professionnel Féminin?

3-Vos attentes futures

3.1 En se projetant vers l'avenir, dans votre vie professionnelle future, comment ce Réseau Professionnel Féminin pourrait-il vous aider à accroître votre réussite?

3.2 Et maintenant, du point de vue du Réseau Professionnel Féminin lui-même, comment ce réseau pourrait-il se développer davantage pour mieux répondre à vos besoins?

3.3 Dans l'avenir, qu'est-ce que vous aimeriez pouvoir dire à de nouveaux membres, institutions ou investisseurs, sur la valeur ajoutée de ce Réseau Professionnel Féminin pour votre entreprise / institution, ou pour votre environnement économique local?

MERCI !

MANAGING TELECOMMUTERS' WORK-FAMILY BOUNDARIES

Kellyann Berube Kowalski, Department of Management & Marketing, Charlton College
of Business, University of Massachusetts Dartmouth, 285 Old Westport Road,
North Dartmouth, MA 02747, (508) 999-8327, kkowalski@umassd.edu

Jennifer Ann Swanson, Department of Business Administration, Stonehill College,
320 Washington Street, Easton, MA 02357, (508) 565-1349, jswanson@stonehill.edu

INTRODUCTION

With the advent of information technology, more and more individuals began working from home and the integration of work and home became easier and more acceptable. A recent study from WorldAtWork [8] found that the number of Americans working from home or remotely at least one day per month rose 39% from 2006 to 2008, from 12.4 million to 17.2 million. It was also reported that the 2008 numbers represented a 74% increase from 2005 [8]. There are obvious benefits to this integration available in the information age. Telecommuting from home enables one to have more flexibility to balance work and family. These workers see telecommuting as more integrative and therefore, as an opportunity to better balance their work and family lives.

Although the view that home and work should be kept separate had been the norm since the industrial revolution, that view has become unrealistic as it is increasingly impossible to keep the two worlds separate in today's information age. Despite the obvious benefits that come from the flexibility associated with telecommuting from home, there are also costs that have to be minimized. When individuals telecommute from home the boundaries between work and family become blurred and this can lead to high levels of work-family conflict. In particular, telecommuting has been linked to a higher level of family-to-work conflict [2] [3].

Our completed paper will address the role integration telecommuters experience by working from home and how this causes boundaries to become more blurred. We will discuss boundary theory as we examine the benefits and costs of role integration. We will propose a framework that suggests strategies for minimizing the costs by managing the physical, temporal, and psychological boundaries between work and family roles.

Below is an extended abstract of the proposed framework.

MANAGING BOUNDARIES

As the role integration telecommuters experience by working from home causes boundaries to become more blurred, it is necessary to actively manage these boundaries. We postulate that telecommuters experience work-family conflict and blurred boundaries because for many becoming a telecommuter is a strategy encompassing high role integration. Although highly integrated roles do allow for the flexibility many telecommuters are seeking, the boundaries between such highly integrated roles must be effectively managed.

According to boundary theory [4] [5] [7] individuals create and maintain boundaries between roles in order to make roles clearer and to create order. Some people allow boundaries to be crossed over more easily than others, whereas some use boundaries to clearly keep roles separated [5]. Boundaries can thus help to keep roles delineated and separate [1] and allow one to clearly know or concentrate on whichever role they are in [5]. In order to minimize the negative effects of the role integration associated with telecommuting, we are suggesting a model of boundary management. Our model proposes different strategies for managing three different types of boundaries between work and family roles: physical, temporal, and psychological.

MANAGING PHYSICAL BOUNDARIES

Physical boundaries separate the physical locations in which work and family activities take place. In telecommuting, the physical boundaries between work and family activities often become blurred and sometimes even nonexistent. For example, in many telecommuting situations, the office computer and the home computer are one in the same and the telecommuter's office doubles as a homework center and/or bill paying location. Often times work and family space become one in the same. In such situations physical boundaries between work and family roles are extremely flexible and permeable as roles are fully integrated. Obviously not all telecommute situations are integrated to such a great extent, but nonetheless it is important to manage the physical boundaries between work and family roles. There are several strategies that can be enacted to help telecommuters to better manage physical boundaries.

Separate Workspace

First and foremost, it is important for telecommuters to have a separate workspace where they conduct their business. The workspace should be in a location that is used solely for business and one that is not also used for family activities. The separate workspace should include a telephone line, computer, fax, etc. that are used just for the business. Family activities, such as homework and bill paying should be done in a separate location and on different equipment.

Client Rules of Interaction

In order to minimize boundary crossing by clients entering family space, client rules of interaction should be enacted. Telecommuters should develop a set of rules regarding the specifications of client interactions, including where and when they can take place. If at all possible, it is best to have a separate entry for the workspace where clients can enter without invading the home space.

Family and Friend Rules of Interaction

In order to minimize boundary crossing by family and/or friends entering family space, family and friend rules of interaction should be enacted. Telecommuters should develop a set of rules regarding the specifications of family and friend interactions. It is often the case when working from home, that family and friends do not perceive telecommuters as "working." It is very

difficult to get work done if family and friends can cross the boundaries between work and family spaces at any time. Just as family and friends would not ordinarily just stop by employees' places of work if they were at a business location, they also should not just stop by telecommuters' home-based workplaces.

Plans Regarding Vacation Time

When telecommuting from home, it is very easy to work 24/7 almost 365 days out of the year – even during vacation. There is always something left to be done and telecommuters are never truly away from work, since their work is home-based. Therefore, it is very important for telecommuters to have a plan regarding vacation time and that if possible vacation be spent at a location away from the home and home-based office.

MANAGING TEMPORAL BOUNDARIES

Temporal boundaries separate the times in which work and family locations take place. In a home-based telecommuting situation, the temporal boundaries between work and family activities often become blurred and overlap to a great extent. Many telecommuters try to take care of home and work activities at the same time. For example, talking on the phone to a client or co-worker while cooking dinner. In such situations temporal boundaries between work and family are extremely flexible and permeable as roles are fully integrated. There are several strategies that can be employed to help telecommuters better manage temporal boundaries.

Time Management

In order to keep work and family separate in terms of time it is important to properly manage that time. Using time management techniques, telecommuters can more effectively keep the temporal boundaries between work and family from blurring or falling down completely. For example, using a planner to keep track of when work activities are scheduled as well as when family activities are scheduled will help to limit the number of scheduling conflicts between the two roles.

Delegation

As with any manager, telecommuters need delegation in order to be most effective and efficient. No one can do it all alone. Telecommuters who effectively delegate both home and work tasks will be able to better balance work and family boundaries.

Family Instrumental Social Support

Social support is "an exchange of resources between at least two individuals and perceived by the provider or the recipient to be intended to enhance the well-being of the recipient" [6, p. 13]. Instrumental social support is defined as helping behaviors such as loaning money or giving one's time and skill or advice. Having instrumental social support from the family has been found to decrease the work-family conflict experienced by employees. Therefore,

telecommuters with a strong family network, which provides instrumental social support, will be better able to balance work and family boundaries.

Outside Domestic Help

If affordable and desirable, telecommuters should hire someone to take care of home chores, like cleaning the house. This will allow telecommuters more time to spend leisure time with family or take care of business instead of domestic chores.

MANAGING PSYCHOLOGICAL BOUNDARIES

Psychological boundaries compartmentalize work and family in the mind. It is extremely easy to physically and temporally be in one role, yet psychologically in the other. For example, telecommuters may have succeeded in effectively managing the physical and temporal boundaries between work and family. Working from a separate office and developing a routine with specific times for carrying out work activities, telecommuters may be in the workspace, conducting business, but thinking about what to make for supper or how the laundry is piling up. Psychological boundaries are very easy to cross and therefore need to be properly managed. There are several strategies that can be utilized to help telecommuters better manage psychological boundaries.

Developing Rituals

Telecommuters can more easily manage the psychological boundaries between work and family roles when they consciously perceive themselves as moving from one role to the other. One way telecommuters can do this is by developing rituals (such as leaving the house and entering through a different door) to help make the transition from one role to the other role easier and distinct.

Getting a Role Model or Mentor

Telecommuters can benefit from having someone who has worked from home successfully balancing work and family roles to talk to and discuss what they are going through.

Setting Realistic Expectations
If telecommuters expect themselves to be perfect in every role, they will be setting themselves up for failure and will only feel guilty about it. Telecommuters need to realize that they don't have to be the ones to do everything. For example, get takeout instead of a home cooked meal and don't feel guilty about it.

Connecting with Colleagues (Business Contacts)
Telecommuting from home can be isolating. Telecommuters may feel disconnected from the office and from colleagues. Therefore, it is important for telecommuters to connect with coworkers and colleagues and/or join professional organizations like the Rotary club, Chamber of Commerce, etc.

Family Emotional Social Support

Emotional social support is defined as caring behaviors such as the provision of trust, empathy, love or evaluative feedback. Telecommuters can benefit from emotional social support such as their families being okay with no homemade dinners, understanding what they are trying to accomplish, and being there to listen when telecommuters need to talk.

CONCLUSION

In conclusion, we believe that our completed manuscript and framework will provide telecommuters with useful tools to better manage the blurring work and family boundaries they experience when working from home.

REFERENCES

[1] Ashforth, B.E. and Humphrey, R.H. (1995). Labeling processes in the organization: Constructing the individual. In L.L. Cummings & B.M. Staw (Eds.), *Research in Organizational Behavior*, vol. 17, pp. 413-461. Greenwich, CT: JAI Press.

[2] Felstead A. and Jewson, N. (2000). In work, at home. Towards an understanding of home Working. Routledge, Taylor, & Francis Group, U.K.

[3] Golden, T., Veiga, J. and Simsek, Z. (2006). Telecommuting's Differential Impact on Work-Family Conflict: Is There No Place Like Home?, *Journal of Applied Psychology*, vol. 91, no. 6, pp. 1340–1346.

[4] Michaelsen, S. and Johnson, D.E. (Eds.) (1997). Border theory: The limits of cultural politics. Minneapolis, MN: University of Minnesota Press.

[5] Nippert-Eng, C.E. (1996). Home and work: Negotiating boundaries through everyday life. Chicago, IL: University of Chicago Press.

[6] Shumaker, S.A. and Brownell, A. (1984). Toward a theory of support support: Closing conceptual gaps, *Journal of Social Issues*, vol. 40, Winter, pp. 11-36.

[7] Zerubavel, E. (1991). The fine line: Making distinctions in everyday life. New York, NY: Free Press.

[8] WorldatWork. (2009). Telework Trendlines.

DOES MOBILE TECHNOLOGY HELP OR HINDER MEETING PRODUCTIVITY?

Kristen M. Dietel, Roger Williams University, 1 Old Ferry Rd., Bristol, RI 02809, (401)254-3175, kdietel311@g.rwu.edu

Scott P. McMann, Roger Williams University, Roger Williams University, 1 Old Ferry Rd., Bristol, RI 02809, (401)254-3175, spmcmann@gmail.com

Susan M. Bosco, Roger Williams University, 1 Old Ferry Rd., Bristol, RI 02809, (401)254-3175, sbosco@rwu.edu

Diane M. Harvey, Roger Williams University, 1 Old Ferry Rd., Bristol, RI 02809, (401)254-3018, dharvey@rwu.edu

ABSTRACT

The preponderance of web-based communication technologies in the workplace has changed behaviors of meeting participants. Specifically, individuals move between face-to-face and person-to-device interactions. Some claim that the use of technology during meetings improves information-gathering and schedule coordination. Others, however, view the use of communication devices as delaying meeting progress. There were 215 respondents to this survey of technology use in the workplace. Initial results indicate that 62 percent either agreed or strongly agreed that personal technology use in a meeting is offensive. According to 58 percent of respondents, the use of mobile devices results in less productive meetings.

Key Words: Generational differences, technology, meetings

INTRODUCTION

The increasing use of mobile technology by individuals in a variety of workplace situations and interactions necessitates a better understanding of the impact of such use on individual productivity, organizational productivity, and the quality of human interactions. The goal of this research was to gain a better understanding of employees' personal opinions of the use of mobile technology in the workplace. We focused on three aspects of technology use: Personal Productivity, Effect on Organizational Productivity, and Social Acceptance.

In addition to these areas, we also found through our literature review that there appear to be generational differences in the use and perception of use by others of mobile devices. Therefore, we attempted to determine where some of those differences might lie. The terminology used for these personal technology devices varies; there appears to be no standard for the decision to use mobile device versus personal technology, for example. In this paper, these terms refer to the use of cell phones, Blackberrys (and similar items such as iPhones), and laptop computers.

PERSONAL PRODUCTIVITY

Personal preference has much to do with the way mobile technology use during meetings is regarded. Some may feel that is entirely appropriate to use a cell phone in the presence of others, while others may deem it unacceptable. Much of their opinion has its source in generational standards, societal expectations and organizational cultures. For example, research shows that 69 percent of Baby Boomers who took part in a survey agreed that PDAs and mobile phones contribute to the decline of proper workplace etiquette [9]. Some individuals make the decision regarding mobile device use based on the size of the meeting. In large meetings, for example, some find it is more productive to engage in electronic work than to listen to presentations that are only tangentially related to their jobs [10].

While this may be true for some, in other situations, individuals found it less productive because they became distracted by their own or others' usage. Laptops, for example, can be helpful to take notes and multitask. However, for some, it can become overwhelmingly distracting. People may have difficulty balancing between listening and becoming lost in the work on their laptops. Others reported finding it off-putting when those around them are typing loudly or looking at irrelevant websites [4].

The use of mobile technology may have less to do with personal preference, however, and more to do with company's culture. At some organizations, such as Microsoft, it would be peculiar to *not* bring a laptop to meetings [4]. The company will have expectations or cultural standards that can help dictate whether or not it is appropriate to use such devices during a meeting. Often these practices will reflect the kind of industry the company is in, or the products they sell. Certain departments may find the use of technology takes away from their practices. Human resources, for example, may require more face-to-face interaction. The use of mobile technology can slow the development of workplace relationships that are necessary to effectiveness. By using technology when one is face-to-face with others, it sends a message that these individuals are not important. It will take time to rebuild that connection once it is damaged by the interruption of

technology [13]. These kinds of relationships cannot afford to be severed in a business world that is becoming more dependent on team-based activities.

Because of the importance of mobile device use as it relates to individual productivity in the workplace, we asked the following research questions:

Research Question 1: How does my personal use of technology affect my productivity?
  Research Question 1A: Does the view of personal use of technology differ by generational membership?

Research Question 2: Does the use of personal technology by a person in a face-to-face meeting negatively impact that relationship?
  Research Question 2A: Does the use of personal technology by a person in a face-to-face meeting negatively impact that relationship regardless of generational membership?

ORGANIZATIONAL PRODUCTIVITY

The use of communication technologies seem to provide connectivity among workers, however, in reality, this may only be an illusion. The use of technologies complicates the communication process, as it removes many of the cues we use in face-to-face encounters [2]. Because of the distance created among employees, the efficiency of the technology is lost as email becomes more about documenting that work was done, rather than doing it better. Duffy [2] proposes that more face-to-face communication is needed to ensure that electronic exchanges result in productive outcomes.

Some individuals find that use of these devices has resulted in more productive meetings because the meeting leader feels more compelled to make the meeting content valuable to the attendees [10]. Others have a different view, however, contending that meetings last longer when people are using devices because information has to be repeated for those who were not paying attention. The rudeness of using these devices and absenting oneself mentally from the meeting can result in poor team relations as well, damaging future work products [7].

In some organizations, workers are required to leave their devices at the door of the meeting room, according to Silverman [10] and Weinstein [13]. These practices are indicative of the type of culture managers wish to establish for their company and within meetings. It is unclear whether many companies are incorporating these policies and what they entail.

The use of these policies may reflect the perception that managers have regarding the improvement in productivity that they expect from the multitasking behaviors they observe. Is the availability of technological capabilities contributing to more productive organizations? For example, one employee cites communication with another person who is engaging in five conversations at once – you know you are only getting 20% of the person's attention – is that productive [10]? One example of the negative result of this type of multitasking at one organization was the loss of a strong job candidate who was put off by the texting of one of the pair of an interview team [5].

Decision making can also be affected by the use of technology. In one survey, a respondent expressed the view that emails delay decision-making, stating "I find that many executives avoid conversation because they may be forced to make a decision or express an opinion. If they can keep communications within e-mail, they can continually pass the buck around or back without having to commit. Management by failure to act may be the new favored process" [6].

There may be a need to further examine the appropriateness of technology use "We're beginning to learn that different methods of communication are more effective at certain tasks than others," said another respondent. "E-mail is great for scheduling and confirming meetings, phone is good for quick conversations that require two-way communications and a memo is preferred for long background pieces. In-person and scheduled meetings are always the best for any discussion requiring true dialogue and consensus" [6]. In these cases, using technology can be effective. "Instant messages and e-mail are communication accelerators," said one executive. "Discussion databases are more efficient for larger groups [6].

> Research Question 3: How does mobile technology affect productivity?
>> Research Question 3A: Does the perception of how technology affects productivity vary by generational membership?

> Research Question 4: What should organizations do about policies for mobile device use?

> Research Question 5: Are organizations enacting policies for mobile device use?

> Research Question 6: What organizational training policies should exist for mobile technology use?
>> Research Question 6A: Do beliefs regarding organizational training policies for mobile device use differ by generational membership?

SOCIAL ACCEPTANCE

The examination of the appropriateness of technology use extends to the perceptions this activity generates for employees when interacting with someone else who is using a mobile device(s). An additional consideration is whether this use is viewed differently depending on whether it occurs in meetings or during individual encounters in the workplace.

Meeting Behaviors/Multitasking

Expectations about the use of personal technology in meetings as well as during face-to-face communication can vary according to generational membership as well. The perception of whether such behavior is rude, tolerable, or efficient will affect the interactions of the individuals involved. For example, it appears that there is a generational gap regarding the manner in which the use of technology during meetings is viewed. Baby Boomers are much more likely to be annoyed by younger workers checking their mobile phones for email and texting during meetings [9]. Both Boomers and Veterans consider the use of mobile devices during meetings "as gauche as ordering out for pizza" [10]. There is not a clear consensus, however, among

younger workers. Some Gen Y workers surveyed also found the use of these devices during meetings to be "rude". Other considered it risky not to respond to emails and texts promptly.

Some workers perceive those who email and text during meetings as trying to retreat from meeting-related obligations. One individual stated that "They (those using phones during meetings) are the first ones to ignore the same emails/texts when they're on their own time" [10]. They are viewed as trying to avoid assignments that may arise from the meeting by not participating in it. Another person questioned whether a worker would be using the phone if he/she was seated next to the boss at the meeting. Some workers, however, are comfortable with their use of cell phones during meetings. Mike McAfee, executive director of the Monroe County Convention and Visitors Bureau, said he would pick up calls and answer messages from family members because one never knows if it's an emergency. He also would not think twice about using his iPhone to take notes and check his schedule in meetings [12].

Regardless of the potentially negative perception it generates, personal technology use during meetings seems to be increasing. A recent New York Times article reported that a third of more than 5,300 workers polled in May by Yahoo Hot Jobs said they often checked e-mails in business meetings. Nearly one-fifth of the respondents said they have been reprimanded by their employers or fellow workers for bad behavior with their wireless device [12].

One way to address meeting behavior with regard to personal technology use is to set ground rules at the beginning of the meeting. An executive interviewed by Pansch [8] stated that he set ground rules at the start of meetings. He allowed cell phone use when necessary to obtain answers to clients' questions or check on information needed to proceed with the meeting's topic. He considered non-meeting related cell phone use to be rude. It is also suggested that in small meetings, one ask permission to use the device rather than just proceed [1]. It is also expected that such use would be meeting-related.

The Etiquette Question

The etiquette question continues as well. In a Robert Half Technology survey, 51 percent of CIOs (sample of 1400) said the rise in poor workplace etiquette can be attributed to the increased use of mobile devices [5]. In this same survey, 67 percent of them said breaches in technology etiquette are more common today than three years ago [11]. The usage patterns that are viewed as rude vary, with one example coming from Judith Sharp, Monroe County tax assessor, who said that it disturbs her when people pay more attention to their smartphones than to people they are with. Greg Main, president of i2E a nonprofit organization that manages the Oklahoma Technology Commercialization Center, said he is most offended by blatant "multitaskers" with whom he is trying to engage in conversation. "It could be a variety of things from cell phones to e-mail," Main said, "It is just rude behavior" [11].

Another breach of tech etiquette identified in this survey is when people leave their cell phone ringers on during a meeting, as noted by 88 percent of respondents. A close second, at 80 percent, is sending instant messages, while sending and receiving e-mail during a meeting is third, with 79 percent saying it's annoying [11]. In Tung's [12] interviews of workers about technology use in the workplace, one respondent stated that "When the person I am speaking

with looks down at his or her BlackBerry, it's the same as turning away to talk to other people. It is very rude. They are not giving you full attention."

Preferred Communication Methods

There appears to be no consensus with regard to whether the use of emails, texts, telephone or face-to-face organizational communication results in more productivity. In a global survey, 67 percent of senior executives and managers said their organization would be more productive if their superiors communicated more often by personal discussion. In contrast to their opinions in this regard, the top personal method of communicating for these same business leaders is e-mail, based on the survey by NFI Research [6]. One respondent stated that "Personal discussion is the foundation of communications." Another said that personal communication provides a foundation and "enables all of the other forms of communication. Having a personal connection builds trust and minimizes misinterpretation and misunderstanding" [6].

The difference in opinion could have some basis in generational membership. Older individuals are more likely to believe that using email is taking the easy way out of interactions, but is not always the most efficient. They believe that time is saved by calling or meeting others face-to-face to answer questions and respond to issues. One survey respondent stated "I often find that when I look the other person in the eyes and ask them something I get far more than I ever would over e-mail." Younger workers are less likely to find texting, emailing, and web surfing distracting. According to Schonfeld [9], these individuals are more likely to find these activities "productive and efficient." " The Blackberry is a great tool for getting information, setting schedules, making appointments and getting more done," said a respondent. "It's increased the pace to get more done; it's doubled the pace. It gives me more opportunities to get business" [8].

> Research Question 7: Is the use of personal technology in meetings considered to be rude?
>> Research Question 7A: Are there generational membership differences in whether it is considered rude to use personal technology in meetings?

> Research Question 8: Is there acceptance of, or participation in, multi-tasking activities during meetings?
>> Research Question 8A: Are there generational differences in acceptance of or participation in, multi-tasking activities during meetings?

> Research Question 9: Is the use of personal technology while talking to others considered rude?
>> Research Question 9A: Are there generational differences in whether use of personal technology while talking to others is considered to be rude?

> Research Question 10: Does more intra-organizational communication take place via email versus face-to-face?
>> Research Question 10A: Are there generational differences in use of email versus face-to-face communication is used in organizations?

METHODOLOGY AND RESULTS

This research study used a snowball sampling method to obtain respondents to a paper survey. The survey was approved by the HSRB of the University and gathered both quantitative and qualitative data. Respondents represent the four generations in the workplace today; however, due to the low response rates for Veterans and Generation X individuals, analysis was done for data from Baby Boomers and Generation Y only. Since the data was primarily categorical, chi-square analysis was most often used.

There were 247 respondents. Males represented 46 percent of the sample. Generational membership consisted of 1 percent Veterans, (born before 1945) 14 percent Baby Boomers (born between 1945 and 1965), 10 percent Generation X (born between 1965 and 1980), and 76 percent Generation Y (born after 1980).

Personal Productivity

Research Question 1: How does my personal use of technology affect my productivity? This question was addressed through three questions asking about the effect on personal productivity of use of mobile devices during meetings, the ability to ignore mobile devices during a meeting and whether the use of these devices during meetings is distracting. There was no clear consensus about the effect of personal productivity with the use of mobile devices during meetings, although responses were more positive - evenly divided among strongly agree, agree, and neutral. Most individuals were able to focus on the matter at hand during a meeting and ignore these devices. It appears that the use of personal technology by others during meetings is a source of distraction with more than half of respondents indicating that they strongly agree or agree. Research Question 1A: Does the view of personal use of technology differ by generational membership? There were no generational differences regarding personal productivity and mobile device use in any of these areas.

Research Question 2: Does the use of personal technology by a person in a face-to-face meeting negatively impact that relationship? There was general agreement in viewing this type of activity as having a negative effect on a personal relationship with results evenly divided among strongly agree, agree, and neutral. Research Question 2A: Does the use of personal technology by a person in a face-to-face meeting negatively impact that relationship regardless of generational membership? There were no generational differences in this situation.

Organizational Productivity

Research Question 3: How does mobile technology affect productivity? In order to answer this question, we asked whether the use of mobile devices during meetings results in less productive meetings. More than half of respondents agreed or agreed strongly. We also asked whether the use of mobile devices during meetings results in longer meetings and more than half of individuals either agree or strongly agree with this. Finally, we asked whether respondents had established their own individual policies about mobile device use during meetings. Half stated they ignored the use of devices and the other half dealt with the situation on a case-by-case basis.

Research Question 3A: Does the perception of how technology affects productivity vary by generational membership? There were no generational differences in perceptions about the loss of productivity during meetings when mobile devices are used by meeting participants. There was, however, a difference (chi-square <.00) in the manner in which individuals dealt with mobile device use in their own meetings with Baby Boomers more likely to resolve it on a case-by-case basis and Gen Y workers just as likely to ignore it as to resolve it case-by-case.

Research Question 4: What should organizations do about policies for mobile device use? In order for us to determine whether organization recognize the potential loss of productivity as a result of the use of mobile devices, we asked whether organizations should have policies regarding the use of technology at meetings and at the workplace. Most, 75 percent indicated that they agreed or strongly agreed that there should be policies governing the use of mobile devices at meetings and even more agreed or strongly agreed that these policies should be instituted for the workplace as a whole. In both cases, there were significant (chi-square <.00) generational differences with more Baby Boomers than Gen Y individuals in agreement.

Research Question 5: Are organizations enacting policies for mobile device use? For 70 percent of the individuals, the company where they work has no policy governing the use of mobile devices. But when asked whether there should be policies, only 52 percent agreed.

Research Question 6: What organizational training policies should exist for mobile technology use? Employee training should include proper use of mobile devices.
Research Question 6A: Do beliefs regarding organizational training policies for mobile device use differ by generational membership? There was a significant difference between generations (chi-square <.00), with more Baby Boomers in agreement with establishing training than Gen Y members.

Social Perception

Research Question 7: Is the use of personal technology in meetings considered to be rude? To answer this question, respondents were asked for their reactions to several workplace situations: mobile device use for personal activities while in a meeting, lack of attention to a meeting presenter and reaction to an individual who checks their mobile device during a conversation. Research Question 7A: Does this view differ by generational membership? There were significant differences among generations in this regard. More Baby Boomers than Gen Y individuals viewed the use of mobile devices during a meeting negatively (chi-square <.05) as well as this use during personal conversation (chi-square <.00), and lack of attention when acting as a meeting presenter (chi-square <.00).

For Research Question 8 - Is there acceptance of, or participation in, multi-tasking activities during meetings – we asked "When attending meetings, do you find it offensive when someone is using a laptop, Blackberry or I- phone for their personal use?" More than half of all respondents indicated they agreed or strongly agreed. To address Research Question 8A - Are there generational differences in the acceptance of, or participation in, multi-tasking activities during meetings? There was a significant difference by generation, with more Generation Y individuals who agree or strongly agree (chi-square <.00) that they had trouble paying attention

to others around them when they texted. This difference is also found when asking respondents whether they have had a cell phone conversation (not texting) during a meeting. Although done by less than 1/3 of individuals, there were significant generational differences (chi-square <.01), with nearly all Baby Boomers indicating rarely or never, and 15 percent of Generation Y indicating always or sometimes. Most respondents, 86 percent, agree or strongly agree that it is rude to talk on a cell phone during a meeting. We also asked whether meeting attendees read email or texts in meetings that are unrelated to the topics at hand. More than half of the respondents sometimes do so. There was a significant difference (chi-square <.00) among generations with twice as many Gen Y individuals indicating sometimes and twice as many Baby Boomers indicating rarely. Some respondents indicated that they text other meeting participants. There was a significant difference between generations (chi-square <.00), with nearly half of Gen Y responding sometimes. More than half of Baby Boomers responded never, but less than 10 percent of Gen Y were in this category. The size of the meeting group does not seem to affect the perception of those using mobile devices to do personal work, with twice as many respondents indicating that it is not acceptable to do this even in a large meeting. There was no significant difference among generations.

In order to answer Research Question 9 - Is the use of personal technology while talking to others is considered to be rude?, and Research Question 9A - Is there any generational difference in whether use of personal technology while talking to others is considered to be rude, we asked respondents four questions. The first was whether they were offended when presenting an important issue to individuals who then did not pay attention to them. More than 75 percent of respondents agreed or strongly agreed with this statement. More than half of Generation Y respondents, however, were neutral about the behavior, significantly different from the Baby Boomers (chi-square <.00). The next question was whether the distraction of a personal device used by one individual when speaking to another would be considered rude. The responses were similar, with more than half indicating that this would be considered rude. There was a significant generational difference here as well, with more Generation Y individuals having no strong response either positively or negatively and approximately 20 percent disagreeing (chi-square <.00). The third question, asking whether relationships with co-workers suffered when these individuals responded to personal devices during face-to-face meetings, yielded no significant result for the responses to the question. Finally, respondents were asked whether they considered the use of mobile devices during personal meetings and conversations to be a sign of the increased level of rude behavior in the workplace. Again, there was no strong sense of agreement. Although approximately half agree or strongly agree with this statement, one-third reported they were neutral.

For Research Question 10 - Does more intra-organizational communication take place via email versus face-to-face? In this case, respondents were equally divided in their responses among agree, neutral, and disagree. Research Question 10A: Are there generational differences in use of email versus face-to-face communication is used in organizations? – There was no significant generational difference in these responses.

DISCUSSION AND CONCLUSION

Perhaps the design of communications technology systems in organizations needs to become a conscious process. More often than not, these systems have evolved without direction and have led to deterioration of organizational communications. Gorman [3] contends that face-to-face exchanges should be the primary mode of communication in the community, with technological communication being secondary. One way of approaching this design is to consider whether a communication is a relationship-building or a process exchange [3].

These results indicate that employers will need to consider generational differences in the areas of personal productivity, organizational productivity, and perceptions of mobile device use in the workplace.

Personal Productivity

According to these results, individuals generally felt that their personal productivity is increased by the use of mobile devices. The ability to access email at times and places convenient to the worker provides a means of accomplishing tasks during more hours of the day. Research can be completed more promptly and efficiently, allowing for better decision-making. These advantages were not generation-specific; therefore, employers appear to be gaining from increased productivity of all workers.

It was interesting to note that although individuals felt that their own productivity is enhanced by the use of personal technology, its use by others in their presence is seen as a distraction. This finding also did not differ by generation. It seems to follow that the use of these devices during meetings will contribute to a loss of personal productivity due to the distraction of multiple person use. Twice as many Gen Y as Baby Boomers reported multitasking during meetings. The perception of this level of activity as the norm by Gen Y workers is not shared by Baby Boomers who view it as a way to absent oneself mentally from organizational activities.

Negative perceptions are shared by each generation with regard to the effect on personal relationships from mobile technology use, no matter the age of the worker. Since workplace relationships enable the accomplishment of individual and team tasks, the damage done to them from technology use would impede work progress. Having awareness of the effects of one's use of devices during personal encounters and reducing/eliminating this activity should enhance relationships and assist in increasing personal productivity.

Organizational Productivity

Although there seems to be a perception provided by the media that mobile devices unequivocally contribute to productivity at all times, respondents in this study did not agree. In fact, the sense from these results is that mobile devices lead to lower productivity in meetings as well as longer meetings. As noted earlier, this could be due to the distraction experienced by individuals when others are using technology alongside them. There are generational differences in participation levels for this meeting behavior with Gen Y individuals more likely to text someone else at the meeting and to access irrelevant internet sites. Conversely, respondents

noted that the using mobile devices could actually enhance the speed at which additional resources could be referenced. This could lend itself to quickly answered questions or concerns, enhancing the overall effectiveness of the meeting. It is possible that individuals' perception of meeting length was affected by personal bias because participants did not feel that they themselves were any less productive, but somehow the meeting as a whole was. Overall, the survey results indicated that individuals in meetings were quick to blame others or disapprove of their technology use, but did not hold themselves to the same standard.

Generational differences did surface when examining whether organizations should try to address the productivity concerns by implementing policies to control employees' use of personal devices during meetings. Both Gen Y and Baby Boomers agreed that policies for device use should be provided for meeting participants.

At the organizational level, more than 50 percent of individuals felt that policies should exist, but few companies actually have them. A greater degree of control is desired by Baby Boomers. Generation Y respondents felt that use issues should be dealt with on a case-by-case basis or ignored entirely. Corporations need to understand that the influx of Gen Y employees is accustomed to using technology in everyday practices. These individuals are not used to having their access to devices limited and seem not to subscribe to any particular decorum. It may be necessary to create certain cultural expectations in order to tame their technology addiction. Gen Y respondents seemed unwilling to admit or recognize such reliance, but it is evident through to day-to-day activity that those who fall into this generation rely heavily on mobile devices to get them through the day. These workers have grown up used to using technology as they please, and when they reach the professional work force will be less willing to conform to organization, especially if they have not been established at the outset or, are laissez-fair. Scarcity theory posits that if people have something in abundance, and then they are introduced to an environment where that abundance is restricted, they will desperately seek that which is limited or gone.

Social Acceptance

The perception of the use of personal devices during meetings is more negative for Baby Boomers than for Gen Y individuals. Gen Y employees indicate that use could be helpful for increasing personal productivity; therefore, trying to control it by general policies is not desirable. The view that multitasking positively impacts employee productivity varies by generation as well. Gen Y workers' opinions were more widely distributed among responses with no clear consensus of agreement or disagreement. Baby Boomers have a negative view of those who use meeting time to accomplish personal work. They prefer that all meeting participants interact with the team on a personal basis in order to better establish the team's relationships. This generation also wants the respect that they feel is their due by virtue of their organizational level. Use of personal technology during these interactions is therefore viewed as personally disrespectful.

There were more Gen Y respondents who were willing to identify themselves as being addicted to personal technology. This level of use may help account for their sense that multitasking contributes to productivity. According to a report by the SHRM done in 2005, people ages 8-18

live "media saturated" lives.   More than a quarter of their 6.5 hours per day of media access involved the use of two or more technologies simultaneously.   Given this trend, which only appears to be growing, it is not surprising to find that multitasking is viewed as ordinary for this generation.

The major conclusion drawn from this research is that of the need for case-to-case determination when considering the use of personal technology in the workplace. Personal opinions were diverse and determined by situation. Upon examining the data, it became evident that people have difficulty making arbitrary decisions about appropriate mobile device use. They prefer to make decisions about their technology use according to the situation. This contingency approach is important to understanding the differences of opinion that exist between generations. Each has a different opinion about which "cases" or "situations" lend themselves to the use of technology and which do not. Perhaps the design of communications technology systems in organizations needs to become a conscious process.   More often than not, these systems have evolved without direction and have led to deterioration of organizational communications [3].

Given the increasing frequency of technology use in corporations, it is clear the time has come to for them to either establish policies, or maintain a laissez-faire attitude about its use. Corporations that choose to follow the policy route may find resistance upon its implementation because those who have become accustomed to using it will feel a greater loss than those who were never given or never took the opportunity in the first place. By implementing policies and establishing expectations now, however, proper etiquette will be created. Hopefully, a social norm for the appropriate use of such technology will develop for future generations.

REFERENCES

[1]Bernoff, J. *Social netiquette: Is it OK to use a phone in meetings?* Advertising Age, 2009. Retrieved from http://adage.com.

[2]Duffy, M. *Communication tools limit effective interaction if co-workers aren't as connected as they assume.* Nation's Restaurant News, 2008, 42, 20.

[3]Gorman, G.E. *Professional community networks.* Online Information Review, 2009, *33(1),* 5-6.

[4]Leigh, H. (2007). *The etiquette of laptops in meetings.* One louder. Retrieved from http://blogs.msdn.com

[5]Mann, H. (2010). *Study links handheld devices to poor workplace manners.* Society for Human Resource Management. Retrieved from www.Shrm.org/hrdisciplines/technology/Articles/pages/TechManners.aspx

[6]Martin, C. (2007). *Productive communication by personal discussion.* CIO. Retrieved from http://www.cio.com/article/103100/Productive_Communication_by_Personal_Discussion

[7]May, E. (2009). *Laptops in meetings – Just don't.* Post on Practical Manager. Retrieved from http://practicalmanager.wordpress.com/category/tools/laptops

[8]Pansch, R. *With some key exceptions, yakking a cell phone in public deemed rude.* Tribune Business News, January 11, 2009.

[9]Schonfeld, E. (2009). *Survey says Baby Boomers think playing with your Blackberry during a meeting is rude.* Techcrunch. Retrieved from http://techcrunch.com/2009/04/15/survey-says-baby-boomers-think-playing-with-your-blackberry-during-a-meeting-is-rude

[10]Silverman, D. (2009). *Is smartphone use in meetings rude or righteous?* Techblog. Retrieved from http://blogs.chron.com/techblog/archives/2009/06

[11]Stafford, J. *With more technology comes more ways to be rude.* Knight Ridder Tribune Business News, November 28, 2004.

[12]Tung, A. *Cell-phone users hanging up on etiquette.* Tribune Business News, June 29, 2009.

[13]Weinstein, B. (2009). *Dear Bev: Is it rude to use your Blackberry in a business meeting? MediaPost News.* Retrieved from www.mediapost.com/publications

THE "CRACKBERRY" USER – ANTICIPATING THE NEXT WAVE

Diane M. Harvey, Roger Williams University, 1 Old Ferry Rd., Bristol, RI 02809, (401)254-3018, dharvey@rwu.edu

Susan M. Bosco, Roger Williams University, 1 Old Ferry Rd., Bristol, RI 02809, (401)254-3175, sbosco@rwu.edu

ABSTRACT

The use of Blackberry devices has grown significantly since their introduction in 1999. They are now viewed as much more than a combination cell phone, email device, and planner. As executives increasingly incorporate the Blackberry into their work and personal lives, it has become known as the "crackberry", a glimpse into its potentially addictive nature. Today's college students, tomorrow's executives, are already heavy users of these devices. This pilot study sought to determine current levels of use of this technology, especially texting, among this population. Results indicate a more balanced perception of its use in the workplace than expected.

Key Words: technology, addiction, liability

INTRODUCTION

The term "crackberry" has a dual identity – on one hand, there is an ominous sense of the addictive nature of the type of technology represented by the term, but on the other, there is an almost comedic image that comes to mind of a technology user who lacks self control. There has been a continuing examination of the uses and abuses of this sophisticated mobile personal technology, which allows users to not only make calls, but to check email, surf the net, and accomplish other personal and work management tasks. While there are advantages in productivity that can be achieved through its use, there is an emerging awareness that employees can become addicted to these devices to the detriment of both their professional and personal lives.

What will happen to the usage patterns of this technology when current college students enter the workplace in the next several years? It is evident that this segment of the population is closely tied to its personal technologies – iPhones, iPods, and cell phones. Are these individuals prepared for the management of their devices in a professional environment?

The purpose of this pilot study was to determine the extent of use of mobile devices by college students. In addition, we wanted to gain a sense of their level of addiction to these devices. If students are already interacting with them in a dysfunctional manner, this indicates a need for some type of intervention during their business education experience.

CURRENT WORKPLACE TRENDS

Many businesses use and issue mobile devices for on-call work assignments, and people use them to organize their lives [2]. In a study at the Sloan School of Management at MIT, it was found that managers are under extreme pressure to be reachable around the clock [6]. Another study by Osterman Research had similar findings [11]. With email becoming so highly integrated into mission-critical business processes, employees are feeling extraordinary pressure to be constantly available. There is more addiction to accessing email during non-work hours, illustrating the important role it has taken in the business workflow [7]. This study found that employees rely so heavily on mobile email availability that if service went down, even for an hour, 85 percent of respondents indicated that it would negatively impact their business [7]. According to Jackson [8], these trends actually undermine the efficiency and connectivity that devices are supposed to create. The loss in efficiency is due to the attempt to live in two moments at once. Processes become interruption driven, resulting in fragmented time with others and the ability to work on only small pieces of projects. One-fourth of 900 respondents in the survey felt they were on a "permanent corporate leash" [8].

There is not widespread agreement about this point, however. Psychiatrists have not made a final determination as to whether this activity is a true addiction [5]. A spokesman for the manufacturer of BB dismissed suggestions that BB are addictive. He claims that by using Blackberrys, workers gain productive time otherwise wasted in transit or waiting for meetings. These devices help us to achieve more, according to James Hart, Vice President of Research in Motion, and creator of the Blackberry [6].

Technology is not only necessary for most regular, business functions; devices such as BlackBerrys are now the status symbol for bosses and ambitious office staff [6]. It is common for face-to-face conversations to be interrupted by a call or text for one party. In addition, meeting attendees are not reluctant to spend their time texting or checking and sending emails.

Addiction to Technology – The Crackberry Phenomenon

The signs are similar to that of any sort of dependence on a drug [2]. According to one user, quitting smoking was easier than quitting BB use. One man feels "phantom vibrations" from the device even when he isn't wearing it [12]. Ringxiety is a state of readiness that develops in cell phone users. These individuals hear others' phones or similar sounds and reach for their phone. Those who experience this phenomenon most often, 67 percent of survey population, also used the phone the most [1]. Other symptoms of Blackberry addiction exhibited by users include denial, withdrawal, and antisocial behavior. The constant checking for emails and text occurs even in the presence of family [6] [14]. Gayle Porter, of Rutgers University School of Business, agrees that the fast pace of technology used to enhance work can become addictive. Employers may face liability for these addictions [3].

There are signs of this addiction that employers should be aware of in order to prevent negative outcomes. One in eight Americans exhibit at least one possible sign of problematic Internet use, according to a 2006 study by the Stanford University School of Medicine in Stanford, California [5]. Some 14 percent of the respondents reported that it was hard to stay away from the Internet

for a several-day stretch. More than 12 percent said they stayed online longer than intended, and nearly 9 percent said they hid their Internet use from loved ones or employers. Respondents had a strong drive to compulsively check e-mail, make blog entries or visit web sites or chat rooms, which is similar to what sufferers of substance abuse or impulse-control disorders experience: A repetitive, intrusive and irresistible urge to perform an act that may be pleasurable in the moment but that can lead to significant problems on personal and professional levels [5].

Symptoms that employers should monitor include: Attempts to hide their use, more use of the technology than is required for one's job, missing deadlines, losing interest in hobbies, losing sleep, preferring to be online, lack of interpersonal skills [5].

Recent articles suggest that employers may be liable for incidents arising from excessive device use. Lawsuits are a growing occurrence for employers who are sued for failing in their duty of care to staff and in following health and safety guidelines [12]. In one case in the US, a business consultant claimed her marriage ended because she was constantly checking messages. She also lost custody of her children. She sued her employer for damages and the company settled out of court [12]. Another employer had to pay substantial damages to a woman so distracted by her Blackberry while driving that she killed a motorcyclist [12]. A Chicago police officer is suing the city for two years of overtime pay for time spent on his Blackberry after work hours [10].

Employers provide help to workers with chemical or substance addictions, states Porter, but addiction to technology can be equally damaging to the mental health of the workers. Psychological and physical stress can result as well [14]. Workplace stress can cause loss of productivity, increased absenteeism, and violence. Physical problems can also result, for example, there are instances of employees developing repetitive stress injuries from excessive Blackberry use. Some of these require surgery to correct [10]. The growing evidence of Internet addiction has led Porter to conclude that Human Resources areas should rethink their approach to Internet abuse. One option to remedy the problem that is worth examining is rehabilitation, if warranted, rather than termination, not only because of the legal ramifications but for the practical benefits [14]. Regardless, current research and empirical evidence from employee assistance program (EAP) providers suggest that employees consumed by the Internet can get help, if managers catch the warning signs early [5].

BUSINESS STUDENTS' USE OF MOBILE DEVICES

Considering that employers are encountering increasing issues with mobile device use and misuse at the workplace, what can be expected from future employees? The current college population is the generation most closely tied to technology. However, it appears that they will be challenged by the younger teenage population and their heavy use of technology, particularly, texting. These teenagers average 100 text messages per day [10] [13]. Another source states that 3200 texts each month are the norm [4]. The number of teens texting daily increased from 38 percent in February, 2008 to 54 percent in September, 2009 [13]. The same signs of addiction experienced by the employees are also evident in teens. Neuroimaging studies show that teens who are texting have brain images that are the same as those of an addict on heroin [4].

Because texting is the communication mode that has experienced the most growth in the last few years, this pilot study examined that behavior in college students. Concerns about technology-related class management problems by business educators are also increasingly focused on texting during class time. In order to begin to determine whether or how educators can prepare students for appropriate workplace use of these types of technologies, it is important to learn the nature of their current usage patterns. In this study, some of the questions were about the use of texting, such as frequency and context. Others inquired about workplace expectations. Finally, students were asked whether they should be prepared in the course of their business education for appropriate technology use in the workplace. This pilot study was conducted to begin to examine these questions in an empirical manner.

METHODOLOGY AND RESULTS

All business students in a capstone course were asked to complete a short anonymous survey about their use of mobile technology. There were 73 respondents, all seniors, 60 percent were males. The focus of the survey was on texting, since this activity has been discussed anecdotally in recent articles on technology use/overuse [6] [14].

Students were asked how often they texted during the course of a day, whether they text during class, during work, and with what frequency. They were asked the ratio of use of texting for school vs. personal communication. There were two open-ended questions, the first being their opinion of texting the workplace and the second being whether professors should take any action about texting in class and whether any preparation should be provided for students regarding workplace use.

Many respondents, 41 percent, estimated that they text less than one hour per day. Another 33 percent indicated that they text between 2 and 4 hours per day. There was a significant difference in responses by gender (chi-square <.02) with more males texting between 2 and 4 hours per day and more females texting less than 1 hour per day. Twenty percent of the total said they text constantly. More texting occurred during work hours than class hours with 73 percent texting during class time compared to 88 percent who texted during work. Most, 78 percent, said they are not addicted to texting. The percent of text use for school versus personal communication was significantly different with 17 percent for school vs. 82 percent for personal communication (F=7.588, $p<.00$).

To examine the response to the open-ended question asking their opinion of texting and the workplace, we performed a content analysis. We first categorized general opinions into whether texting was acceptable versus unacceptable. As seen in the literature, there was no clear consensus, with 25 comments being positive and 15 negative. A sample of the comments in each category is provided in Table 1.

TABLE 1

Opinions on texting in the workplace

Positive	Negative
Texting should be allowed as long as it is not detrimental to the work that needs to be done	Texting is absolutely a distraction.
As long as it doesn't negatively affect your work	When you are working you should not be texting
It's a good way to communicate without taking 5-10 minutes to make a phone call	Takes away your focus from working
It is an essential component for communication in the business world	It is rude and inappropriate. You should be focused on the job that you are getting paid for
If it doesn't affect the effectiveness or efficiency of the employee then it is fine	Should be frowned upon since it is a distraction

Most of the positive comments included concerns about ensuring that workflow was not interrupted and that productivity and efficiency were not negatively impacted. The respondents exhibited awareness that texting could be distracting and cause loss of focus. In addition, some mentioned the office atmosphere or job duties as limiting factors. For example, if the office culture is casual or if the person works alone or in a cubicle where he/she is more isolated from others at the workplace. Limiting factors for texting were the need to interact personally with other workers or customers and operating machinery or equipment.

Negative comments included concerns about poor etiquette and effect on personal interactions. Texting while talking to someone else was cited as an example of rude workplace behavior. Another concern was the appearance of unprofessionalism that texting could give to others. The consequences of loss of focus and increased distraction on the quality and/or quantity of work produced were considered other undesirable outcomes to texting.

Responses to the open-ended question about what faculty should do about making students aware of the use of texting in the workplace were also reviewed using content analysis. Only 5 students responded that faculty should not take any action in this regard. Many more, 21, indicated that they would welcome some instruction or guidance about texting in a professional environment. Most of these concerns were focused on the appropriate use of texting, such as, when it would/would not be considered acceptable. Some felt a business etiquette discussion or presentation would be helpful. Others would include this topic as part of general business communication instruction, to incorporate considerations like situational factors, use of formal language vs. slang, efficiency and effectiveness of the medium.

CONCLUSIONS AND RECOMMENDATIONS

From this initial pilot study, it appears that, although students text in both work and personal environments, most do not do so constantly. The mode for amount of time texting was less than one hour per day. Therefore, they are using this technology in a manner that provides balance to other activities in which they are involved. Very few students, only 22 percent, claimed to be

addicted to texting. Although this result could be the result of social response bias, the amount of time spent texting seems to support this relatively low number.

There is a large divide between the amount of texting occurring for school, which is in this case, their workplace, and that occurring for personal reasons. Very little texting, only an average of 17 percent, involves school-related issues. Therefore, the use of this technology has not yet become an essential element of their "work" day. The amount of texting that is done for personal communication is somewhat of a concern. Texting does not provide the means for anything but the most mundane of messages. The extent to which this mode of communication replaces, versus supplements, personal exchanges would be of interest to examine further. Since one of the complaints lodged in the workplace about technology use is the increasing tendency for workers to communicate through email and text rather than in person, indications from this survey are that this trend will continue, and probably increase, with the upcoming cadre of employees as well.

Awareness of the positive and negative aspects of texting in the workplace indicates an appreciation for the place of technology in an organizational environment. Most of the students as seniors have done professional internships and they exhibited a strong sense of the potential that texting has to be abused in the workplace. They were careful to qualify their responses about the use of texting with concern about making sure one's work was being completed in a timely and effective manner. It was also considered as just another tool that could help to accomplish work more quickly – saving time on phone calls was mentioned by several. Of course, this preference over using the phone could be a sign of less developed interpersonal skills. A major negative was that it was rude and unprofessional which does reflect some awareness of the personal impact of this technology. Concern over the element of distraction and lack of focus that use of this technology could introduce into the work environment was also expressed. The acknowledgement that certain types of work were better suited to allowing texting was evident in these responses as well.

In response to the question of what type of preparation students desire from their business education with regard to texting use in the professional world, the major area where guidance is sought is in the area of personal relations. When to use and not use texting was specifically mentioned. An additional concern the emerged was the general lack of information about communication channels and modes for business purposes. Therefore, it seems that some type of classroom preparation in the area of business discourse and specifically in technological tools in communication is desired.

In summary, the awareness that students showed about the positives and negatives of texting in the workplace was in contrast to the addictive use of mobile devices that appears to be currently occurring. Perhaps because this technology is not a novelty for the Gen Y generation, they are able to view it more critically. They are quick to acknowledge the ability of technology to help them to be more efficient; however, its weaknesses also seem apparent to them.

One reason for this difference could be due to their generational membership. Many individuals in the workplace today are from the Baby Boomer generation, which is known for having workaholic tendencies. In some of the articles we reviewed for this study, the relationship

between mobile device addiction and other workaholic behaviors was noted. Since the generation of current college students desires more work-life balance, they may not have as much of an inclination to be addicted to their workplace technology. Perhaps the ability of Gen Y individuals to already see both the positive and negative aspects of technology use reflects a reluctance to become as strongly tied to the workplace as older workers have become. Therefore, employers will possibly find themselves managing employees who seek limitations on expectations for technology use for work, rather than seek to expand its instrusion further into the rest of their lives. In a world where organizations are increasingly competitive, seeking the ability to respond at all times to customer needs, the reluctance of younger workers to accept this demand may generate some basis for conflict.

In contrast, the boundaries for appropriate personal use of texting are nearly nonexistent for this generation, with 73 percent texting during class time and 88 percent texting during work. They express an awareness of when texting is appropriate, but their actions indicate otherwise, since they are engaged in it regardless of the situation or place. College students need to be prepared for the perceptions that await them in the work environment with regard to the use of mobile devices. These perceptions are not generally positive; with the expectation by older workers that texting is an activity that reduces productivity. The Director of Career Services at Westminster College in Missouri, Meg Langland, trains students to prepare to limit their use of texting and email tools. She feels it is essential for them to plan for the new rules that will likely face them in the workplace [9]. Our findings support this recommendation that rules about technology use should be introduced and discussed to bring currency to business school education.

REFERENCES

[1]Clark, J. (2010) *Do you suffer from "Ringxiety"?* Retrieved from http://communication.howstuffworks.comringxiety.htm

[2]*Crackberry: Are you addicted to your BB?* (2010). Retrieved from http://www.printfriendly.com/get?url=http://hubpages.com/hub/Crackberry-Addiction

[3]*Employers could be liable for Crackberry addiction.* Cellular-News, 2010. Retrieved from http://www.cellular-news.com/story/18961.php

[4]Feyerick, D. & Garrett, D. (2010). *Can teen texting become an addiction?* Retrieved from http://amfix.blogs.cnn.com/2010/04/20/can-teen-texting-become-an-addiction/

[5]Fox, A. *Caught in the web*. HRMagazine, 2007, *52(12),* 35-39.

[6]Goodchild, S. & Hodgson, M. *Crackberry addicts: Why the workers who can't switch off are suing their employers.* The Independent, 2010. Retrieved from http://license.icopyright.net

[7]*Inappropriate behaviors: As email becomes part of the mission-critical business workflow, 95 percent check email after work hours; 76 percent admit to driving while texting.* PR Newswire, New York: February 17, 2010.

[8]Jackson, M. (2007). *Keeping work where it belongs.* Retrieved from www.boston.com

[9]Mann, H. *Study links handheld devices to poor workplace manners.* Society for Human Resource Management, 2010. Retrieved from www.Shrm.org/hrdisciplines/technology/Articles/pages/TechManners.aspx

[10]Miller, M. (2010). *When texting becomes an addiction.* Retrieved from www.cbsnews.com

[11]Osterman, M. (2008). *How frequently do you check mobile e-mail?* Retrieved from http://www.networkworld.com/newsletters/2008/081108msg1.html

[12]Rainey, M. (2007). *Crackberry addiction isn't pretty.* Retrieved from http://www.Bloggingstocks.com/2007/04/19/

[13]Schrock, R. (2010). *US: Texting: the teenage girls' addiction.* Retrieved from universityworldnews.com/

[14]Taylor, J. *Business: Work/Life balance: Part I.* Psychology Today, 2009. Retrieved from www.psychologytoday.com/print/32565

DISCUSSION OF PARABOLIC MASK
FOR CUMULATIVE SUM CONTROL CHARTS

Donald S. Holmes, Stochos Inc. P.O. Box 247, Duanesburg, N.Y. 12056.
(518) 895-2896, dsholmes@stochos.com
A. Erhan Mergen, Rochester Institute of Technology, Saunders College of Business
Decision Sciences, 107 Lomb Memorial Drive, Rochester, N.Y. 14623-5608.
(585) 475-6143, emergen@saunders.rit.edu

ABSTRACT

In this paper we will discuss the use of parabolic control limits for cumulative sum (Cusum) control charts and an approach to correct the average run length (ARL) problem that may result from using a "three sigma" parabolic mask. Result of a simulation study will be reported to correct this problem.

Keywords: Cumulative sum control charts, parabolic mask, average run length.

DISCUSSION

Cusum control charts are the result of earlier developments: sequential analysis and control charts. Cusum charts are an extension of the seminal work on control charts by Shewhart [19], who introduced the idea that statistical analysis, could be used to determine whether or not an ongoing process' quality level could be identified as being stable. Identifying causes of instability and eliminating them would result in a less variable and hence higher quality level. Cusum charts were introduced by Page [15, 16]. They were a natural follow up to the growth in interest in sequential analysis at the time. These charts became known as Cusum Control Charts.

Shewhart's approach was to use charts for both the process center and the process width, using what he referred to as "Control Charts": an X-bar chart (to monitor the center of the distribution of measurements generated) and a Range or standard deviation chart (to monitor the width of the distribution of process measurements).

Page's Cusum control charts introduced a somewhat different idea. The "control limits" for the process property being plotted were neither fixed in location nor parallel straight lines. The limits consisted of a *"V-shaped mask"* that was always positioned relative to the most recently acquired data point. The placing of the *V-mask* on the latest data point provided an answer to a different question: "Is the process now doing about the same as it had been doing at previous points in time?" For example, a point outside of the V mask at a point 20 time periods previous to the current point would indicate that the process is currently further from the target than it had been 20 time periods ago.

Since its introduction by Page, academicians and practitioners have been researching different aspects of these charts. These control charts display real strength in detecting small shifts in the

average or trend conditions (see, for example, Goldsmith and Whitfield [7], Page [16], Lucas [13, 14], Hawkins [8], Woodall [21], Gan [5]). Economic design of Cusum charts were also studied by Chiu [3], Goel and Wu [6], Taylor [20] and others. Some authors, for example Atienza, et al. [1] and Lu and Reynolds [11], looked into using Cusum in autocorrelated processes. For other uses of Cusum charts, such as monitoring a proportion and process variance, see, for example, Reynolds and Stoumbos [18], Chang and Gan [2] and Page [17].

The logic behind the Cusum chart can be explained with a simple example. Suppose the following six measurements were made on the output of a process: $X = 4.1$, $X_2 = 4.0$, $X_3 = 4.1$, $X_4 = 4.2$, $X_5 = 4.3$, $X_6 = 4.3$. Further suppose that the goal value (k) is the nominal value of 4.2. The values to be plotted on the Cusum control chart would be obtained as shown in Table 1 below:

Measurement Number, i	Measurement Value, X_i	Deviation from Goal, $X_i - k$	Cumulative Sum of Deviations, $\Sigma X_i - k$
1	4.1	4.1 – 4.2 = -0.1	-0.1
2	4.0	4.0 – 4.2 = -0.2	-0.3
3	4.1	4.1 – 4.2 = -0.1	-0.4
4	4.2	4.2 – 4.2 = 0	-0.4
5	4.3	4.3 – 4.2 = 0.1	-0.3
6	4.3	4.3 – 4.2 = 0.1	-0.2

Table 1. Cumulative sum example

This example shows the nature of the Cusum chart: If the process is below the goal, the cumulative sums decrease and the graph shows a downward trend. This is true for points 1, 2, and 3. If the process is on goal, the cumulative sums neither increase nor decrease and the graph remains horizontal. This is true for point 4. If the process is above the goal, the cumulative sums increase and the graph shows an upward trend. This is true for points 5 and 6.

Thus, control limits should be built to detect excess variation (i.e., out-of-control situations). The control limits (V-shaped, also known as "V mask") for the Cusum charts were initially described in terms of the two parameters d and Ө (Page [15]) as it is diagrammed below in Figure 1.
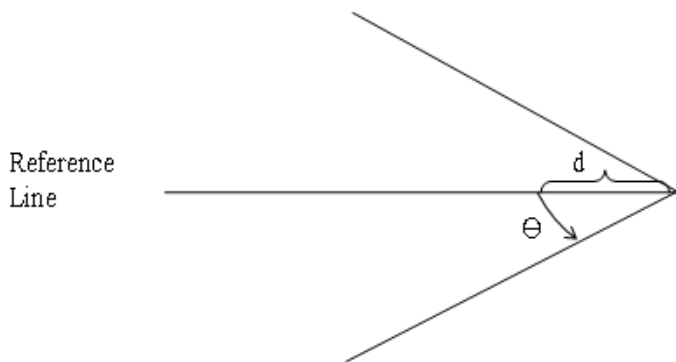
Figure 1. V-Mask

The point which is d units to the left of the vertex is placed on the cumulative point representing the sum of the deviation from the average (or from the goal that is of interest). Ө is the angle between each of the limits of the V shaped limits and the horizontal line. The reference line is placed in the horizontal position. Points outside the V shaped limits indicate that the process is out of control.

Parabolic mask:
Holmes [9] proposed use of parabolic limits on the Cusum charts. Lucas [12, 13] developed "a modified V-mask," which adds a parabolic section to the V-mask. Lucas [12] reported that parabolic-shape mask would work better than a V-mask for detecting large changes quickly. Later Holmes and Mergen [10] developed an average value scale for the Cusum charts which enables the user to read the average value of the process deviation from the average (or goal).
Holmes [9] used a 3-standard deviation approach to define the parabolic mask. Lucas, on the other hand, used a risk level approach to define the mask. When using the parabolic mask, the vertex is placed on the most recent point and any points to the left of the vertex which fall outside the mask are taken as indications that the process is out of control. Figure 2 displays a Cusum chart with parabolic limits. Cumulative sums outside the parabolic mask indicate that they are significantly different than the most recent cumulative sum value.

In this paper we will address an issue related to average run length (ARL) when the parabolic mask is used using our prior works on Cusum charts. We deal with the case where the Cusum charts are built based on a capability standard deviation and used for checking the stability of a process. Holmes [9] proposed the following to determine the parabolic limits for Cusum:

$$\pm 3\sigma_x \sqrt{n} \text{ or } \pm 3\sigma_x n^{0.5} \tag{1}$$

where σ_x is the capability standard deviation and n is the number of periods prior to the point where the mask is placed. The 3 sigma mask proposed by Holmes [9] and Holmes and Mergen [10] above uses the square root of the number of data points behind mask in calculating the control limits. It has been noted by some statisticians that the 3 sigma mask has shorter average run lengths (ARLs) than comparable control charts when there has been no process change. That

is to say, the alpha risk is greater for the 3 sigma parabolic mask than it is for X-bar charts using 3 sigma limits.
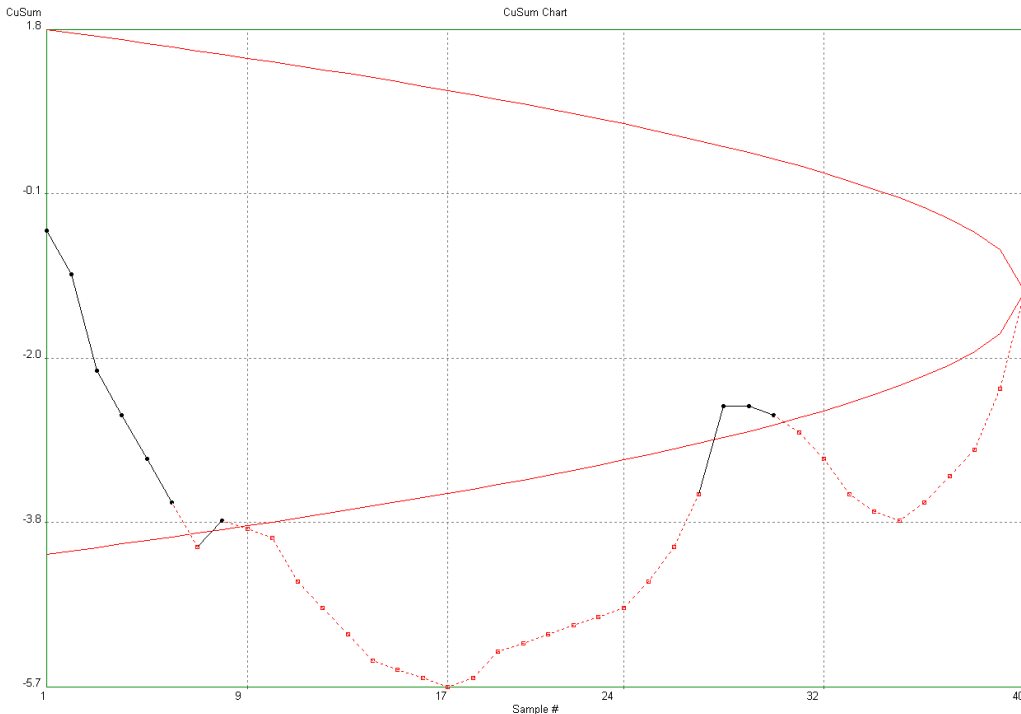
Figure 2. Cusum chart with parabolic limits

This issue has several potential sources:
a. The successive sums used in Cusum control charts are, by their nature, positively autocorrelated. This leads to the need to reduce the width of the control limits, i.e., the effect of the autocorrelation is to reduce the width of the limits (Deligonul and Mergen [4].

b. The use of the limits, using either the V-mask or the parabolic mask, requires the comparison of a finite number of past observations. This leads to the need to increase the width of the limits as points at the wider end of the mask are evaluated.

Thus we made a large number of simulation runs at various roots n greater than 0.5. The value of the root which produced ARL of about the same as those for a 3 sigma X-bar chart (i.e., i.e., type I error of about 0.0028 when the process is in control) turned out to be 0.66 (Computer program is available from the authors upon request.) This program gives the number of out-of-control signals out of 3000 simulation trials when the process is in control (i.e., number of false signals). Average of several 3000 simulation trials yields type I error (subject to minor fluctuation given the random number seeds selected). We therefore suggest that, should you elect to use a parabolic mask for your Cusum control chart, the limits calculated using equation (1) be modified as shown in equation (2):

$$\pm 3\sigma_x n^{0.66} \tag{2}$$

CONCLUSION

In this paper we discussed an issue regarding the ARL of Cusum chart when parabolic mask is used. We believe that parabolic mask is easier to build and understand for the user than the V-mask since it uses the usual approach of multiples of the standard deviation. Practitioners should consider the use of Cusum chart given its power of detecting small changes in the process. Our next topic for research would be to check the ARL of the Cusum chart with parabolic limits for various shift conditions in the mean.

REFERENCES

[1]   Atienza, O.O., Tang, L.C. & Ang, B.W. *A Cusum Scheme for Autocorrelated Observations*. Journal of Quality Technology, 2002, 34(2), 187-199.

[2]   Chang, T.C. & Gan, F.F. *A Cumulative Sum Control Chart for Monitoring Process Variance*. Journal of Quality Technology, 1995, 27(2), 109-119.

[3]   Chiu, W.K. *The Economic Design of Cusum Charts for Controlling Normal Means*. Applied Statistics, 1974, 28(3), 420-433.

[4]   Deligonul, Z.S. & Mergen, A.E. *Dependence Bias in Conventional p-charts and Its Correction with an Approximate Lot Quality Distribution*. Journal of Applied Statistics, 1987, 14(1), 75-81.

[5]   Gan, F.F. *An Optimal Design of Cusum Quality Control Charts*. Journal of Quality Technology, 1991, 23(4), 279-286.

[6]   Goel, A.L. & Wu, S.M. *Economically Optimum Design of Cusum Charts*. Management Science, 1973, 19(11), 1271-1282.

[7]   Goldsmith, P.L. & Whitfield, H. Average Run Lengths in Cumulative Chart Quality Control Schemes. Technometrics, 1961, 3, 11-20.

[8]   Hawkins, D.M. *A Fast Accurate Approximation for Average Run Lengths of Cusum Control Charts*. Journal of Quality Technology, 1992, 24(1), 37-43.

[9]   Holmes, D.S. *V-Notch Control*. ASQC Convention Transactions, Cincinnati, OH, 1962, 263-267.

[10]   Holmes, D.S. & Mergen, A.E. *An Average Value Scale for Cumulative Sum Controls Charts*. Northeast Decision Sciences Institute Proceedings (formerly Northeast American Institute for Decision Sciences), Williamsburg, VA, 1986, 286-287.

[11]   Lu, C.W. & Reynolds, M.R. *Cusum Charts for Monitoring an Autocorrelated Process*. Journal of Quality Technology, 2001, 33(3), 316-334.

[12]   Lucas, J.M. *A Modified "V" Mask Control Scheme*. Technometrics, 1973, 15(4), 833-847.

[13]   Lucas, J.M. *The Design and Use of V-Mask Control Schemes*. Journal of Quality Technology, 1976, 8(1), 1-12.

[14]   Lucas, J.M. *Combined Shewhart-Cusum Quality Control Schemes*. Journal of Quality Technology, 1982, 14(2), 51-59.

[15]   Page, E.S. *Continuous Inspection Schemes*. Biometrika, 1954, 41(1/2), 100-115.

[16]   Page, E.S. *Cumulative Sum Charts*. Technometrics, 1961, 3(1), 1-9.

[17]   Page, E.S. *Controlling the Standard Deviation by Cusums and Warning Lines*. Technometrics, 1963, 6, 307-316.

[18]   Reynolds, M.R. & Stoumbos, Z.G. *A Cusum Chart for Monitoring a Proportion When Inspecting Continuously*. Journal of Quality Technology, 1999, 31(1), 87-108.

[19]   Shewhart, W.A. Economic Control of Quality of Manufactured Product. D. Van Nostrond, New York, 1931, page 299.

[20]   Taylor, H.M. *The Economic Design of Cumulative Sum Control Charts*. Technometrics, 1968, 10(3), 479-488.

[21]   Woodall, W.H. *The Design of Cusum Quality Control Charts*. Journal of Quality Technology, 1986, 18(2), 99-102.

BETA-GEOMETRIC MODEL FOR ESTIMATING THE PRODUCT QUALITY IN THE REPETITIVE INSPECTION PROCEDURE

Young H. Chun and **Edward F. Watson**
E. J. Ourso College of Business, Louisiana State University, Baton Rouge, LA 70803
chun@lsu.edu and *ewatson@lsu.edu*; (225) 578-2506

ABSTRACT

A complex product such as a software document is often inspected more than once in a sequential manner to detect more "faults". For each fault in the software document, the probability that it will be detected during each review cycle is usually assumed to be an unknown "constant" that should be estimated as well. In many practical situations, however, some faults are easily detected, while others are much more difficult to be found. In the paper, we propose a "beta-geometric" inspection model in which the heterogeneity in detection probability is described by a beta distribution. In a numerical study, we show that our beta-geometric model clearly outperforms traditional estimation methods that are based on the assumption of the constant detection probability.

Keywords: Inspection; Product quality; Reliability; Maximum likelihood estimation

1. INTRODUCTION

Although inspection is one of the important and effective tools that serve the task of assuring product quality, "inspection error" is inevitable in any inspection process. Due to human error and technical problem, not every defect (or non-conformity) is detected during a round of inspection cycle. That is why some complex products (e.g., software system) are inspected multiple times in a sequential manner in order to improve the outgoing quality.

Such a multiple inspection plan is also known as a sequential defect removal sampling [2], a repetitive testing [11], a repeat inspection [10], or a sequential review [14]. After a given number of multiple inspections, we need to estimate the number of defects still remaining in the product. Conversely, we also want to determine the number of multiple inspections that ought to be conducted to achieve a certain level of outgoing quality [5].

As an example of the repetitive inspection, consider a software system that contains an unknown number N of "faults". A complete test and correction cycle is referred to as a "review" as in Rallis and Lansdowne [14], and the software system will be reviewed more than once in a sequential manner. For each fault, the "detection probability" is p, which is the probability that the fault will be detected, if not detected earlier, during the current review cycle. After each review, the number of faults x_i detected during the review cycle i is recorded, and those faults are removed or corrected prior to the next review. After a series of k independent reviews, we have a record $\mathbf{x} = \{x_1, x_2, \ldots, x_k\}$ of the number of faults detected and corrected during each review cycle.

Based on the inspection results $\mathbf{x} = \{x_1, x_2, \ldots, x_k\}$, we need to estimate the total number of faults N or, equivalently, the number of faults R_k ($=N- x_1- x_2- \ldots -x_k$) still remaining in the software

system. Assuming that the detection probability p is a known constant, Rallis and Lansdowne [14] treated N as a Poisson random variable. As a prior distribution of N, Chun [5] proposed a negative binomial distribution and developed a Bayesian model that determines the number of inspections needed to achieve a certain level of product quality.

Assuming that both N and p are unknown constants, Chun [8] proposed the maximum likelihood estimators. Bonnett and Woodward [2] also treated both N and p as unknown parameters that can be estimated by a non-linear regression method. Recently, Chun [6] pointed out a non-convergence problem of the non-linear regression method, and proposed a Gibbs sampling method that can find the point and interval estimates of both N and p by a Monte Carlo simulation.

In the aforementioned articles, however, the detection probability p of a fault is assumed to be (i) a known constant that is given *a priori* or (ii) an unknown constant that ought to be estimated. In many practical situations, however, each fault has a different probability of being detected; some faults can be found easily, while others are much more difficult to be detected.

The purpose of this article is to propose an improved inspection model that considers the "heterogeneity" in detection probability p. Specifically, we assume that the detection probability p is distributed as a beta distribution with parameters a and b. By changing its beta parameters, we can describe a wide variety of distributions with different shapes and scales. In a numerical analysis, we show that our "beta-geometric" model clearly outperforms traditional estimation methods such as (i) the maximum likelihood method and (ii) the conditional maximum likelihood method.

In the next section, we define several terms and formally introduce the notation that will be used throughout the paper. We first introduce the method of maximum likelihood in Section 3 and propose the conditional maximum likelihood estimator in Section 4. Assuming that the detection probability follows a beta distribution, we develop a beta-geometric model in Section 5. In Section 6, we evaluate their performances in a Monte Carlo simulation, followed by concluding remarks in Section 7.

2. PRELIMINARIES

Suppose that there are N faults in a software document, where $N = 0, 1, \ldots, \infty$. During a round of review, each fault is discovered with probability p and not detected with probability q ($=1-p$). In search theory, q is known as the "overlook" or "wink" probability [1]. We first assume that the total number of defects N and the detection probability p are unknown constants that should be estimated. Later in the beta-geometric model, non-homogeneous detection probabilities are described by a beta distribution with parameters a and b. In such a case, the beta parameters should be estimated along with N.

The software document will be reviewed k times by a software engineer. These reviews are "sequential" in the sense that a fault discovered during one round of review is removed or corrected prior to the next review so that the same fault cannot be detected more than once [14]. The sequential reviews are also statistically independent with each other so that the detection probability p (or its probability distribution) remains unchanged throughout the review process.

The sequential model should be contrasted with the "parallel" review model, in which the same copies of software documents are reviewed by several inspectors at the same time independently. In such a case, some faults detected by one inspector could also be discovered by other inspectors. Based on those inspection results, we may be able to estimate the number of faults

still remaining in the software document. One of the well-known parallel review models is the "capture-recapture" estimation method. Readers are referred to Briand, El Emam, Freimut, and Laitenberger [4] and Chun [7] for a comprehensive analysis of capture-recapture models in the context of software inspections.

As shown in the tree diagram in Figure 1, let x_i be the number of faults detected and corrected during the ith review. Then, the inspection history during the k reviews is summarized by $\mathbf{x}=\{x_1, x_2,\ldots, x_k\}$. In most practical situations, the histogram of x_i over i is skewed to the right, showing a longer tail dwindling over the review cycle i. For notional convenience, let s_k ($= x_1 + x_2 +\ldots + x_k$) be the *cumulative* number of faults discovered during the first k reviews. Furthermore, let R_k ($=N-s_k$) be the number of faults still remaining in the software document after the k reviews.
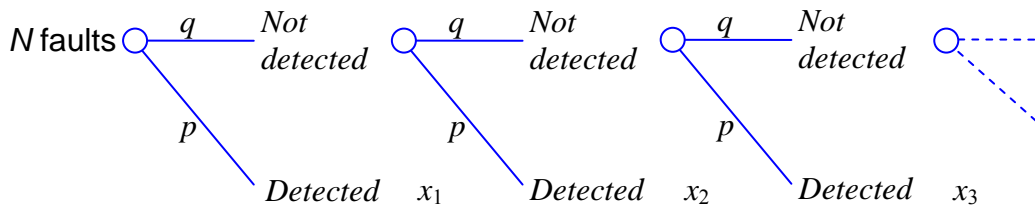
Figure 1. Repetitive inspection procedure

In the next section, we consider the maximum likelihood method that estimates N (or R_k) based solely on the inspection results $\mathbf{x}=\{x_1, x_2,\ldots, x_k\}$. In the traditional estimation method, we simply assume that the detection probability p is an unknown constant that should be also estimated along with N.

3. METHOD OF MAXIMUM LIKELIHOOD

At the beginning of the ith review, we still have $N-s_{i-1}$ faults remaining in the software document. Then, the number of faults x_i that will be discovered during the ith review follows a binomial distribution:

$$P[X_i = x_i \mid N, q] = \binom{N - s_{i-1}}{x_i} q^{N-s_i} (1-q)^{x_i}, \quad \text{for } i = 1, 2, \ldots, k, \tag{1}$$

where $q = 1-p$.

When the inspection results $\mathbf{x}=\{x_1, x_2,\ldots, x_k\}$ are available after k reviews, the likelihood function of N and q is expressed as follows:

$$L(N,q) = \prod_{i=1}^{k} P[X_i = x_i \mid N, q] = \frac{N!}{(N - s_k)! \prod_{i=1}^{k} x_i!} q^{\sum_{i=1}^{k}(N-s_i)}(1-q)^{s_k} \tag{2}$$

The maximum likelihood estimates (MLE) of N and q are the ones that maximize the likelihood function in (2). However, the optimal values of N and q which maximize the likelihood function $L(N, q)$ in (2) also maximize its log-likelihood function $\ln L(N, q)$. Therefore, rather than maximizing the likelihood function itself, it is more convenient to maximize its natural logarithm:

$$\ln L(N,q) = \ln \frac{N!}{(N-s_k)!} + \ln q \sum_{i=1}^{k}(N-s_i) + s_k \ln(1-q). \tag{3}$$

Note that

$$\frac{N!}{(N-s_k)!} = N(N-1)...(N-s_k+1) = \prod_{j=1}^{s_k}(N-s_k+j). \tag{4}$$

Thus, the log-likelihood function in (3) is expressed as

$$\ln L(N,q) = \sum_{j=1}^{s_k}\ln(N-s_k+j) + \ln q \sum_{i=1}^{k}(N-s_i) + s_k \ln(1-q). \tag{5}$$

The first-order derivative of the log-likelihood function in (5) with respect to q is

$$\frac{d}{dq}\ln L(N,q) = \frac{1}{q}\sum_{i=1}^{k}(N-s_i) - \frac{s_k}{1-q}. \tag{6}$$

Setting it equal to zero, we can derive the maximum likelihood estimator of q as follows:

$$\hat{q} = \frac{\sum_{i=1}^{k}(N-s_i)}{\sum_{i=1}^{k}(N-s_{i-1})}. \tag{7}$$

By plugging $\hat{q}$ in (7) into the log-likelihood function in (5), we can formulate the problem of finding the maximum likelihood estimate of N as a single-parameter maximization problem. Any iterative method, such as Newton-Raphson method, can easily find the optimal solution $\hat{N}$ that maximizes the log-likelihood function in (5).

4. CONDITIONAL MAXIMUM LIKELIHOOD ESTIMATOR

Still assuming that the detection probability p is an unknown *constant*, we consider in this section another traditional method of estimating N and p. The so-called "conditional maximum likelihood estimator" was originally proposed by Sanathanan [15], who used it to estimate the number of unknown trials in a multinomial probability distribution. We modify the estimation method for the repetitive inspection model and compare its performance later with those of other estimation methods.

For a certain fault in the software document, the probability that it will be detected and removed during the ith review follows a geometric distribution with parameter q:

$$P[i \mid q] = q^{i-1}(1-q). \tag{8}$$

From (8), the probability that the fault will not be discovered during the k review cycles and will be still remaining in the software document is

$$P[i > k \mid q] = q^k. \tag{9}$$

After k review cycles, the inspection results are $\mathbf{x}=\{x_1, x_2,..., x_k\}$ and $(N-s_k)$ faults are still remaining in the software document. Thus, it directly follows from (8) and (9) that the likelihood function of N and q is

$$L(N,q) = \frac{N!}{(N-s_k)!\prod_{i=1}^{k}x_i!}(q^k)^{N-s_k}\prod_{i=1}^{k}[q^{i-1}(1-q)]^{x_i}. \tag{10}$$

Note that this likelihood function is equivalent to the previous one in (2).

We found that the likelihood function in (10) can be divided into two separate parts as follows:

$$L(N,q) = \left[\frac{N!}{(N-s_k)!\,s_k!}(1-q^k)^{s_k}(q^k)^{N-s_k}\right] \times \left[\frac{s_k!}{\prod\limits_{i=1}^{k}x_i!}\prod_{i=1}^{k}\left[\frac{q^{i-1}(1-q)}{(1-q^k)}\right]^{x_i}\right]. \tag{11}$$

The first part is the likelihood based on the probability of s_k and the second part is the likelihood based on the conditional probability of x_1, x_2, …, and x_k given s_k. Since the first likelihood function in (11) is a binomial distribution,

$$L_1(N\,|\,q) = \left[\frac{N!}{(N-s_k)!\,s_k!}(1-q^k)^{s_k}(q^k)^{N-s_k}\right], \tag{12}$$

the maximum likelihood estimate of N conditional upon q is simply shown to be

$$\hat{N}(q) = \frac{s_k}{1-q^k}. \tag{13}$$

The second likelihood function in (11) is a multinomial distribution,

$$L_2(q) = \frac{s_k!}{\prod\limits_{i=1}^{k}x_i!}\prod_{i=1}^{k}\left[\frac{q^{i-1}(1-q)}{(1-q^k)}\right]^{x_i}, \tag{14}$$

which is independent of N. Thus, the maximum likelihood estimator of q is the one that maximizes the likelihood function in (14) or, equivalently, its log-likelihood function:

$$\ln L_2(q) = \sum_{i=1}^{k}x_i[(i-1)\ln q + \ln(1-q) - \ln(1-q^k)]. \tag{15}$$

From the first-order derivative of the log-likelihood function in (15), it can be easily shown that the maximum likelihood estimator $\hat{q}$ is the solution to the following equation:

$$\frac{\sum\limits_{i=1}^{k}s_i}{s_k} = \frac{\sum\limits_{i=1}^{k}(1-q^i)}{1-q^k}, \tag{16}$$

which can be further simplified as follows:

$$\frac{\sum\limits_{i=1}^{k}s_i}{s_k} = \frac{k}{1-q^k} - \frac{q}{1-q}. \tag{17}$$

After finding $\hat{q}$ that satisfied the equation in (17), we can easily obtain $\hat{N}$ from (13). Sanathanan [15] showed that the unconditional and conditional maximum likelihood estimates are asymptotically equivalent. As expected, the performances of those two estimation methods are shown to be very similar in our Monte Carlo simulation in Section 6.

5. BETA-GEOMETRIC MODEL

5.1. Beta distribution

In the traditional estimation methods, the detection probability p (=1-q) is assumed to be an unknown *constant*. To represent the heterogeneity in detection probabilities, we now assume

that the probability p of being detected during each review cycle is distributed as a beta distribution with parameters a and b:

$$f(p\,|\,a,b) = \frac{1}{B(a,b)}\, p^{a-1}\,(1-p)^{b-1}, \qquad \text{for } 0 < p < 1, \tag{18}$$

where the beta function is

$$B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}. \tag{19}$$

By changing the parameter values a and b in (18), we can represent a wide variety of variations in the detection probability p. If $a = b = 1$, for example, the beta distribution represents the standard uniform (or rectangular) distribution with equal probabilities over the range $(0, 1)$. If $a = 1$ and $b = 2$ (or $a = 2$ and $b = 1$), the beta distribution becomes a triangular distribution.

5.2. Likelihood function

For a certain fault in the product, the probability that the fault will be discovered and removed during the ith inspection cycle follows a *geometric* distribution with parameter p:

$$P[i\,|\,p] = (1-p)^{i-1}\, p. \tag{20}$$

Likewise, the probability that it will not be found during the first k inspection cycles and will be still remaining in the product is

$$P[i > k\,|\,p] = (1-p)^{k}. \tag{21}$$

From (18) and (20), the probability that the fault will be successfully discovered during the ith review cycle is a beta-geometric distribution as follows:

$$\begin{aligned}
P[i\,|\,a,b] &= \int_{p=0}^{1} P[i\,|\,p]\, f(p\,|\,a,b)\,dp \\
&= \frac{B(a+1, b+i-1)}{B(a,b)} \\
&= a\,\frac{\Gamma(b+i-1)}{\Gamma(b)}\,\frac{\Gamma(a+b)}{\Gamma(a+b+i)}.
\end{aligned} \tag{22}$$

Note that a gamma function in (22) has the following property: $\Gamma(c+1) = c\,\Gamma(c)$ for any constant c. Thus, it can be further simplified as

$$P[i\,|\,a,b] = a\,\frac{\displaystyle\prod_{j=0}^{i-2}(b+j)}{\displaystyle\prod_{j=0}^{i-1}(a+b+j)} = \frac{a}{b+i-1}\prod_{j=0}^{i-1}\frac{b+j}{a+b+j}. \tag{23}$$

From (18) and (21), the probability that a certain fault will still remain undetected after k inspection cycles is

$$\begin{aligned}
P[i > k\,|\,a,b] &= \int_{p=0}^{1} P[i > k\,|\,p]\, f(p\,|\,a,b)\,dp \\
&= \prod_{j=0}^{k-1}\frac{b+j}{a+b+j}.
\end{aligned} \tag{24}$$

The numbers of faults discovered during the k review cycles are $\mathbf{x} = \{x_1, x_2,\ldots, x_k\}$, and the

number of undetected faults still remaining in the software document is $N\text{-}s_k$. Thus, it follows from (23) and (24) that the likelihood function of N, a, and b is

$$L(N,a,b\,|\,\mathbf{x}) = \frac{N!}{(N-s_k)!\prod\limits_{i=1}^{k} x_i!}\left(\prod_{j=0}^{k-1}\frac{b+j}{a+b+j}\right)^{N-s_k}\prod_{i=1}^{k}\left(\frac{a}{b+i-1}\prod_{j=0}^{i-1}\frac{b+j}{a+b+j}\right)^{x_i}. \tag{25}$$

5.3. Re-parameterization

We may find the estimates $\hat{N}$, $\hat{a}$ and $\hat{b}$ that maximize the log-likelihood function of (25) simultaneously. However, we propose to transform the parameters a and b in the beta distribution of p so that the solution algorithm for the log-likelihood function is more computationally efficient.

Note that the expected value and the variance of the beta random variable p in (18) are

$$E[p] = \frac{a}{a+b} \text{ and } Var[p] = \frac{a}{(a+b)^2}\frac{b}{(a+b+1)}, \tag{26}$$

respectively [12]. Thus, if the dispersion of detection probabilities p is very small, the estimates of a and b may approach infinity, while $E[p] = a/(a+b)$ is fixed at a constant.

To avoid such a divergence problem, we need to re-parameterize a and b as follows:

$$a = \frac{\pi}{\theta} \text{ and } b = \frac{1-\pi}{\theta}. \tag{27}$$

The new parameter value $\pi = a/(a+b)$ represents the *average* detection probability, whereas $\theta = 1/(a+b)$ shows the *dispersion* of the distribution. If $\theta = 0$, for example, the beta distribution represents the case in which the detection probability p is a constant. As θ increases, the variance of p increases or, in other words, the magnitude of heterogeneity in detection probability p increases.

With the new parameters π and θ, the likelihood function of N, a, and b in (25) becomes

$$L(N,\pi,\theta\,|\,\mathbf{x}) = \frac{N!}{(N-s_k)!\prod\limits_{i=1}^{k} x_i!}\left(\prod_{j=0}^{k-1}\frac{1-\pi+j\theta}{1+j\theta}\right)^{N-s_k}\prod_{i=1}^{k}\left(\frac{\pi}{1-\pi+(i-1)\theta}\prod_{j=0}^{i-1}\frac{1-\pi+j\theta}{1+j\theta}\right)^{x_i}. \tag{28}$$

It can be easily shown that, in an extreme case in which θ equals zero, the likelihood function in (23) simply reduces to that in (10). Thus, our beta-geometric model is a more generalized version of the maximum likelihood method in Section 3.

For a given inspection record $\mathbf{x} = \{x_1, x_2, \ldots, x_k\}$, the maximum likelihood estimates of N, π, and θ can be directly found from (28) by maximizing the three parameters simultaneously. In this paper, however, we modify the three-parameter optimization problem as a two-parameter problem by expressing N as a function of π and θ.

Note that the likelihood function in (28) can be expressed as a combination of two separate functions as follows:

$$L(N,\pi,\theta\,|\,\mathbf{x}) = \frac{N!}{(N-s_k)!\,s_k!}\left(\prod_{j=0}^{k-1}\frac{1-\pi+j\theta}{1+j\theta}\right)^{N-s_k}\left(1-\prod_{j=0}^{k-1}\frac{1-\pi+j\theta}{1+j\theta}\right)^{s_k}$$

$$\times \quad \frac{s_k!}{\prod\limits_{i=1}^{k} x_i!} \quad \prod_{i=1}^{k} \left(\frac{\dfrac{\pi}{1-\pi+(i-1)\theta} \prod\limits_{j=0}^{i-1} \dfrac{1-\pi+j\theta}{1+j\theta}}{1 - \prod\limits_{j=0}^{k-1} \dfrac{1-\pi+j\theta}{1+j\theta}} \right)^{x_i} . \tag{29}$$

Since the first part in (29) is a binomial distribution, the maximum likelihood estimator of N conditional upon π and θ is simply

$$\hat{N}(\pi,\theta) = \frac{s_k}{1 - \prod\limits_{j=0}^{k-1} \dfrac{1-\pi+j\theta}{1+j\theta}} . \tag{30}$$

The second part in (29) is a multinomial distribution with k classes:

$$L(\pi,\theta \mid \mathbf{x}) = \frac{s_k!}{\prod\limits_{i=1}^{k} x_i!} \quad \prod_{i=1}^{k} \left(\frac{\dfrac{\pi}{1-\pi+(i-1)\theta} \prod\limits_{j=0}^{i-1} \dfrac{1-\pi+j\theta}{1+j\theta}}{1 - \prod\limits_{j=0}^{k-1} \dfrac{1-\pi+j\theta}{1+j\theta}} \right)^{x_i} , \tag{31}$$

which is independent of N. Its log-likelihood function is given as follows:

$$\ln L(\pi,\theta \mid \mathbf{x}) = \sum_{i=1}^{k} x_i \left[\ln \frac{\pi}{1-\pi+(i-1)\theta} + \sum_{j=0}^{i-1} \ln \frac{1-\pi+j\theta}{1+j\theta} \right] - s_k \ln \left(1 - \prod_{j=0}^{k-1} \frac{1-\pi+j\theta}{1+j\theta} \right). \tag{32}$$

For a given inspection history $\mathbf{x} = \{x_1, x_2, \ldots, x_k\}$, any optimization software can find the maximum likelihood estimates $\hat{\pi}$ and $\hat{\theta}$ that maximize the log-likelihood function in (32). In a simulation study in the next section, we simply use Microsoft Excel - Solver to obtain $\hat{\pi}$ and $\hat{\theta}$ in (32) and plug those maximum likelihood estimates into (30) to find another estimate $\hat{N}$.

6. PERFORMANCE EVALUATION

6.1. Detection probabilities

In the simulation study in which we compare the performance of the beta-geometric model with those of traditional estimation methods, we assume that there are $N=100$ faults in a software document. To represent the actual situation in which the probability p of being discovered is different from fault to fault, we consider four different cases as shown in Figure 2.

Among the four cases in Figure 2, the detection probabilities in (a) and (b) are symmetrical around 0.5, while (c) and (d) are skewed to the right and left, respectively. Furthermore, the cases (a) and (b) have the same mean, while (a) has a larger variance. The detection probabilities in (c) and (d) have the same variance, but different means.

Note that random numbers u distributed uniformly between 0 and 1 can be used to generate random numbers p of any desired probability distribution by passing them through the inverse cumulative distribution function of the desired distribution. For more information about the inverse transform method, readers are referred to Law and Kelton [13].

For each of the four distributions in Figure 2, the probability density function (PDF), the cumulative distribution function (CDF), and the inverse cumulative distribution function are given below.
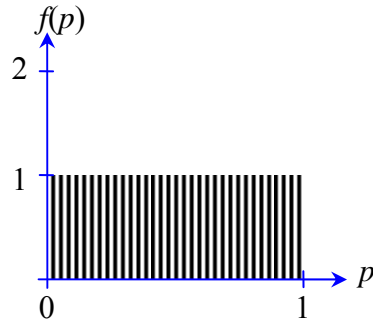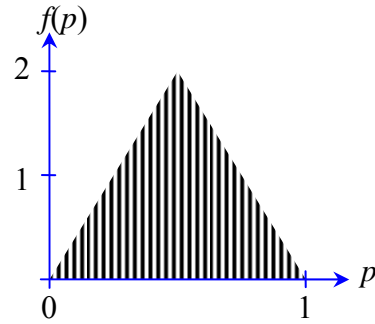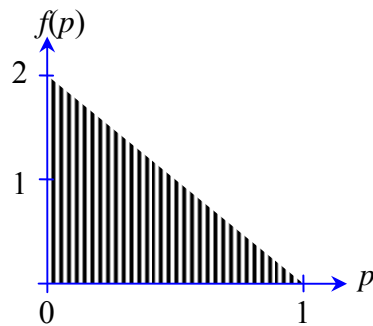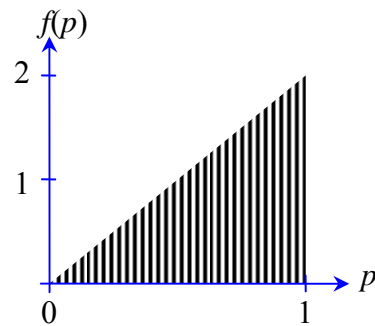
Figure 2. Various distributions of the detection probability p

(a) Rectangular distribution with $E[p]=1/2$

$$f(p) = 1 \qquad\qquad \text{for } 0 < p < 1$$
$$F(p) = p \qquad\qquad \text{for } 0 < p < 1$$
$$p = u \qquad\qquad \text{for } 0 < u < 1$$

(b) Triangular distribution with $E[p]=1/2$

$$f(p) = \begin{cases} 4p & \text{for } 0 < p < 0.5 \\ 4(1-p) & \text{for } 0.5 < p < 1 \end{cases}$$

$$F(p) = \begin{cases} 2p^2 & \text{for } 0 < p < 0.5 \\ 4p - 2p^2 - 1 & \text{for } 0.5 < p < 1 \end{cases}$$

$$p = \begin{cases} \sqrt{0.5u} & \text{for } 0 < u < 0.5 \\ 1 - 0.5\sqrt{2(1-u)} & \text{for } 0.5 < u < 1 \end{cases}$$

(c) Triangular distribution with $E[p]=1/3$

$$f(p) = 2(1-p) \qquad\qquad \text{for } 0 < p < 1$$
$$F(p) = 2p - p^2 \qquad\qquad \text{for } 0 < p < 1$$

$$p = 1 - \sqrt{1-u} \qquad\qquad \text{for } 0 < u < 1$$

(d) Triangular distribution with $E[p]=2/3$

$$f(p) = 2p \qquad\qquad \text{for } 0 < p < 1$$
$$F(p) = p^2 \qquad\qquad \text{for } 0 < p < 1$$
$$p = \sqrt{u} \qquad\qquad \text{for } 0 < u < 1$$

In the simulation study, we first generated a standard uniform random number u_j, $0 < u_j < 1$, for $j = 1, 2, \ldots, 100$, using the random number generator, =RAND( ), in Microsoft Excel. We then transformed it to the random number $\hat{p}_j$ that represents the detection probability of a certain fault j.

6.2. Parameter estimation

In the simulation study, we assume that each of the 100 faults in the software document is subject to review cycle up to $k=10$ times. The exact detection time i of the fault with p follows a geometric distribution, and its cumulative distribution function is $1-(1-p)^i$, where $i = 1, 2, \ldots, 10$. Thus, with another standard uniform random number v from Microsoft Excel, we simulated the detection time i of the fault with p as the smallest integer larger than or equal to

$$i = \left\lceil \frac{\ln(1-v)}{\ln(1-p)} \right\rceil . \qquad\qquad (33)$$

Any faults with i larger than 10 have not been discovered during the $k=10$ review cycles.

For $N=100$ faults in the software document, we then counted the number of faults that have been detected during the ith review cycle. The inspection results are simply given by $\mathbf{x} = \{x_1, x_2, \ldots, x_{10}\}$ for one simulation run. The beta-geometric model, along with other traditional estimation methods, is used to estimate the true parameter value $N=100$ for each simulation run. When $\mathbf{x} = \{44, 17, 8, 6, 4, 8, 0, 1, 0, 1\}$, for example, the estimates of N obtained by the maximum likelihood method, conditional maximum likelihood method, and beta-geometric model are shown to be 89.00, 89.45, and 96.01, respectively.

Table 1. Estimates of N by three methods in four different cases
(The true parameter value is $N = 100$.)

Distribution of p	Performance Measures	Estimation Methods		
		MLE	Conditional MLE	Beta-Geometric
(A) Rectangular with $E[p]=1/2$	Average $\hat{N}$	89.679	89.959	99.702
	MAPE	10.321%	10.041%	4.187%
(B) Triangular with $E[p]=1/2$	Average $\hat{N}$	97.180	97.386	99.848
	MAPE	2.820%	2.635%	2.048%
(c) Triangular with $E[p]=1/3$	Average $\hat{N}$	83.999	84.723	99.756
	MAPE	16.001%	15.277%	6.867%
(d) Triangular with $E[p]=2/3$	Average $\hat{N}$	98.300	98.315	100.095
	MAPE	1.700%	1.691%	1.600%
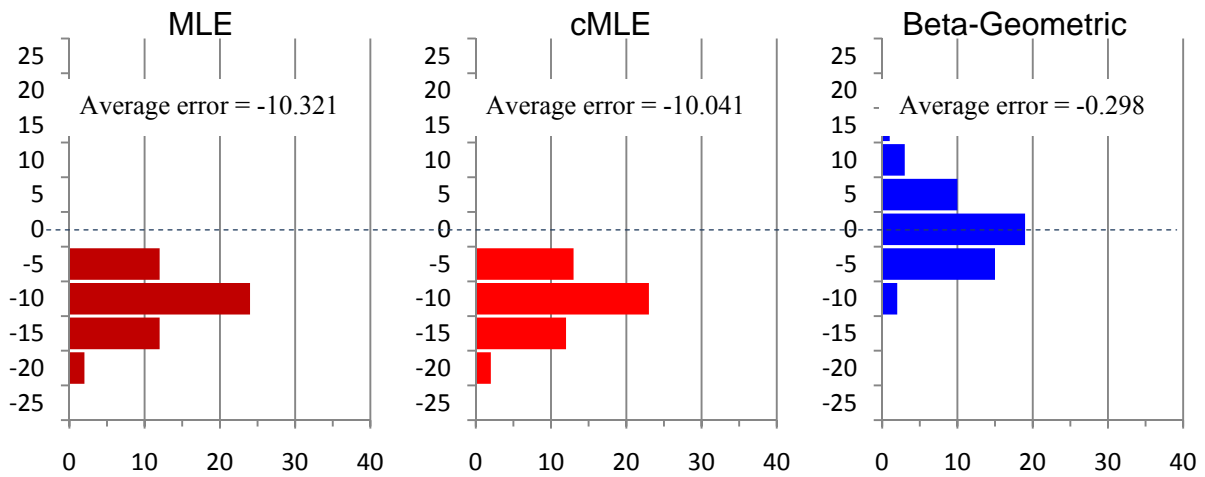
MLE

Average error = -10.321

cMLE

Average error = -10.041

Beta-Geometric

Average error = -0.298

(a) Uniform distribution with E[p] = 1/2

MLE

Average error = -2.820

cMLE

Average error = -2.614

Beta-Geometric

Average error = -0.152

(b) Triangular distribution with E[p] = 1/2

MLE

Average error = -16.001

cMLE

Average error = -15.227

Beta-Geometric

Average error = -0.244
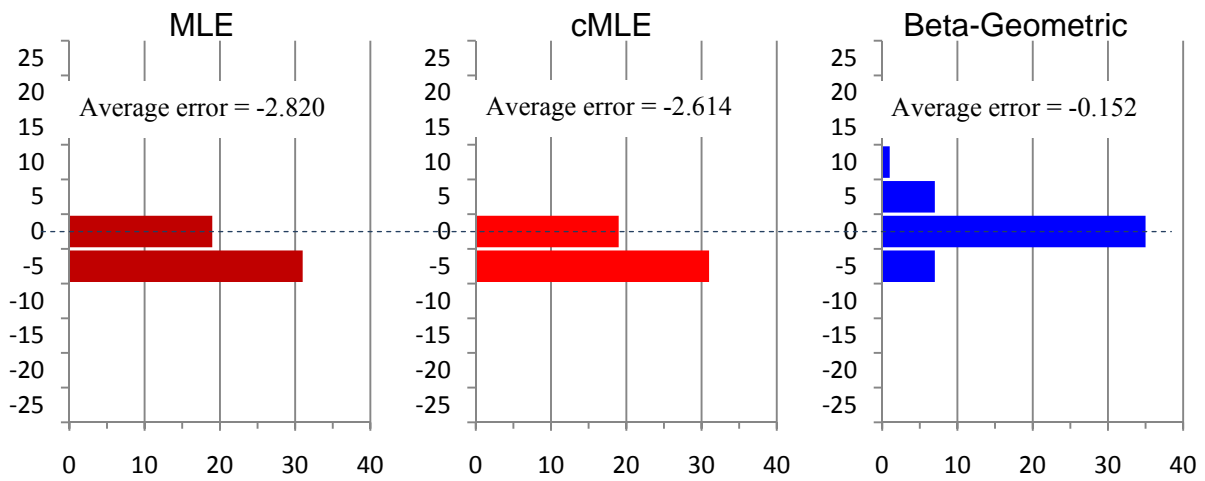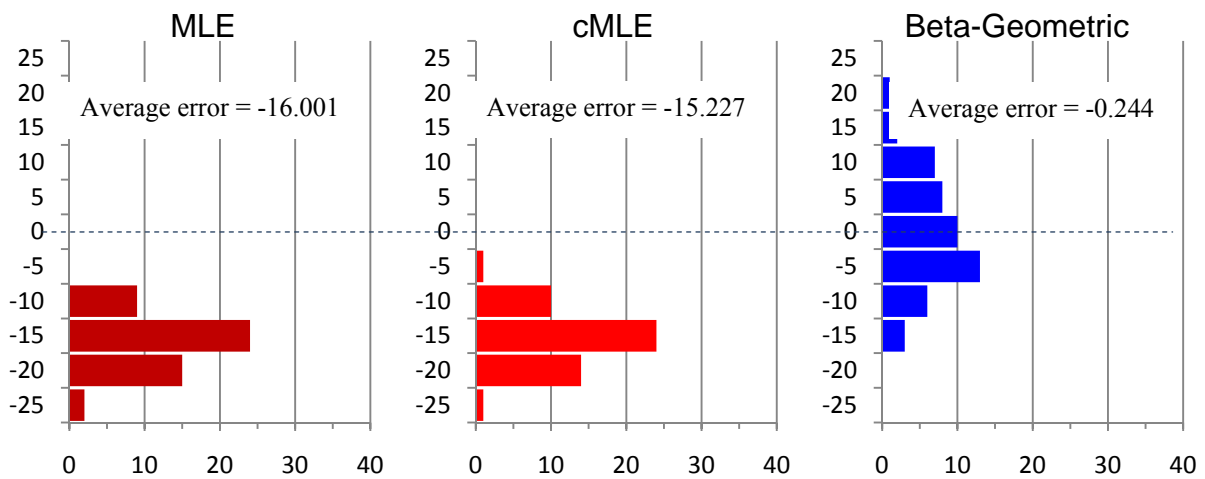
(c) Triangular distribution with E[p] = 1/3
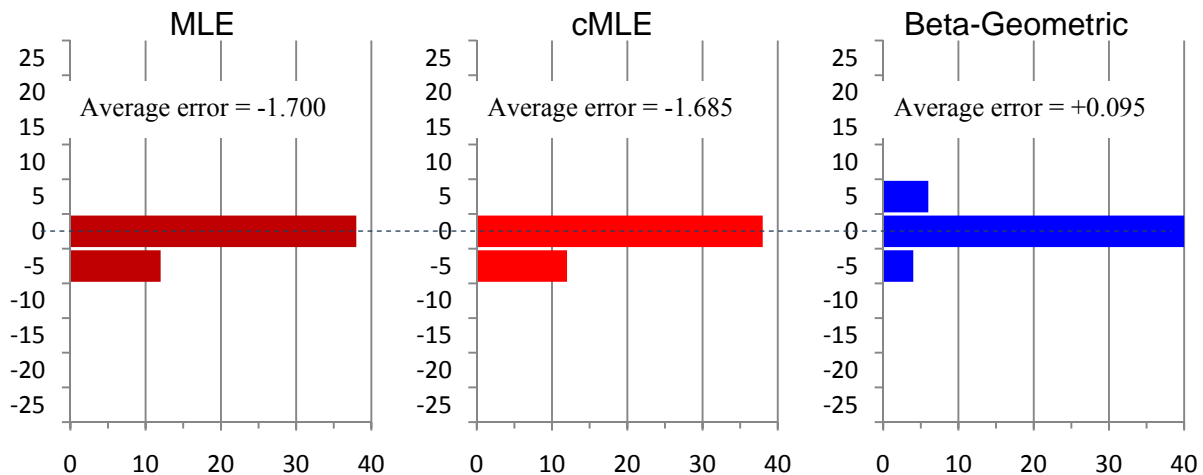
(d) Triangular distribution with E[p] =2/3

Figure 3. Estimation errors of three estimation methods for various types of distribution p

After 100 simulation runs, we then calculated the average estimate of $\hat{N}$ and the mean absolute percentage error (MAPE) for each estimation method. The performance measures are summarized in Table 1. To check the unbiasedness of each estimation method, we also draw histograms of $\hat{N}$ in Figure 3.

6.3. Simulation results

As shown in Table 1, the conditional MLE is slightly better than the MLE in terms of the average estimate and the mean absolute percentage error. However, the beta-geometric model clearly outperforms the traditional estimation methods in all four cases. When the detection probabilities are uniformly distributed between 0 and 1, for example, both the MLE and the conditional MLE severely underestimate the true number of faults N as 89.679 and 89.959, respectively. On the other hand, the average of the 100 estimates obtained by the beta-geometric model is 99.702, which is very close to the true parameter value N=100. Its mean absolute percentage error is 4.187%, which is much better than those of the MLE and the conditional MLE.

According to the histograms of the estimation errors in Figure 3, the traditional methods consistently underestimate the true number of faults when the detection probability is not a constant, but a random variable. Only the beta-geometric model handles the heterogeneity in detection probability very well, giving almost unbiased estimates in all four cases.

7. CONCLUDING REMARKS

In many practical situations, the probability of being detected during each inspection cycle is not the same among different types of defect. That is why we propose in the paper the beta-geometric inspection model in which the heterogeneity in detection probability is simply described as a beta distribution. In a Monte Carlo simulation, we show that our inspection model

clearly outperforms the maximum likelihood method and the conditional maximum likelihood method, predicting the total number of defects N with less biases and smaller variances.

Those simulation results are not unexpected, given that the maximum likelihood method is a special case of our beta-geometric model. The only drawback of the beta-geometric model is that we need to estimate three parameter values (a, b, and N), rather than two (p and N) as in the traditional methods. As shown in the simulation study, however, the computational complexity is not a big problem even with Microsoft Excel. Thus, in estimating the product quality after multiple inspections, there is no reason not to prefer our beta-geometric model over the traditional estimation methods.

In the paper, we focus on the problem of estimating the number of defects (i.e., non-conformities) in a complex product such as a software document. With slight modifications, our beta-geometric model can be applied directly to the problem of estimating the number of defective items (i.e., non-conforming items) in a batch of items such as IC chips [11]. In such a case, we need to estimate the average defective rate as well as the detection probability.

Another application area which is seemingly unrelated to the multiple inspection plan is the consumer response modeling in direct marketing [3]. When survey forms, solicitation letters, or discount coupons are sent out to n people, many of them will not respond immediately or never respond at all. Based on the daily response records during the past k days $\mathbf{x} = \{x_1, x_2, \ldots, x_k\}$, we can estimate the total number of respondents as well as the daily response rate. In a similar situation, we can also determine the total number of survey forms n that should be distributed so that we could achieve a certain level of responses during a given time period.

REFERENCES

[1] Berry, D.A. and Mensch, R.F. (1986) Discrete search with directional information. *Operations Research,* **34**, 470–477.

[2] Bonett, D. G. and Woodward, J. A. (1994). Sequential defect removal sampling. *Management Science*, 40, 898-902.

[3] Bose, I. and X. Chen. 2009. Quantitative models for direct marketing: A review from systems perspective. *European Journal of Operational Research* 195, 1-16.

[4] Briand, L. C., El Eman, K., Freimut, B. G. and Laitenberger, O. (2000). A comprehensive evaluation of capture-recapture models for estimating software defect content. *IEEE Transactions on Software Engineering*, 26, 518-540.

[5] Chun, Y. H. (2009). Improving product quality by multiple inspections: Prior and posterior planning of serial inspection procedures. *IIE Transactions*, 41, 831-842.

[6] Chun, Y. H. (2008). Bayesian analysis of the sequential inspection plan via the Gibbs sampler. *Operations Research,* 56, 235-246.

[7] Chun, Y. H. (2006) Estimating the number of undetected errors in software design document: A correlation matrix model for multiple inspectors. *European Journal of the Operational Research*, 175, 1180~1192.

[8] Chun, Y. H. (2005). Serial inspection plan in the presence of inspection errors: Maximum likelihood and maximum entropy approaches. *Quality Engineering*, 17, 627-632.

[9] Chun, Y. H. and Sumichrast, R. T. (2007) Bayesian inspection model with the negative binomial prior in the presence of inspection errors. *European Journal of the Operational Research*, 182, 1188-1202.

[10] Elshafei, M., M. Khan, and S. O. Duffuaa. (2006). Repeat inspection planning using dynamic programming. *International J. of Production Research*, 44, 257-270.

[11] Greenberg, B. S. and Stokes, S. L. (1995). Repetitive testing in the presence of inspection errors. *Technometrics*, 37, 102-111.

[12] Johnson, N. L., Kemp, A. W., and Kotz, S. (2005) *Univariate Discrete Distributions*, 3nd ed. John Wiley & Sons, New York.

[13] Law, A. M. and W. D. Kelton. *Simulation Modeling and Analysis*, 2nd edition. McGraw-Hill, New York, 1991.

[14] Rallis, N. E. and Lansdowne, Z. F. (2001). Reliability estimation for a software system with sequential independent reviews. *IEEE Transactions on Software Engineering*, 27, 1057-1061.

[15] Sanathanan, L. (1972). Estimating the size of a multinomial population. *Annals of Mathematical Statistics*, 43, 142-152.

THE GENDER GAP: A SERVQUAL EMPIRICAL STUDY OF TRANS-ATLANTIC AIRLINES PASSENGERS

Kien-Quoc Van Pham, Humboldt State University, School of Business, 1 Harp Street Arcata CA 95521, 707-825-0452, kv71@humboldt.edu

ABSTRACT

The SERVQUAL model is used to identify and explore gender differences in service quality expectations and perceptions of passengers on Trans-Atlantic flights. In-flight SERVQUAL based surveys were administered and collected. The study finds no gender based statistical significant difference in the overall importance ratings of the SERVQUAL five operationalized service quality dimensions with the exception of the Assurance dimension. Further investigation was conducted to isolate which of the twenty two service quality attributes would evidence any statistically significant disparity. Identified gender gaps in service quality perceptions and expectations should prove beneficial to airlines intent on pursuing a gender segmentation strategy.

INTRODUCTION

Recent Asian, and American airlines promotional campaigns targeting women passengers driven by the projected increase in the number of women travelers, especially , business women and the paucity of gender specific research in this particular industry indicates the need to empirically validate this emerging trend for gender market segmentation beyond the traditional business economy class one. Furthermore, the advent and sustained global proliferation and growth of low costs airlines, increasing transaction costs transparency and information diffusion with the internet, i.e. virtual airfares comparisons given an itinerary via travel portals, airlines websites, and subsequent legacy purchasing channels "disintermediation" all have contributed to the acceleration of the "commoditization" of airlines services. In a "commoditized", highly competitive environment, the pursuit of a differentiation strategy via quality services is an imperative.

GENDER GAP AND SERVICES MARKETING

The application of the segmentation concept is a timeless topic of critical import in marketing literature and practice. Management has long condoned and sustained the practice of relying on intuition and on traditional segmentation techniques based on socio-demographic variables for both products and services marketing.

Services Characteristics

Shostack (1977) wrote about some of the fundamental differences between the marketing of goods and services. Services are thought to be processes and are less standardized than goods, partly because of their reliance on interpersonal interactions [31]. Zeithaml (1981) posited that services are characterized by experience and credence properties (i.e. those that can only be

evaluated after some consumption or are difficult to evaluate even with some trial) more than search properties, whereas goods are characterized more by search and experience properties [35]. As services are distinctively characterized by their intangible, heterogeneous, inseparable, and perishable nature, academics and practitioners have acknowledged that service marketing is inherently different from product or goods marketing. The services business sectors, e.g. banking, retail, medical, hospital, tourism, travel, entertainment, etc. nowadays are being challenged to come up continuously with more innovative ways to compete and sustain their basic "raison d'etre" in light of commoditization.

For many services, customer-contact employees influence the interaction quality that reflects how the service is delivered [4]. Yet, the airline industry and all of its constituent passenger air-carrying actors have deemed appropriate to sustain principally a segmentation strategy predicated upon the differentiation between business and economy class passengers. Teichert, Sheu, and Von Wartburg in their latest study (2007) of the airlines' marketing practices suggested the need for more alternative segmentation approaches, ones that would capture: a) the preference heterogeneity among customers and b) provide for a better understanding of consumer preferences [32].

Agentic Versus Communal Gender Behaviors

Research has indicated that, in general, most individuals possess assumptions that are biased by gender specific stereotypes [10] [14]. Calls for research testing the socialized aspects of gender specific behaviors led to differentiations between agentic and communal gender role behaviors [11] [12] [21] [22] [18]. Agentic behaviors refer to give-and-take tendencies [11] [12]. Agentic behaving individuals are likely to be described as assertive and ones who would utilize resources as leverage for obtaining a goal. Agentic behaving individuals would "demonstrate" self-sufficiency, independence, and are dominant, aggressive and task-oriented [5] [11] [12]. Men are more likely to display agentic behaviors than their female counterparts [12]. Communal behaviors refer to a concern with interpersonal relationships [11]. An individual who displays communal behaviors is likely to be described as caring, empathetic and nurturing. Communal behaving individuals do show sympathy, and are socially oriented, helpful and expressive [5] [12]. Eagly et al (2003) found that women are more likely to display communal behaviors than their male counterparts [12].

This agentic versus communal distinction between the two sexes impacts how each gender (the more emotional females and the more rational males) observes, evaluates and relates to the environment. It has been suggested that women have a greater concern for social context and relationships, assigning more importance to interpersonal relationships relative to men. The emphasis women place on emotional as opposed to rational factors is reflected in their approach to evaluating an experience. Most women are characterized as emotion-dominated "Feeling" types while most males are designated as logic-dominated "Thinking" types according to Myers – Briggs Type Indicator (MBTI) research. This stream of research suggests that "Thinking" type people (males) evaluate experiences on the basis of rational factors whereas "Feeling" types (females) tend to rely on affective processes.

There is also some evidence that females and males differ in their approaches to forming relationships with service providers and in evaluating service encounters. As women are socialized to maximize the interpersonal aspects of their relationships, they emphasize the process component of the service delivery. They are also influenced more by relational information as opposed to efficiency and accuracy cues. Men, on the other hand, tend to focus more on the successful delivery of the core service. As long as the appropriate outcome is attained, the process or the delivery style is not as important to them. Indeed, according to Hall and Carter (1999), gender stereotypes are incorporated into customer expectations regarding good service [17a]. Research also shows that women are more likely to be frontline service providers who assist the customer. Women customers also rate employees more favorably if they are of the same gender, and they tend to more likely identify with the frontline employees.

Service Encounter: Core and Relationship Components

Studies have already converged in the conceptualization of a service encounter as an interchange between a service provider and a client, in which the client experiences both the core service component and a relationship component [1] [2] [3] [4][27][29][30]. The core of a service is the part of the service we think of when we name the service. The relationship aspect of a service describes the interpersonal process by which the service is delivered and is thought to be especially important in customer interactions with professional service providers [9] [30]. Customers may view those employees who are perceived as reliable, responsive, and caring as "friends" [26]. When employees are perceived as "friends" with the ability and desire to provide excellent service, this is expected to foster confidence in the individual and trust in the organization. Likewise, when service firms are perceived to be reliable in fulfilling the service promise, this should enhance trust [26].

Yet, the airline industry has been one that has been rather "reluctant", "slow" to address such gender issues. These issues were identified formally by a qualitative exploratory survey of this industry in general by Westwood, Pritchard, and Morgan (2000) with a particular focus on women business travellers [34]. The authors who labeled this industry's marketing practices as "gender blind" noted that this flying population segment expressed concern over inadequate levels of comfort and the sexist attitudes of airlines personnel [34]. Weber (2005) in her airlines alliances service quality perceptions survey reaffirmed two of the previously isolated significant differences as women travelling assigned greater importance to: 1) respectful treatment and, 2) better assistance in case of problems [33].

SERVICE QUALITY AND SERVQUAL

In the services marketing literature, the widely used definition of service quality is to meet customers' expectations [25]. According to Parasuraman et al, in their review of the quality theory literature, service quality can neither be conceptualized nor evaluated by the traditional methods of goods quality because services possess three characteristics: intangibility, heterogeneity and inseparability [25] [26] [27]. They have defined and conceptualized service quality as a form of attitude, which results from a comparison of customers' expectations with perceptions of performance, and developed SERVQUAL to measure service quality.

The SERVQUAL scale instrument is based on a gap model [25] [26], which suggests that the gap between the customers' expectations and their perceptions of actual performance drives the perception of service quality. Both the original version of SERVQUAL [26] and its revised versions [27] [28] contain five dimensions:

1. Tangibles assess the appearance of a company's physical facilities, equipment and personnel, the elements of the service environment that impact upon perceived service quality, i.e. cleanliness of premises, staff appearance and the appropriateness of other "tangible" items, i.e. computers or phones.
2. Reliability measures the ability to perform the service dependably and accurately, i.e. consistency in meeting service promises, keeping schedules or appointment times, completing tasks on time, ensuring that outcomes are met.
3. Responsiveness represents the willingness to help customers and provide a prompt service, the ability of the service to respond to individual customer requirements, e.g. specifying delivery times, altering aspects of the delivery process, and ensuring that customers remain involved.
4. Assurance assesses the knowledge and courtesy of employees and their ability to impart, inspire confidence. This dimension would also include staff training in the use of tools and knowledge of their service processes, customer interaction, and the perception that the service is competent.
5. Empathy measures the caring and individualized attention provided to customers, embodying access, communication and understanding. This combined dimension is really about the communication style of the service organization through its service personnel, its communications including leaflets, instructions, signage and people management.

These five dimensions were derived from ten overlapping dimensions, which were regarded as essential to service quality by Parasuraman et al exploratory research [25]. Service quality can be defined as a consumer's overall impression of the relative efficiency of the organization and its services. Understanding exactly what customers expect is the most crucial step in defining and delivering high-quality service according to Zeithaml et al [35] (36) [37]. As in other sectors, the problem in the airline sector is whether management can correctly perceive what customers want and expect. Expectations serve as a major determinant of a consumer' service quality evaluations and satisfaction [24]. The ''voice of the customer'' should be factored into the service design process, and after delivery, service providers should monitor how well customers' expectations have been met. For this purpose, SERVQUAL is one of the best models for evaluating customers' expectations and perceptions [30 [7]; it remains as the most commonly used diagnostic model for evaluating service quality [20] [23].

Other scholars have measured airline service quality utilizing various quality dimensions other than SERVQUAL's. Gourdin (1988) defined airline service quality in terms of three items: safety, timelines and price [16]. Elliott and Roach (1993) proposed food and beverage, timely luggage transport, seat comfort, the check in process, and in-flight service dimensions [13]. Haynes and Percy (1994) factored in the processing of luggage, seat cleanliness, and the check-in process, the convenience of transit, timeliness, and handling of customer complaints as standards of service quality [17b]. Gilbert and Wong (2003) incorporated flight patterns into tangibles,

breaking it further into employees and facilities, and added customization [15]. Pakdil and Aydm (2007) made use of flight patterns also, with the addition of image and availability [30].

This proposed exploratory study seeks to investigate, identify and diagnose gender specific service quality issues with the SERVQUAL model/instrument, hypothesizing that service quality expectations and perceptions, and dimensions (core and relational) importance ratings will vary according to gender. Core business service quality dimensions are identified as Tangibles, Reliability. Core relational/relationship (certain authors refer to these as core peripheral) service quality dimensions are defined by Responsiveness, Assurance and Empathy. Furthermore, women passengers will assign greater importance to these relationship/relational aspects relative to men.

METHODOLOGY AND FINDINGS

In-flight service quality surveys based on the SERVQUAL twenty two service attributes for expectations and perceptions (Appendix I) representing the five dimensions (Table I) and the attendant basic socio-economic demographic variables were administered and collected from passengers on Trans-Atlantic flights for a number of airlines resulting in a total overall usable sample of 642 passengers, comprised of 422 male and 220 female passengers.

In terms of scale and instrument reliability, all twenty two service attributes (using a 7 points Likert scale) were validated with Cronbach's Alphas of .949 for perceptions and .933 for expectations respectively. Data analysis was conducted with both parametric and non-parametric statistical methods given the mix of ordinal and interval scale ratings and to allow for the presence of skewness and kurtosis.

Overall, when queried as to the most important, second most important and lest important service quality dimensions, both genders deemed Reliability as the most important one (56.5% of responses), followed by Responsiveness (40.5%), and Tangibles (44.4%) least. This finding for Reliability and Tangibles corroborates Zeithaml et al (1990), and Sultan and Simpson (2000) studies [35] [36] [37] [28], but Pakdil and Aydm (2007) found Responsiveness superseding Reliability as the most important [30]. Gender differences in the relative importance percentage weights distribution of the five dimensions (all five totaling 100%) were not statistically significant with the exception of the Assurance dimension (t-value of -2.449), one of the three identified relational/relationship dimensions (Table 1). This finding is confirmed with Mann-Whitney U test with p value of 0.024.

In this study, the other relational/relationship dimensions, Responsiveness and Empathy which were expected to be conceptually gender sensitive in terms of relative import proved statistically inconclusive. Empathy also did not appear to be as significant a service quality dimension indicator according to both genders, as indicated by prior service research

Table 1

Dimensions	Gender	N	Mean	Levene's (variances) F	Levene's (variances) Sig.	t-test (means) t	t-test (means) df
Tangibles	Male	418	.1577	.037	.847	.916	634
	Female	218	.1500				
Reliability	Male	419	.2933	1.999	.158	-.100	636
	Female	219	.2945				
Responsiveness	Male	419	.2109	.531	.467	-.062	636
	Female	219	.2113				
Assurance	*Male*	*417*	*.1725*	*1.206*	*.273*	*-2.449*	*632*
	Female	*217*	*.1896*				
Empathy	Male	419	.1621	.437	.509	.660	636
	Female	219	.1574				

Subsequent analysis of the twenty two corresponding SERVQUAL operational service attributes (Appendix 1) for gender differences in ratings yielded the following:

Gender Differences in Service Quality Expectations

For the Tangibles dimension, the 4th operational descriptor "An excellent airline's materials associated with its service, e.g., pamphlets or statements, will be visually appealing" proved to be statistically significant (t value of -2.571), likely due to diverging canons of esthetics. In terms of Responsiveness, all four service attributes exhibited divergence as hypothesized: "Employees of excellent airlines will tell customers exactly when services will be performed" (t-value -2.367), "Employees in excellent airlines will give prompt service to customers" (t-value -2.219), "Employees of excellent airlines will always be willing to help customers" (t-value 2.902), "Employees in excellent airlines will never be too busy to respond to customer requests" (t-value -2.854). Likewise, for all four Assurance descriptive statements, the gender gap was confirmed statistically as indicated previously by its overall relative importance rating: "The behavior of employees in excellent airlines will instill confidence in customers" (t-value -3.346), "Customers of excellent airlines will feel safe in their transactions" (t-value -3.247), "Employees in excellent airlines will be consistently courteous with customers" (t-value -3.507), and "Employees in excellent airlines will have the knowledge to answer customers' questions" (t-value -4.911). Empathy-wise, only "Excellent airlines will have the customer's best interests at heart" (t-value -2.652) was significant for gender differential.

All of the above gender-based differences were confirmed with non-parametric tests, Mann-Whitney U and Wilcoxon W, with Z values of: -2.333 for the Tangible dimension, -1.9888, -1.992, -2.823, and -3.483 respectively for Responsiveness, -4.077, -3.826, -3.652, and -5.402 for Assurance, and -2.850 for Empathy.

Gender Differences in Service Quality Perceptions

With the exception of "An excellent airline's materials associated with its service, e.g., pamphlets or statements, are visually appealing" for Tangibles (t-value -2.671), and "Excellent airlines have

operating hours convenient to all their customers" for Empathy (t value -2.082), both genders were congruent in their relative and respective perceptions of the airlines' service quality SERVQUAL criteria. Mann-Whitney U and Wilcoxon W non-parametric tests elicited two more statistical gender-based divergent perceptions beyond the ones indicated with parametric tests: "An excellent airline's employees are neat appearing" (Z value -2.841) for Tangibles as above, "Employees in excellent airlines are consistently courteous with customers" (Z value -1.955) for Assurance, "Excellent airlines have operating hours convenient to all their customers" (Z value -1.921) as above, and "Excellent airlines have the customer's best interests at heart" (Z value -1.946).

Chau and Kao (2009) study of SERVQUAL Gap-5 in airline services between Taipei and London found no support for the gender gap [6]. The authors stated that "this is unsurprising, as there is no obvious reason why one gender would judge a service quality dimension any differently from the other", and that their data analysis "does not suggest strongly that demographic factors relate to customers' evaluation of service quality-with only income levels and occupation being significant" [6]. The authors concluded that "the findings are generalizable to other contexts that share similar characteristics as tight regulation, open competition, and where service quality is paramount" [6]. This author sincerely beg to differ as this empirical analysis of Trans-Atlantic passengers and findings does affirm the need to address women passengers' differential expectations of service quality, in particular for Responsiveness and for Assurance.

The debate continues to rage about the pros and cons of the SERVQUAL model versus its main contender, Cronin and Taylor's (1992) SERVPERF model [8], the latter one advocating the elimination of expectations for pure performance scoring (perceptions). However, this research endeavor did not elicit as many gender based differences with Perceptions/Performance scores vis–a-vis Expectations. Therefore, given the exploratory, investigative nature and objectives of this study for/of potential gender differences in airline services, the SERVQUAL model has performed exceedingly well in this author's opinion.

MANAGERIAL IMPLICATIONS AND STUDY LIMITATIONS

American Airlines launched their women only web portal in 2007 with great fanfare. All Nippon Airways (ANA) announced on February 24, 2010 their designating one restroom as female-only on most international routes. Yet, as reported by the Wall Street Journal (Eastern edition, September 30, 1996), airlines in Asia were offering personalized prices by age, race, and gender with Air India even providing "ladies fare" to shopping destinations since then. These gender-based operational and promotional campaigns and strategies are definitely justified for differentiation purposes in this globally commoditized service industry, and empirically defensible according to this study. Lagace (2005) commented that "despite their numbers as a growing force in the marketplace, women business travellers are still often shoehorned into a model designed for men" [19]. She quoted American Airlines Vice President Peggy Sterling's remarks "Women who travel are clearly a rising population and all the statistics show it", who also acknowledged that for American Airlines, "female travelers…complain to American that flight attendants are not responsive enough to their needs compared to men" while both genders expect the same quality service and efficiency [19].

This study affirms the need for airlines to differentiate between genders for service quality purposes, to specially focus on the relational, "affective" components as identified for Responsiveness and Assurance for women travelers, and to continuously monitor service expectations for both to ensure positive "moments of truth". Since other service research have advocated the need to ensure satisfactory levels for the core service components to avoid service quality derogation, it is rather critical to prioritize in which service areas Management must deploy the needed resources. Hopefully, this study has provided a starting "blueprint". However, the findings of this investigation are specific to Trans-Atlantic flights and to the passengers surveyed at this point in time. Research have shown that expectations are definitely not immutable, and a longitudinal follow up is needed to validate the elicited gaps for Management to sustain the suggested focus on the affective, relational components of the services being rendered for service quality assurance purposes. The growing passenger air traffic on Trans-Pacific routes also does point to the need to assess gender-based service issues/gaps for this corridor and to also assess the impact of nationality and culture on airline service quality for customers' delights, and most importantly for competitive differentiation purposes.

APPENDIX I
SERVQUAL Dimensions and Operational Descriptive Statements
Expectations (future tense "will"), Perceptions (present tense "have", "are")

Service Quality Dimension	Statement No./Code	Descriptive statement
Tangibles	(1) TAN1	Excellent airlines will have modern-looking aircraft.
	(2) TAN2	The office, terminal and gate facilities of excellent airlines will be visually appealing.
	(3) TAN3	An excellent airline's employees will be neat appearing.
	(4) TAN4	An excellent airline's materials associated with its service, e.g., pamphlets or statements, will be visually appealing.
Reliability	(5) REL1	When excellent airlines promise to do something by a certain time, they will do so.
	(6) REL2	When a customer has a problem, an excellent airline shows a sincere interest in solving it.
	(7) REL3	Excellent airlines will perform the service right the first time.
	(8) REL4	Excellent airlines will provide their services at the time they promise to do so.
	(9) REL5	Excellent airlines will insist on error-free records.
Responsiveness	(10) RSP1	Employees of excellent airlines will tell customers exactly when services will be performed.
	(11) RSP2	Employees in excellent airlines will give prompt service to customers.
	(12) RSP3	Employees of excellent airlines will always be willing to help customers.
	(13) RSP4	Employees in excellent airlines will never be too busy to respond to customer requests.
Assurance	(14) ASR1	The behavior of employees in excellent airlines will instill confidence in customers.
	(15) ASR2	Customers of excellent airlines will feel safe in their transactions.
	(16) ASR3	Employees in excellent airlines will be consistently courteous with customers.
	(17) ASR4	Employees in excellent airlines will have the knowledge to answer customers' questions.
Empathy	(18) EMP1	Excellent airlines will give customers individual attention.
	(19) EMP2	Excellent airlines will have operating hours convenient to all their customers.
	(20) EMP3	Excellent airlines will have employees who give customers personal attention.
	(21) EMP4	Excellent airlines will have the customer's best interests at heart.
	(22) EMP5	The employees of excellent airlines will understand the specific needs of their customers.

REFERENCES

[1] Berry, L. L. *Relationship Marketing*. In L. L. Berry, G. L. Shostack, & G. D. Upah (Eds.), 1983, Emerging perspectives on services marketing (pp. 25-28). Chicago: American Marketing Association.

[2] Bitner, M. J. *Servicescapes: The impact of surroundings on customers and employees*. 1992, Journal of Marketing, 56, 57-71.

[3] Bitner, M. J., Booms, B. H., & Tetreault, M. S. *The service encounter: Diagnosing favorable and unfavorable incidents*. Journal of Marketing, 1990, 54, 71-84.

[4] Brady, M.K. and Cronin, J.J. Jr. *Some new thoughts on conceptualizing perceived service quality: A hierarchical approach*. Journal of Marketing, 2001, 65 (3), 34-49

[5] Carli, L. L. *Gender differences in interaction style and influence*. Journal of Personality and Social Psychology, 1989, 56, 565-576

[6] Chau, V.S., and Kao Y.K. *Bridge over troubled water or long and winding road?Gap-5 in airline service quality performance measures*. Managing Service Quality, 2009, Vol. 19, No. 1, 106-134, Emerald

[7] Chen, C. F. *Investigating structural relationships between service quality, perceived value, satisfaction, and behavioral intentions for air passengers: Evidence from Taiwan*, Transportation Research, 2008, Part A, Vol. 42, 709-717.

[8] Cronin, J. Joseph, Jr., and Steven A. Taylor (1992).*Measuring Service Quality; A Reexamination and Extension*. Journal of Marketing, 56 (July), 55-68.

[9] Crosby, L. A., & Stephens, N. *Effects of relationship marketing on satisfaction, retention, and prices in the life insurance industry*. Journal of Marketing Research, 1987, 24, 404-411.

[10] Deaux, K. *From individual differences to social categories: Analysis of a decade's research on gender*. American Psychologist, 1984, 39, 105-116.

[11] Eagly, A. H., & Johnson, B. T. *Gender and leadership style: A meta-analysis*. Psychological Bulletin, 1990, 108, 233-256.

[12] Eagly, A. H., Makhijani, M. G., & Klonsky, B. G. *Gender and the evaluation of leaders*. Psychological Bulletin, 1992, 111, 3-22.

[13] Elliott, K. M. and Roach, D. W. *Service quality in the airline industry: are carriers getting an unbiased evaluation from consumers?* Journal of Professional Services Marketing, 1993, Vol. 9, 71-82.

[14] Ely, R. J. *The Power in Demography: Women's Social Constructions of Gender Identity at Work*. Academy of Management Journal 38 (1995): 589-634.

[15] Gilbert, D., & Wong, R. *Passenger expectations and airline services: a Hong Kong based study*. Tourism Management, 2003, 24(5), 519-532.

[16] Gourdin, K. *Bringing quality back to commercial travel*. Transportation Journal, 1998, Vol. **27,** 23-29.

[17a] Hall, J. A., Carter, J. D. *Gender-stereotype accuracy as an individual difference*. Journal of Personality and Social Psychology,1999, *77*, 350-359

[17b] Haynes, R., & Percy, J. *Perception paradox: Airline service quality issues*. Proceedings of Decision Sciences Institute 1994 Annual Meeting, 3, 1950-1952

[18] Iacobucci, D. and Ostrom, Amy. *Gender Differences in the impact of Core and Relational Aspects of Services on the Evaluation of Service Encounters,*1993, Journal of Consumer Psychology, 2(30, 257-

[19] Lagace, M. *The Hidden Market of Female Travelers*, HBS Working Knowledge, 2005

[20] Lam S.S.K. and Woo S. *Measuring Service Quality: A Test-Retest Reliability Investigation of SERVQUAL*. Journal of the Market Research Society, 1997, 39(2), 381-96

[21] Meyers-Levy, J. *Gender differences in information processing*, in P. Cafferata & A. Tybout (Eds.), Cognitive and affective responses to advertising (pp. 219-260). MA: Lexington Books, 1989.

[22] Meyers-Levy, J., & Maheswaran, D. *Exploring differences in males' and females' processing strategies*. Journal of Consumer Research, 1991, 18, 63-70.

[23] Mittal, Banwari and Walfried M. Lassar. *Why Do Customers Switch? The Dynamics of Satisfaction versus Loyalty*. Journal of Services Marketing, 1998, 12(3), 177-194

[24] O'Connor et al., 200

[25] Parasuraman, A., Leonard L. Berry, and Valarie A. Zeithaml. *A Conceptual Model of Service Quality and Its Implications for Future Research.* Journal of Marketing, 1985, 49 (Fall), 41-50

[26] Parasuraman, A., Leonard L. Berry, and Valarie A. Zeithaml (1988). *SERVQUAL: A Multiple-Item Scale for Measuring Consumer Perceptions of Service Quality.* Journal of Retailing, 1988, 64(1), 12-40.

[27] Parasuraman, A., Leonard L. Berry, and Valarie A. Zeithaml. *Refinement and Reassessment of the SERVQUAL Scale.* Journal of Retailing, 1991, 67 (Winter), 420-450

[28] Parasuraman, A., Leonard L. Berry, and Valarie A. Zeithaml. *Research note: More on improving quality measurement.* Journal of Retailing, 1993, 69 (Spring), 140-147

[29] Parasuraman, A., Leonard L. Berry, and Valarie A. Zeithaml. *Reassessment of Expectations as a comparison Standard in Measuring Service Quality: Implications for Further Research.* Journal of Marketing, 1994, 58 (January), 111-124.

[30] Pakdil, F. and Aydm, O. *Expectations and perceptions in airline service: An analysis using weighted SERVQUAL scores*. Journal of Air Transport Management, 2007, Vol. 13, 229-237.

[31] Shostack, G. L. *Service positioning through structural change*. Journal of Marketing, 1977, 51, 34-43.

[27] Solomon, M. R., Surprenant, C., Czepiel, J. A., & Gutman, E. G. *A role theory perspective on dyadic interactions: The service encounter*. Journal of Marketing, 1985, 49, 99-111.

[28] Sultan, F., & Simpson, M.C. *International service variants: Airline passenger expectations and perceptions of service quality*. Journal of Services Marketing, 2000, 14(3), 188-216

[29] Surprenant, C. F., & Solomon, M. R. *Predictability and personalization in the service encounter*. Journal of Marketing, 1987, 51, 86-96.

[30] Swartz, T. A., & Brown, S. W. *Consumer and provider expectations and experiences in evaluating professional service quality*. Journal of the Academy of Marketing Science, 1989, 17, 189-195.

[31] Teas, R. Kenneth. *Expectations, Performance Evaluation and Consumers' Perceptions of Quality.* Journal of Marketing, 1993, 57 (October), 18-34

[32] Teichert, T., Shehu, E., and Wartburg, I.V. *Customer segmentation revisited: The case of the airline industry*. Transportation Research Part A, 2008, 227-242. Elsevier

[33] Weber, Karin. *Travelers' Perceptions of Airline Alliances*. Journal of Travel Research, 2005, 43(3), 257-265,

[34] Westwood, S., Pritchard, A., and Morgan, N.J. *Gender blind marketing: busineswomen's perceptions of airline services.* Tourism Management, 2000, 21, 353-362. Elsevier

[35] Zeithaml, V. A. *How consumer evaluation processes differ between goods and services*. In J. H. Donnelly & W. R. George (Eds.), Marketing of services (pp. 186-189). Chicago: American Marketing Association, 1981.

[36] Zeithaml, V.A., Parasuraman, A., and Berry, L.L. *Delivering quality service – balancing customer perceptions and expectations*. New York: The Free Press, 1990
[37] Zeithaml, V. A., Berry, L. L. and Parasuraman, A. *The behavioral consequences of service quality*. Journal of Marketing, 1996, Vol. 60**,** 31-52.

OPERATIONALIZING SERVICE QUALITY: CUSTOMERS' PERSPECTIVE

Linda Boardman Liu, Simmons College School of Management
300 The Fenway, Boston, MA 02115
617-521-2412 / linda.liu@simmons.edu

ABSTRACT

This research examines the relationships between technical quality, functional quality, and overall quality in the service operations environment. We capture individual customer assessments of service quality both in terms of what activities and measures are important to service quality, as well as specific assessment of a particular service encounter. Our findings provide insights into how to manage service quality.

Introduction & Literature Review

Customers select services that will both satisfy their needs and provide a pleasant experience. Service companies try to do a good job for their customers yet customers are frequently disappointed and companies often do not understand why. For services, there are many approaches to defining quality but common to all is that the concept of quality is important as service quality is clearly linked to customer satisfaction (examples include [2], [1], or [6]).

Harvey [5] differentiated between quality of results and quality of process. He proposed that results are the objective, technical aspect of quality and that the process quality includes four of the Parasuraman, Berry & Zeithaml [7] dimensions: empathy, responsiveness, assurance, and tangibles. Grönroos [3] described a dichotomized service quality framework consisting of technical and functional quality. Grönroos connected these two dimensions together as "experienced" quality. More recently, Grönroos wondered if he should have called it "technical and functional features" which supports the idea that technical and functional quality are distinct from each other and contribute each in their own way to overall quality [4].

Recognizing that service quality is a critical goal of the service delivery system, managing the system to deliver this quality experience is a critical, if not *the* critical, operations objective. What specific aspects of the service delivery system actually result in service quality continues to be elusive. If service quality is only assessed through customer perception, then each specific service encounter will have to be unique, as each customer is unique. But managing a system that is completely customized to each customer's expectations is simply not viable for most service providers. Therefore, the service process needs to be designed and managed to achieve specific, directly measured elements that will result in the indirectly measured, customer perceived service quality. Understanding customers' expectations is challenging and each customer may have different expectations – from other customers, and even from encounter to encounter. Managing to an average level of quality satisfies no one specifically, and setting service quality perception goals that are "the best" may be impossible to achieve, challenging to manage, and prohibitive from both a cost and process perspective.

This research examines the relationships between technical quality, functional quality, and overall quality in the service operations environment. We capture individual customer assessments of service quality both in terms of what activities and measures are important to service quality, as well as specific assessment of a particular service encounter. Our findings provide insights into how to manage service quality.

Research Setting
This research is set in the call center environment, a mass service process type by Schmenner's classification [8]. Call centers are deployed throughout the world as a cost-effective way of enabling customer-company interactions. A call center provides front-line contact to customers. An inbound call center is generally accessed by a customer initiating a call to the company. The call is routed to a group of agents usually with the help of a call distribution platform and/or voice response unit in order to route the caller to the 'right' agent, depending on the firm's choice of segmentation.

Survey Data Collection
We conducted a survey to explore the operational definition of service quality via the customers' conceptualization of service quality. We partnered with an organization that provided customer-level contact information immediately after a service encounter. We collected customer-response data related to the specific service encounter via an on-line survey distributed within four days after the service encounter. The data set includes the customer's perception of technical quality, functional quality, and overall service quality. The data from this study were analyzed to identify any relationships between and among the quality constructs.

Survey Respondent Demographics
Two hundred thirty-one individuals provided contact information to participate in the survey. Of the 207 emailed surveys, 53 were completed. Of the 24 telephone contacts attempted, 10 contacts were made, and all 10 surveys were completed. We paneled the data from the two respondents groups, online and telephone, and conducted t-tests to determine if there was a difference in the groups. No significant difference was found between the two groups' responses therefore we have combined the groups into one data set. With only one-touch allowed for data collection, we obtained a 30.4% response rate.

We initially asked the respondents an open-ended question to solicit general ideas and concepts relative to a quality call center experience. Respondents were asked "When you call a company and have an interaction with a service representative, what makes that encounter a quality experience? List up to three things that happen during a call that make you think a company provides high quality service." Respondents provided 174 individual statements, with 61 of the 63 respondents providing at least one statement. We used open coding to analyze and reduce these statements (Table 1). Four dominant themes emerge from these statements. First, they expect the agent to have a certain kind of attitude. Secondly, customers expect that the agent will be capable to resolve the issue. Third, customers expect that the encounter will be expedient. Finally, they expect the agent to be genuine. Although not a dominant theme, 10 respondents specifically stated that the call needed to be answered by a "live person" or "human."

Table 1 – Customer Expectations

Expectation	Key words	# Responses
Attitude	Friendly…courteous…polite…tone...attitude…helpful… easy… interested… pleasant	51
Resolve	Knowledge…answered question…information available …get answer…resolves issue…competence	46
Expedient	Speed…wait time…efficient…hold time	31
Genuine	Listen…understanding…empathy…unscripted	27

Customer Evaluation: Measures & Experiences

Survey participants were presented a list of typical call center measurements and asked to indicate how important these measures were relative to service quality on a 7-point scale, ranging from "very unimportant" to "very important." See Table 2 for the average rating and standard deviation for each measure. Customers believe that low wait time is important for service quality, measured as service level ("most of the time when you call, the call is answered quickly"), wait time ("how long you wait until the CSR answers the phone"), queue length ("how long you wait until you hang up and call later") and access ("whether you get a busy signal"). In addition, customers believe that first call resolution ("whether your needs are met on your first call") is also important relative to service quality. From the customer perspective, these five measures are statistically equal. Following these measures, customers perceive the amount of time spent on the phone as being important to service quality (mean=5.63). Of the measures evaluated, customers rank call monitoring as the least important measure relative to service quality (mean=5.00). Of note, all of the measures are relatively important from the customer perspective, with 5.0 being the lowest mean on a 1 – 7 point scale.

Table 2– Customer Importance Rating of Call Center Measures

Measure	Importance Rating (1 – 7)	Standard Deviation
How long you wait until the customer service representative answers the phone.	6.40	1.056
Whether your needs are met on your first phone call.	6.38	1.113
The total time you are on the phone interacting with the customer service representative.	5.63	1.299
Whether you get a busy signal when you call the company.	6.29	1.122
How long you wait for someone to answer before you hang up and call again another time.	6.37	.927
Most of the time when you call, the call is answered quickly.	6.43	.979
Whether the company has someone monitoring calls to ensure quality.	5.00	1.558

Customers were asked to identify the nature of the relationship between these call center measures and service quality. Table 3 presents the response frequencies for each measure in order of highest consistency. There is high consistency in the perception that having to wait to reach a CSR is related to low service quality, as is having to call back. Interestingly, actual call

duration and the customer perception of its relationship to service quality is mixed. A majority (33 of 57 respondents) indicate that longer encounters increases service quality. Given the emphasis on speed this is an interesting finding.

Table 3 – Call Center Measures Relationship to Service Quality

Description	Up	Down
If most of the time when you call the call is answered quickly, service quality goes ...	59	1
When you wait longer for someone to answer before you hang up and call again another time, service quality goes ...	1	59
When you have to call back to get your needs satisfied, service quality goes ...	2	59
If you get a busy signal when you call the company, service quality goes ...	2	57
When you wait longer for a customer service representative to answer the phone, service quality goes...	3	57
If the company has someone monitoring calls to ensure quality, service quality goes ...	53	4
When you are on the phone longer interacting with the customer service representative, service quality goes ...	33	24

Participants were asked to evaluate the importance of various experiences identified as being part of a quality service encounter relative to service quality using a 7-point scale, ranging from "very unimportant" to "very important". Table 4 presents the average ratings. Most important to these customers is that the customer service representative is professional and knowledgeable. The customer service representative is able to resolve the inquiry in one call, having answered the call quickly, without technology issues and without transferring the customer.

Table 4 – Call Center Experiences Relationship to Service Quality

The customer service representative…	Importance Rating (1 – 7)	Standard Deviation
is easy to understand, polite, professional, courteous & friendly.	6.92	.272
has technology that supports calls, and system issues don't slow down the process.	6.60	.610
is knowledgeable, has all the information needed and is able to answer your questions.	6.89	.317
does not transfer you or only transfer you once to someone who answers your question.	6.32	.839
resolves your issue in one call, the first time you call.	6.76	.465
answers the call quickly, promptly, in a reasonable amount of time.	6.73	.545

Technical Quality, Functional Quality, Overall Quality

The last set of survey questions were particular to the specific service encounter recently completed. The respondents were asked to evaluate their experience with this particular company in terms of service quality using a 5-point semantic differential scale, defined as 1 = Poor, 2 = Fair, 3 = Good, 4 = Very Good, and 5 = Excellent. This scale was used in order to be consistent with other surveys deployed by the focal organization. These customers rate the overall quality, technical quality (the service being done correctly) and functional quality (a satisfying interaction) very high (Table 5). We compared these average ratings using t-tests, and found no statistical difference between the ratings. We conducted exploratory factor analysis on these ratings to determine if these three constructs were evaluating different aspects of service quality. Using the three variables identified (overall quality, technical quality and functional quality) we found that they loaded onto a single factor that explained 86.94% of the variation inherent in the elements (Table 6).

Table 5 – Evaluation of Overall Quality, Technical Quality, and Functional Quality

How would you rate this company on…	Average Rating (1 – 5)	Standard Deviation
The *overall quality* of your recent call?	4.79	.410
The quality of the *service being done correctly* during your recent call?	4.81	.398
The quality of your *interaction being satisfying* during your recent call?	4.79	.410

Table 6 – Exploratory Factor Analysis of Customer Service Quality Evaluation

Total Variance Explained						
Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.608	86.942	86.942	2.608	86.942	86.942
2	.292	9.733	96.675			
3	.100	3.325	100.000			

Component Matrix[a]	
	Component 1
Overall Quality	.914
Done Correctly	.967
Interaction Satisfying	.914

[a] 1 components extracted
Extraction Method: Principal Component Analysis

DISCUSSION

Customers have clear ideas about what kinds of experiences should occur in a service encounter. We coded these expectations into four themes: provider attitude, resolve the issue, expedience, and genuine affect. Attitude, expedience, and affect align nicely with the broad concept of functional quality, defined as providing a positive interaction, or how the service is provided. Resolving the issue is just another way of saying meeting the output performance requirements: technical quality. Customers identify wait time measures and first call resolution as equally important measures of service quality, failing to prioritize technical or functional quality one above the other. How these various measures affect service quality was consistent with our expectations, with one exception: average handle time. The value-added nature of being directly engaged with the service provider to resolve the issue seems to eliminate the need for speed, as long as the issue is resolved through that single encounter. With a majority of the respondents indicating that longer interactions with the service provider actually increases service quality, a focus on resolution rather than call duration is appropriate for this process type. Our exploratory factor analysis established that from the customer perspective, technical quality, functional quality, and overall service quality are indistinguishable one from the other.

REFERENCES

[1] Aga, M. and O. V. Safakli (2007). "An Empirical Investigation of Service Quality and Customer Satisfaction in Professional Accounting Firms: Evidence from North Cyprus." Problems and Perspectives in Management 5(3): 84 - 98.

[2] Chen, C.-N. and S.-C. Ting (2002). "A Study Using the Grey System Theory to Evaluate the Importance of Various Quality Factors." The International Journal of Quality and Reliability Management 19(6/7): 838-860.

[3] Grönroos, C. (1988). "Service Quality: The Six Criteria of Good Perceived Service Quality." Review of Business 9(3): 10-13.

[4] Grönroos, C. (2001). "The Perceived Service Quality Concept - A Mistake?" Managing Service Quality 11(3): 150-152.

[5] Harvey, J. (1998). "Service Quality: A Tutorial." Journal of Operations Management 16: 583-597.

[6] Hume, M., G. Sullivan Mort, et al. (2006). "Understanding Service Experience in Non-profit Performing Arts: Implications for Operations and Service Management." Journal of Operations Management 24: 304-324.

[7] Parasuraman, A., V. A. Zeithaml, et al. (1988). "SERVQUAL: A Multiple-Item Scale for Measuring Consumer Perceptions of Service Quality." Journal of Retailing 64(1): 12 - 40.

[8] Schmenner, R. (1986). "How Can Service Businesses Survive and Prosper?" Sloan Management Review 27(3): 21-33.

CRITICAL SUCCESS FACTORS FOR SHARED SERVICES: A RESEARCH AGENDA

Shouhong Wang
Charlton College of Business, University of Massachusetts Dartmouth
Dartmouth, MA 02747-2300 USA swang@umassd.edu

Hai Wang
Sobey School of Business, Saint Mary's University
Halifax, NS B3H 2W3 CANADA hwang@smu.ca

ABSTRACT

Shared services have been widely spread in the government and private sectors. This research-in-progress article proposes a research framework for future studies on critical success factors of shared services in the aspects of strategy identification, collaborative partnership networking, optimal shared services process re-designing, and new policies and regulations.

1. INTRODUCTION

Shared services have been widely spread in private enterprises [10][21] and the government sector [6][28]. Shared service is the standardization and consolidation of common functions across multiple organizations to reduce information process duplication and increase information and knowledge sharing. The cross-organizational dimension of shared services makes up distinctive characteristics in contrast to other contemporary management practices. Many organizations have discovered that implementing shared services requires tremendous organizational support to make shared services workable for their specific situations [20].

Services for common functions in individual organizations can be shared to reduce business process duplication and increase knowledge sharing through standardization and consolidation of these service processes. Generally, accounting and financial management, human resources management, acquisition transactions, and customer relation management are the designated lines of business processes for shared services. Shared services are often mistakenly implemented as outsourcing. In fact, the differences between shared services and outsourcing are significant in many aspects [22]. Generally, shared services bring in long-term stable competitive advantages, while outsourcing involves much uncertainty [8]. In terms of general organizational structure, the shared service center, the provider of shared services for the partner organizations, is formed and governed by the partners, while the relation between the service provider and the outsourced firm in outsourcing is bilateral. The major strategies behind shared services are long-term cost saving and knowledge sharing beyond sourcing business processes for the short term. Effective shared services are achieved by standardization of processes to reduce process duplication across the entire shared service network. Human resource management, coordination of the shared service network, and risk sharing in shared services are

unique to outsourcing. Table 1 shows a comparison of shared services and outsourcing.

Managerial Aspects	Shared services	Outsourcing
General organizational structure	Networking with shared service center	Bilateral relation with the outsourced firm
Strategies	Long-term stable cost saving and knowledge sharing	Cost saving for the short term
Risk sharing among partners	Yes	No
Coordination of process re-engineering and standardization of processes	Yes	No
Reduce process duplication	Yes	No
Knowledge sharing among partners	Yes	No
Central control (e.g., policies)	Yes	No
Leverage of information technology	Yes	No
IT human resource changes	Re-deployment and training	Staff reduce

TABLE 1. Shared Services vs. Outsourcing

The importance of organizational support to better fit the changing environment of competition is a familiar theme in the organizational theory field [25]. Nowadays people need a research framework to fully understand shared services. Given the unique characteristics of shared services, research [26] has suggested that strategy of shared services management, collaborative partnership network design, optimal service process design, and policy and regulation design are the major aspects of organization support for shared services. Others aspects of organizational support, such as leadership, control span, and culture could be crucial for the success of shared services in general, but are secondary to these major aspects. This research-in-progress article discusses shared services from the organizational perspective, and proposes a research agenda by developing a set of hypotheses for future studies on critical success factors of organizational support in these aspects for shared services.

2. RESEARCH HYPOTHESES DEVELOPMENT

2.1. Strategies of Shared Services

In general, two types of strategic opportunities can be counted on as a result of shared services. One is cost saving through business process reengineering [2]. By sharing common non-core business functions such as accounting, financial transactions, human resource management, and customer services would dramatically improve business performance of all shared services partner organizations. Another promising advantage of shared services is information and knowledge sharing. For instance, shared distribution services allow the partners to share dependable marketing information [9]. Briefly, goals related to shared services include (1) cost

saving of business processes and improve services in the long run; (2) building a long-term strategic alliances with other organizations to share information and knowledge; and (3) establishing leadership through a focus on core functions.

In general, there are two methods that can be used to define the strategy for the organization [4][5][16]. One is a top-down value chain analysis [19]. Value chain analysis provides a framework for information technology planning and strategy formulation. This framework is appropriate for shared services since it is useful for assessing the values of shared services. In performing value chain analysis, an organization can define its strategic direction by pursuing shared services such as lowering transaction costs, providing more services to customers, and share knowledge with the shared services partners. The second method is bottom-up non-core function analysis to identify potential benefit and risk of sharing these non-core functions with other organizations.

Hypothesis 1: Organizations' clear visions of strategies for shared services positively contribute to corporate success in shared services.

2.2. Strategic Partnership in Shared Services

Research [7] has indicated that successful implementation of shared services requires new organizational forms of sourced service consortia organizations based on organizational analysis in understanding contracting, markets and the nature of cooperation. Shared services drive the partner organizations to form a network with the nucleus of shared services center. A shared service center facilitates the partner organizations and provides shared services and knowledge. Each individual partner organization is a stockholder of the center and acts as part of governance of the center. A shared service network is a virtual organization that actualizes value-adding, resource sharing, and risk sharing partnerships.

Organizations tend to perform only those functions that they do best and arranging for other non-core functions to be performed by other companies. The resulting collection of independent, mostly single-function business units forms a network organization [12][14]. The rational behind shared services is that organizations tend to do fewer things better and with less, to be strong competitors. Organizations perform only those core functions for which they have expert skills, and share with alliance organizations those non-core services that can be performed more economically while are governed by the shared services network.

Shared services create sophisticated networks that have all characteristics of inter-organizational networks as well as intra-organizational networks. The independent partner organizations of the shared services form an inter-organizational network. The network is governed by a board which is elected by the partner organizations. As the day-to-day operations of the shared services are coordinated by the shared service center, and the shared service center is in turn governed by the partner organizations, the center and its governance organizations assemble intra-organizational networks.

Partners and shared service center are the fundamental components for supporting a shared service network. Although little research on hidden costs of shared services network

construction has been done, it is certain that partner selection and shared service center formation are demanding and costly.

Shared services partners are long-term business alliances. A shared services partner must meet four criteria. First, the partner organization provides similar non-core functions that can be shared. Second, the partner can perceive the potential benefit of the shared services. Third, the partner is willing and capable to take the risk involved in the development of the shared services. Fourth, the partner possesses knowledge for the implementation of the shared services. In the organizational theory community there is growing agreement about the basic characteristics of the network organization [15][23]. To understand fully how networks are designed and operated, and where their applications lie, one must study organizational support in the shared services context.

Hypothesis 2: Organizations' long-term business partnership positively correlates with corporate success in shared services.

2.3. Business Process Optimization in Shared Services

Shared services would optimize the operational business processes at the shared service network level. Few research reports can be found in the literature on management of the environment of shared services; yet, workflow management [3][29] is considered a technique for organizations to manage the new technological environment of shared services. Workflow management can be used to re-examine shared services related business process and the relationships between business units. The flow of shared services determines the shared services structure, eliminating the duplicated processes for the network. Workflow management permits the partner organizations to share the understanding of the shared services, and to support the shared processes to meet the shared services strategies.

According to [11][13][17][18][24], every organized human activity gives rise to two fundamental and opposing requirements: the division of labor into various tasks to be performed and the coordination of these tasks to accomplish the activity. The two fundamental requirements can be fulfilled by optimal service process design in shared services. Optimal service process design includes optimal workflow design and optimal human resource design.

(1) IT enabled optimal workflow design
IT enabled optimal workflow design for shared services involves elimination of valueless business processes, negotiation for unique business processes, standardize the service processes, and knowledge sharing of business process reengineering. There are three levels of optimal workflow design. At the individual partner level, the service requirements are modeled in workflows based on maximum value adding. At the shared services network level, duplicated business processes are eliminated. At the shared service center level, all required workflows are optimized so that the services are provided at lowest costs.

Hypothesis 3a: Radical business process workflows re-design positively contributes to corporate success in shared services.

(2) Optimal human resource design

Human resource design becomes important more than ever for the organizations engaged in shared services. The shared service center employs people who are knowledgeable of the shared services. Hiring new skilled people, training the current employees to fit the new environment, and potential downsizing as a result of shared services are all new challenges for the organizations engaged in shared services.

Hypothesis 3b: Re-design of human resource structures positively correlates with corporate success in shared services.

2.4. New Policies and Regulation in Shared Services

Shared services create a special environment for partner organizations. The biggest issue in shared services is cost distribution among the partner organizations [1]. Mutual agreement on the cost distribution formula must be established for the shared service network. The bottom line for shared services is that every partner organization must receive a fair share of cost saving.

Shared services create two new sets of policy and regulation issues in addition to cost distribution. One is related to the funding and governance of the shared service center. The other is related to the rules of access of shared services. It is extremely important to realize that any bylaw of shared services does not guarantee the success of shared services, but the success of shared services must be backed up by a bylaw.

Organizational redesign process defines the course of action and requires varying amounts of improvisation [27]. New shared services introduce new policies and regulation for the partner organizations in the form of bylaws of the service network. The bylaws specify the rights and responsibilities of the partner organizations for the shared services. Generally, shared services create three categories of new policies and regulations.

(1) Governance of the shared service network and center

The configuration of shared service collaborative network is the summation of service operation network structure, negotiation network structure, and value-chain network structure. The shared service center is the coordinator of the collaborative network and represents the common interest of all partner organizations. The responsibility of a shared service center is to maintain and enhance the shared services and develop cohesion among the partners of the share services. The governance board of the shared service network exercises the charters of the board, acquires financial capital for the center, hires the general manager of the center, and monitors the operations of the shared services.

Hypothesis 4a: Strong governance board and service center positively contribute to corporate success in shared services.

(2) Cost distributions of the shared services

The shared service network must have a mutually agreed cost distribution scheme for the partner organizations. The distribution scheme allocates and reserves the shared resources.

Hypothesis 4b: Effective cost distribution scheme positively correlates with corporate success in shared services.

(3) Shared service access authorities and ethical codes
Stable and reliable shared services are always regulated by service access authorization rules and ethical codes related to the shared services.

Hypothesis 4c: Unambiguous ethical codes and access authorities positively contribute to corporate success in shared services.

A conceptual model that articulates the factors and their relationships in shared services is depicted in Figure 1.
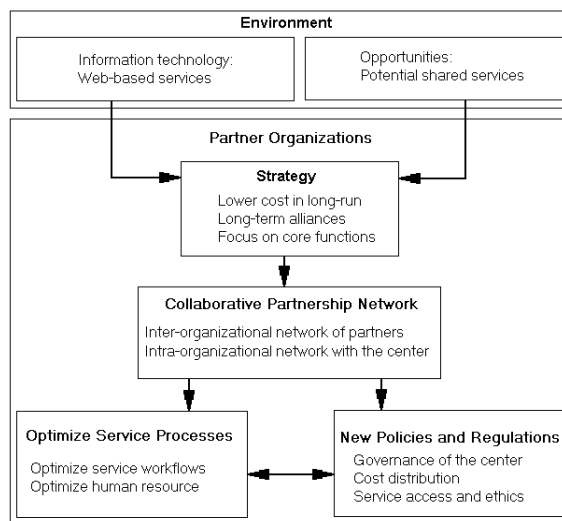
FIGURE 1. Conceptual Model of Shared Service Management

3. PROPOSED RESEARCH FRAMEWORK FOR FUTURE STUDY

In this section, we describe our plan of hypothesis tests to further investigate the critical success factors of organizational support for shared services in the aspects of strategy identification, collaborative partnership networking, optimal shared services process re-designing, and new policies and regulations. The research framework is depicted in Figure 2.

3.1. Proposed Variables

Dependent Variable – Tangible cost saving:
Estimated cost saving for all shared services partners contributed by shared services.

Dependent Variable – Perceived long-term competitive advantages:
Level of long term competitive advantages contributed by shared services.

Independent Variable – Clear vision of strategies of shared services:
Level of clarity of shared service strategies.
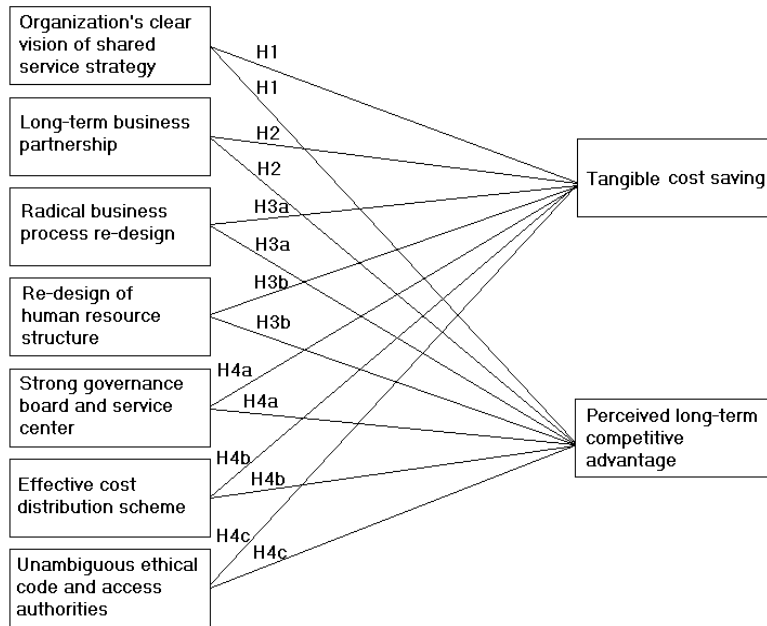
FIGURE 2. Research Framework of Critical Success Factors of Shared Services

Independent Variable – Long term business relationships among shared services partners:
Level of commitment on long term business relationships among shared services partners.

Independent Variable – Business process re-design:
Level of business process re-design for shared services.

Independent Variable – Human resource structure re-design:
Level of human resource structure re-design for shared services.

Independent Variable – Effective governance and service center for shared services:
Level of effectiveness and satisfaction of governance and service center for shared services.

Independent Variable – Effective cost distribution scheme:
Level of fairness and satisfaction of cost distribution scheme.

Independent Variable – Ethical code and access control for shared services:
Level of completeness and satisfaction of ethical code and access control for shared services.

3.2. Proposed Sampling and Research Procedure

A sample population will be examined in the sample selection process. Organizations will be selected from the following groups:
● Government organizations that are engaged into shared services.
● Large international business firms that have developed shared services.
● E-commerce companies that are using shared services.

The participants will be company senior executives or CIO who well understand shared services

in their organizations. Data will be gathered through a questionnaire. A structural outline of the questionnaire will be developed, including

● General Questions: Information about the organization.
● Questions related to estimation of cost saving contributed by shared services.
● Questions related to long term competitive advantages contributed by shared services.
● Questions related to shared service strategies.
● Questions related to long term business relationships with shared services partners.
● Questions related to IT enabled business process re-design in the organization for shared services.
● Questions related to human resource structure re-design in the organization for shared services.
● Questions related to effectiveness and satisfaction of governance and service center for shared services in the organization.
● Questions related to fairness and satisfaction of cost distribution scheme in the organization.
● Questions related to completeness and satisfaction of ethical code and access control for shared services in the organization.

An internal validity test of the questionnaire using the Cronbach Coefficient Alpha method will be completed. Statistical tests will be conducted on the relationships between the independent variable and the dependent variables which are pertinent to the hypotheses.

4. DISCUSSION

Outsourcing has been an important managerial solution to competition in the business field. However, outsourcing may not be a practicable option for the government and many enterprises. This is because significant cost saving is difficult to achieve through domestic outsourcing. Offshoring to other developing countries might also have many obstacles such as significant differences of social systems, cultures, and languages, as well as high risks. We believe that shared service can be another solution to competition for many organizations. Research into shared services is still in its infancy. This research-in-progress article has discussed the potential impact of shared services on organizations, and has proposed a framework for future research into shared services by proposing a set of hypotheses to investigate critical success factors for shared services. Our future study is to collect data and test these hypotheses.

REFERENCES

[1] Baron, O., D. Beyer, and G. R. Bitran, Pricing of shared computer services, *Journal of Revenue and Pricing Management*, 2005, 4(1), 54-65.
[2] Davenport, T. H., and J. Short, The new industrial engineering, *Sloan Management Review*, 1990, 31(4), 11-27.
[3] Ferreira, D. M. R., and J. J. P. Ferreira, Towards a workflow-based integration architecture for business networking, *Business Process Management Journal*, 2005, 11(5), 517-531.
[4] Galbraith, J. R., *Organization Design*. Addison-Wesley, Reading, MA, 1977.
[5] Galbraith, J. R., Designing the innovating organization, *Organizational Dynamics*, 1982, 10(3), 5-25.
[6] Janssen, M., and A. Joha, Motives for establishing shared service centers in public administrations, *International Journal of Information Management*, 2006, 26(2), 102-112.

[7] Kakabadse, A., and N. Kakabadse, Sourcing: New face to economies of scale and the emergence of new organizational forms, *Knowledge and Process Management*, 2000, 7(2), 107-118.

[8] King, D. R., Implications of uncertainty on firm outsourcing decisions, *Human Systems Management*, 2006, 25(2), 115-124.

[9] Krempel, M., Restructuring business for the new Europe, *Strategic Direction*, 2000, 16(5), 13-15.

[10] Leavell, H., International working capital management, *The Business Review*, 2006, 5(1), 233-239.

[11] Mackenzie, K. D., *The Organization Hologram: The Effective Management of Organizational Change*. Kluwer Academic Publishers, Boston, MA, 1991.

[12] Miles, R. E., and C. Snow, Organizations: new concepts for new forms, *California Management Review*, 1986, 28, 62-73.

[13] Mintzberg, H., *The Structuring of Organizations*. Prentice Hall, Englewood Cliffs, NJ, 1979.

[14] Mohrman, S. A., and T. G. Cummings, *Self-designing Organizations: Learning How to Create High Performance*. Addison-Wesley, Reading, MA, 1989.

[15] Mowshowitz, A., Virtual organization, *Communications of the ACM*, 1997, 40(9), 30-37.

[16] Nadler, D., and M. Tushman, *Strategic Organization Design: Concepts, Tools, and Processes*. Scott, Foresman and Company, Glenview, IL, 1988.

[17] Pasmore, W. A., *Designing Effective Organizations: The Sociotechnical Systems Perspective*. Wiley, New York, 1988.

[18] Pasmore, W. A., and S. Mlot, Developing self-managing work teams: an approach to successful integration, *Compensation & Benefits Review*, 1994, 26(4), 15-23.

[19] Porter, M. E., and V. E. Millar, How information gives you competitive advantage, *Harvard Business Review*, 1985, 63(4), 149-160.

[20] Rison, R. P., How to reduce the cost of HR and continue to provide value, *Human Resource Planning*, 2005, 28(1), 14-17.

[21] Rolia, J., L. Cherkasova, M. Arlitt and V. Machiraju, Supporting application quality of service in shared resource pools, *Communications of the ACM*, 2006, 49(3), 55-60.

[22] Sako, M., Technology strategy and management outsourcing versus shared services, *Communications of the ACM*, 2010, 53(7), 27-33.

[23] Snow, C. C., R. E. Miles and H. J. Coleman, Jr., Managing 21st century network organizations, *Organizational Dynamics*, 1992, 20(3), 5-20.

[24] Trist, E. L., The evolution of sociotechnical systems, in A. H. Van De Ven and W. F. Joyce, Ed., *Perspectives on Organization Design and Behavior*. Wiley, New York, 1982.

[25] Wang, S., Impact of information technology on organization, *Human Systems Management*, 1997, 16(2), 83-90.

[26] Wang, S., and H. Wang, Shared services beyond sourcing the back offices: organizational design, *Human Systems Management*, 2007, 26(4), 281-290.

[27] Weick, K. E., Organization design: organizations as self-designing systems, *Organizational Dynamics*, 1977, 6(2), 31-46.

[28] Williams, C. A., Advancing federal financial management, *The Journal of Government Financial Management*, 2006, 55(4), 52-57.

[29] WMC, Workflow Management Coalition, 2010, <http://www.wfmc.com> [accessed November 8, 2010].

OPERATIONAL INTEGRATION OF NEW SERVICES: VARIATIONS AND REACTIONS

Marie-Pierre Spooner, ESG Business School, University of Quebec in Montreal
spooner.marie-pierre@uqam.ca

In this turbulent era, many financial institutions are competing and developing new services to increase their market share, their customer base and outrun their competitors. The content of the new service is only one of the challenges. The ability to deploy it with rigor and speed in the financial institution's operational system is another challenge. Many new services are developed with an eye to marketing but with a limited understanding of the operational system. A majority of the new service development (NSD) litterature comes from marketing and considers the upfront phases of new services: planification and development. The implementation phase where the final integration of the new service is done is rarely explored. Therefore, it is important to deepen the understanding of this phase where variations often arise from the initial planification and development plans. Those variations trigger multiple reactions from managers: some of them incorporate modifications in order to follow the initial plan, others deny any changes or try to find a temporary patch to the processes creating even more variations, etc.

A qualitative study was chosen to explore this subject further. Two longitudinal studies were used to deepen the understanding of the mechanisms of variations and reactions during the operational integration of new services. The studies were held in two different financial institutions integrating two types of new services; a new product and a new process (Booz, Allen et Hamilton, 1982 ; de Brentani, 2001 ; Haksever *et al.*, 2000 ; Heany, 1983 ; Johne et Storey, 1998 ; Johnson *et al.*, 2000 ; Lovelock, 1984 ; Menor, Tatikonda et Sampson, 2002 ; Scheuing et Johnson, 1989 ; Tax et Stuart, 1997). Triangulation of data was obtained by the gathering of interviews, documents and observations. An approach inspired by grounded theory (Corbin et Strauss, 2008 ; Locke, 2001) was used to analyze the data. A first round of preliminary coding gave hundreds of codes pertaining to variations and reactions. A second round of coding, axial coding, limited the codes to categories refering to the types of variations and reactions specific to different phases of the operational integration: turbulence and dormancy. In order to relate the categories as well as their relationships, a systemic model of operational integration of new services in financial institutions will be proposed.

A STATISTICAL ANALYSIS OF MOTORCYCLE CRASHES IN MARYLAND

Mansoureh Jeihani [a]
Assistant Professor
Phone: 443-885-1873
Fax: 443-885-8324
Email: Mansoureh.Jeihani@morgan.edu

Gholamhossein Mazloomdoost [a]
Graduate Student
Phone: 443-885-3348
Fax: 443-885-8324
Email: mazloomdoost@gmail.com

a: Department of Transportation and Urban Infrastructure Studies
Morgan State University
1700 E. Cold Spring Lane, Baltimore, MD 21251

ABSTRACT

The goal of this study was to identify recurring or common road characteristics found at the site of Maryland motorcycle crashes. To achieve this, we integrated motorcycle crash and road inventory data from 1998 through 2007 obtained from Crash Outcome Data Evaluation System (CODES) and the Maryland State Highway Administration (SHA), respectively. In order to specify the minimum number of variables that explain most of the observed variance, a categorical principal component analysis was employed. In addition, ordinal logistic models were developed to estimate the number of motorcycle crashes for road segments within each road class.

Keywords: Crash Analysis, Safety, Statistics, Motorcycle, Categorical principal component analysis

INTRODUCTION

Commuting and recreational motorcycle use in the United States has been on the rise since the mid-1990s, with motorcycle registrations increasing 61 percent between 1996 and 2005 (NHTSA, 2006). As the number of motorcyclists increases, it is important that the safety issues associated with this mode of travel be addressed. Motorcycle riders and passengers are much more vulnerable to injury in crash situations. While crash fatalities decreased from 1990 to 1997, fatalities in the U.S. have increased every year for the past 10 years (NCHRP, 2008).

Motorcycle crashes are becoming more and more frequent in Maryland. Although some studies about different aspects of motorcycles crashes in Maryland have been performed and different crash data are collected, a comprehensive data set and a comprehensive engineering analysis of motorcycle crashes is missing. The goal of this study is to find the road characteristics that have significant effect on motorcycle crashes. The results of this study will help engineers and safety officials to develop solutions for this matter.

Different aspects of motorcycle crashes have been investigated in the literature. However, there are few studies about the engineering aspects of motorcycle crashes. Some of the studies related to motorcycle crashes are summarized below.

Cook (2009) investigated motorcycle crashes using NHTSA's Crash Outcome Data Evaluation System (CODES). There are sixteen states in the CODES data network, and it includes information from 83,527 motorcycle crashes from 2003 to 2005. Fifty-one percent of the motorcyclists involved in those accidents did not use helmets. Forty percent of all motorcyclists involved were admitted to a hospital. Helmeted motorcyclists were less likely to experience facial and head injuries than un-helmeted motorcyclists. The percentage of hospital treated cyclists that used helmet and suffered from a traumatic brain injury (TBI), was 14. Conversely, the percentage for cyclists with TBI that did not use helmet was 21 percent. A logistic regression analysis indicated that helmets significantly reduce head or facial injury, TBI, and in-hospital death.

According to Haque et al. (2008), motorcycles are only 19 percent of total vehicle population in Singapore, but they are involved in 54 percent of intersection crashes. The report also showed that motorcycles are more exposed at signalized intersections than other vehicles.

Zhang and Prevedouros (2005) conducted a web-based survey of 2,000 motorists to find the effect of rainy conditions on driver behavior. Drivers' perception of accident risk was higher in rainy weather, especially in heavy traffic, regardless of their age, gender, driving experience, education, and car type. Respondents stated that they drove 4.9 percent slower on wet roads and 11.1 percent slower when it was raining.

Garber and Kassebaum (2008) utilized fault tree analysis to identify the causal factors of all vehicle crashes on 143 two-lane highways in Virginia. Their data set included approximately 10,000 crashes from 2001 to 2004. The majority of these crashes were run-off-the-road (ROR). They found that the main causal factors for ROR crashes were curvature and annual average daily traffic (AADT). They recommended the development and implementation of a plan for correcting the geometric deficiencies of the significant causal factors at sites with high ROR crashes.

In order to find if existing training programs were reducing accident probabilities, Savolanien and Mannering (2007) estimated statistical models using a sample of Indiana motorcyclists in 2005. Statistical models for speed-choice and helmet usage behavior were also estimated. They concluded that motorcyclists who took the training course were more likely to be involved in an accident than those who did not.

Kostyniuk (2005) stated that the number of licensed motorcyclists of age 45-64 increased by 41 percent from 1997 to 2002 in Michigan. Although all vehicle crashes decreased by 7 percent and their associated fatalities decreased by 8 percent, motorcycle crashes increased by 20 percent and their fatalities increased by 27 percent in this period. Also, the crash rate per licensed motorcyclist was found to have increased by 32 percent. During this period, 45 percent of motorcyclists (per year) who were involved in a crash did not have a motorcycle driving license. Kostyniuk identified three major trends: 1) increase in the number of motorcycles, motorcyclists, motorcycle crashes and deaths 2) aging of the motorcyclist population 3) low level of motorcycle licensing. The author also showed that fast speed was the most frequently recorded hazardous action followed by clear distance, reckless driving, and careless/negligent driving. Kostyniuk indicated that the majority of crashes occurred on two-lane rural roads. It was also found that of the motorcycle crashes in Michigan from 1997-2002, most occurred on dry roads (89-94 percent), in good weather (73-81 percent), during the day (68-71 percent), and away from controlled intersections (70-72 percent). Kostyniuk suggested increasing the knowledge and skill of motorcyclists, including licensing and teaching other drivers how to drive near motorcycles.

Turner and Hagelin (2005) examined motorcycle trends in Florida before and after the helmet law change in 2002. The law permits cyclists over 21 years old not to use helmet if they have at least $10,000 insurance. Helmet use declined from 99.5 percent in 1998 to 52.7 percent in 2002. Helmet use was lowest among riders of cruiser-style motorcycles. While crash and injury rates per registered motorcycle and per motorcycle vehicle-mile-travel (VMT) declined after the helmet law change, fatal crash rates increased.

Green et al. (2008) analyzed traffic crash data in Kentucky and found that motorcycle crashes increased significantly in 2007 compared to the average of 2003-2006. The helmet usage dramatically decreased after the helmet requirement law was repealed in 1998. Also, the number of injury and fatal motorcycle crashes increased following the law change.

Rinde (1977) conducted a before-and-after study in California to find the relationship between shoulder width and crash rates. Three different shoulder widths were considered, and all of them showed a reduction in crash rates after widening.

In an attempt to improve traffic operators, the Minnesota Department of Transportation evaluated the statistical relationship between vehicular crashes and highway access (Preston, 1998). Preston found a relationship between access density and crash rates in 90 percent of the highway categories. The relationship explains why there is high access density in high crash rates sites near roadways and why there is low access density in low crash rate sites. The statistical analysis found that traffic volumes and traffic speeds do not have an effect on the crash rates, but roadway classes did have an impact on this analysis.

The U.S. General Accounting Office (2003) named human factors, roadway environment, and vehicle factors as the three factors that contribute to accidents. Of the three, human factor is the largest.

DATA INTEGRATION AND ANALYSIS

We joined two data sources, the Maryland Crash Outcome Data Evaluation System (CODES) and the road inventory databases, to analyze the motorcycle crashes in Maryland.

CODES is a project funded by the National Highway Traffic Safety Administration (NHTSA) to study motor vehicle crashes in conjunction with other healthcare databases. The CODES program resource center in Maryland, the National Study Center for Trauma/EMS at the University of Maryland in Baltimore, provided the motorcycle crashes data in the state from 1998 to 2007. This data included 14,434 data records.

The road inventory data presents all road characteristics in Maryland and includes more than 172,000 road segments. In order to conduct the engineering analysis, we performed a spatial joint of the road characteristics and the motorcycle crash data. The final database includes 6736 data points. Some characteristics of the joint data are explained as follows.

Accident Trends

Motorcycle registration in Maryland has nearly doubled from 2001 to 2008 (Figure 1). The number of motorcycle crashes has been increasing since 1998 and has almost doubled from 1998 to 2007. We calculated the crash rate as the number of crashes per number of registered motorcycles. Figure 2 shows that the crash rate decreased from 2001 to 2003, slightly increased from 2003 to 2005, and decreased from 2005 to 2007. The crash rate determined for 2007 was the lowest observed yet.
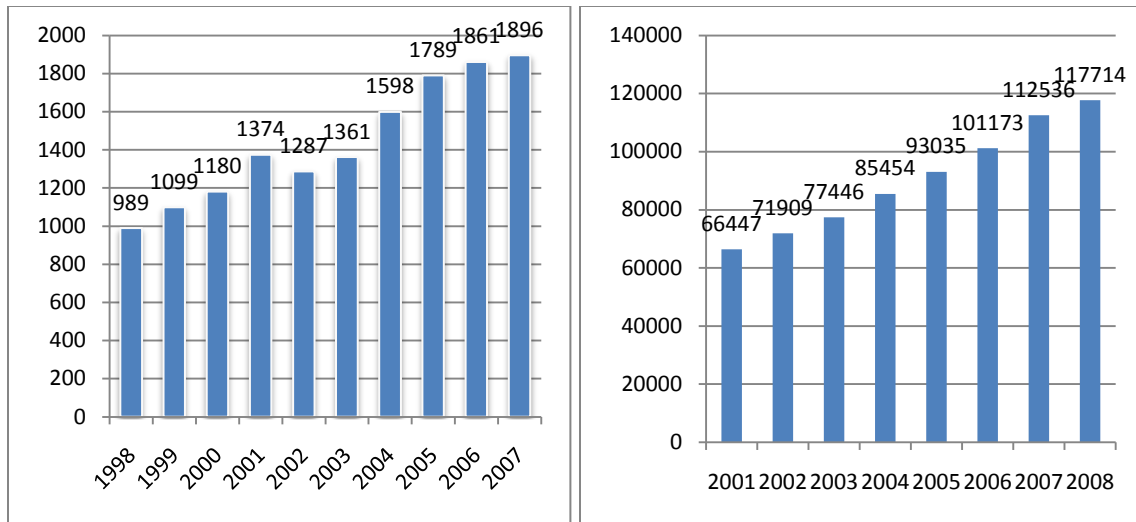
FIGURE 1- Motorcycle Crashes (left) and Registration (right) in Maryland, 2001-2008
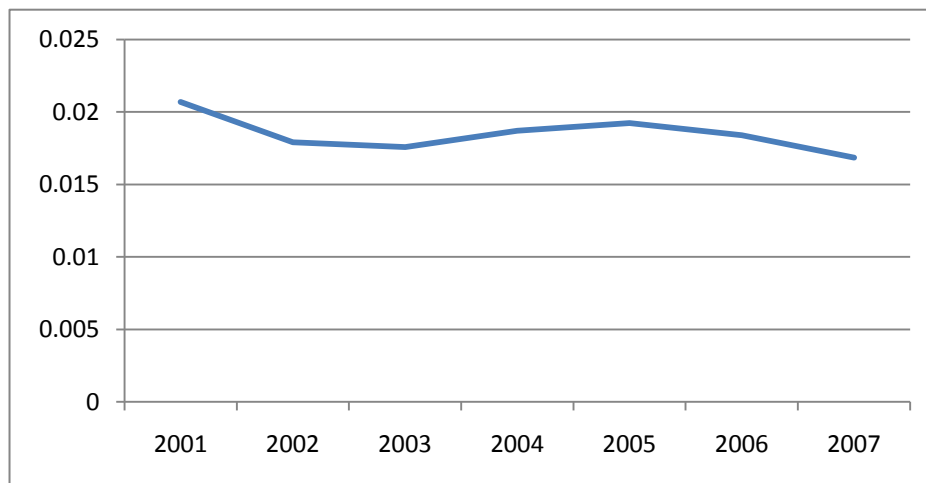
FIGURE 2- Motorcycle Crash Rate in Maryland, 2001-2007

Injury Severity

For this analysis, crash injury severity was categorized as fatal, injured, possibly injured, or permanently disabled. As presented in Figure 3, the percentage of injured motorcyclists more than doubled from 1998 to 2007. The fatality percentage has increased over these years as well. However, percentage of drivers disabled in crashes has fluctuated. Comparing these results with

the national fatality rate (Fatality Analysis Reporting System, FARS) indicate that Maryland's fatality rate is typical of national trends.
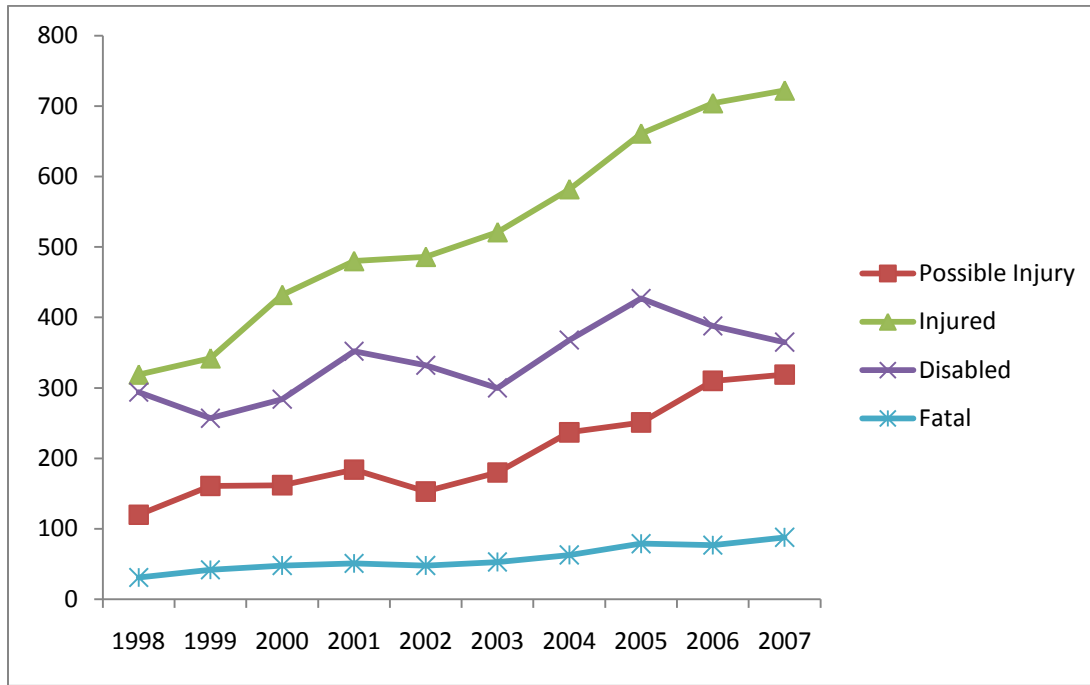
FIGURE 3- Injury Severity in Motorcycle Crashes in Maryland

A preliminary analysis of crash data indicates that most motorcycle crashes happened on state roads (61 percent) with no access control (84 percent) and a speed limit of 40-55 mph (50 percent). These roads are mostly urban roads (83 percent), non-divided (57 percent), two-way roads with two through marked lanes (71 percent) and no auxiliary lanes (91 percent).

Accidents typically occurred during the day (68 percent) when weather conditions were sunny/ cloudy (96 percent) and the road surface was dry (93 percent). The majority of crashes were single-vehicle collisions (42 percent) in which the motorcycle was moving straight in a constant speed (53 percent) far from an intersection (72 percent). The drivers, who were mostly 25-to-40 (41 percent) or 40-to-55 years of age (30 percent), men (93 percent), were in normal condition (79 percent) and using helmets (71 percent).

Prince George's County, Baltimore County, Baltimore City, Anne Arundel County and Montgomery County have had the highest percentage of the motorcycle crashes.

METHODOLOGY

After the preliminary analysis, we performed a statistical analysis on the combined dataset. A conventional statistical analysis would not produce acceptable results because the crash data does not have normal distribution, the dataset has a lot of categorical data, and many values are missing.   Therefore, we studied and tested different methods and software in order to make a reasonable analysis of the motorcycle crash data. We utilized categorical principal component analysis (CATPCA) to explain the current situation and to find the most important factors that caused motorcycle crashes in Maryland. We then used generalized linear model (GLM) to predict the number of accident on each road type. The variables in our model are as follows.

- Area Type (Rural/Urban): the accident happened on a rural or urban road
- Road Class: the road on which the crash happened is classified into four categories: Freeway & Interstate, Arterial, Collector, and Local
- Age: driver's age
- Sex: driver's gender
- Weekday: The day of accident. We grouped the data into 3 major categories: "Monday and Friday", "Tuesday to Thursday" and "Saturday and Sunday". We separated Monday and Friday from other weekdays, since they are the first and last day of the weekdays and drivers may have different behavior.
- Acc_Time: Time of accident. We categorized accident's time to "AM peak" which includes accidents from 7 am to 9 am, "Mid-day off peak" for 9 am to 4 pm, "evening peak" including accident between 4 pm to 7 pm, "evening off peak" for 7 pm to 11 pm and "night" for all accident happened during 11 pm to 7 am.
- Surf_Cond: Surface condition which is "Dry" and "Not Dry".
- Weather: weather condition which is defined as if it's "clear/cloudy" or "Not clear".
- Light: accidents took place in daylight or in dark
- Speed_Limit: speed limit
- GOV_Ctrl: Government Control, based on who controls and maintains each specific road segment. We classified it to "state highway", "county" and "others".
- Intersect: Y and N in intersect indicate if the accident took place at or near intersection or not.
- IRI: International roughness index
- Road_Div: road division is classified to: "not divided", "divided with barrier", "divided with no barrier" and "others" as separate category.
- AADT: average annual daily traffic
- Meadian_TY: type of the median is classified into seven groups: "curbed", "positive barrier", "unprotected", "none(undivided)", "center TLA (undivided)", "roundabout" and "painted".
- Meadian_W: Median Width

- Lt_Out_Sh: Indicating type of the left outside shoulder as: "curbed", "surface with concrete or bituminous" and "no shoulder or curbs".
- Rt_Out_Sh: Indicating type of the right outside shoulder as: "curbed", "surface with concrete or bituminous" and "no shoulder or curbs".
- Lt_In_Sh: Indicating type of the left outside shoulder as: "curbed", "surface with concrete or bituminous" and "no shoulder or curbs".
- Rt_In_Sh: Indicating type of the right outside shoulder as: "curbed", "surface with concrete or bituminous" and "no shoulder or curbs".
- Lt_Out_Sh_W: width of the left shoulders
- Rt_Out_Sh_W: width of the right shoulders
- Lt_In_Sh_W: width of the left shoulders
- Rt_In_Sh_W: width of the right shoulders
- Lane: number of through lanes

Factor Analysis and Categorical Principal Component Analysis

Factor analysis is a statistical method that identifies the underlying variables that explain the pattern of correlations within a set of observed variables. Factor analysis is generally utilized in variable reduction (i.e., to specify the minimum number of variables that explain most of the observed variance). Factor analysis can also be used to generate hypotheses regarding causal mechanisms or to screen variables for subsequent analysis.

Categorical principal component analysis (CATPCA) is factor analysis for categorical data that quantifies categorical variables while reducing the dimensionality of the data. This method reduces variables to a smaller set of uncorrelated variables (components) that represent most of information in the original variables. CATPCA is most useful when a large number of variables prohibit a reasonable relationship between the dependent variable and the independent variables. Due to the large amount of data, we utilized CATPCA to reduce the number of categorical variables used in the regression model. A summary of our analysis is presented in Tables 1a and 1b.

The component loading table (Table 1a) shows how much each variable is related to each dimension. The dimension is defined by the user. We tried different dimensions and concluded the best model was the one with 10 dimensions. The table is colorized so that one can easily follow the relationships with colors. Blue cells have the highest relation to their dimension.

The final grouping of variables as the result of factor analysis (optimal scaling) is presented in Table 1b. Age, median width, road class, road division, intersection, and rural/urban were placed in a separate category to be used as individual inputs for the regression because they did not correlate with any of dimensions. Also, no variable could be associated with category 10, so it was removed in the final table (Table 1b). Based on the results, two pairs of categories – Surface

Condition and Weather, and Light and Accident Time – essentially described the same thing. As a result, we used only one variable (e.g. Surface Condition, Light, etc.) from each pair in the regression analysis. The Left and Right Shoulder Types and Shoulder Widths were combined since left and right shoulders were always in the same group. Similarly, we used only Out-Shoulder-Type and Out-Shoulder-Width since they are similar to In-Shoulder-Type and In-Shoulder-Width; In-Shoulder-Type and In-Shoulder-Width were also missing many data points. Although AADT, median type, speed limit, and number of lanes are in the same group, we used them separately in the regression analysis because they are important factors in motorcycle crashes. Therefore, by using regression analysis we can find which of the above factors have more affect on the number of accidents.

TABLE 1A- Component Loadings

Column1	Dimension									
	1	2	3	4	5	6	7	8	9	10
Road_Div	-.502	.313	.117	.021	.321	.025	.051	-.074	-.081	.075
Lt_In_Sh	-.180	.802	.311	-.007	.053	-.003	.207	.274	-.013	-.022
Rt_In_Sh	-.179	.798	.322	-.005	.054	-.002	.207	.271	-.004	-.018
Intersect	-.164	-.247	-.017	-.057	-.003	.036	.057	.199	.472	-.526
IRI	-.144	-.306	-.363	-.045	.547	.124	.393	.151	-.136	.137
Light	-.055	.052	-.001	-.147	-.061	.760	-.188	.084	.067	-.044
Age	-.050	.153	-.101	.043	-.066	.272	.325	-.426	.361	-.399
Acc_Time	-.036	.015	-.017	-.102	-.063	.722	-.242	.195	-.024	.233
Sex	.000	.022	-.007	-.042	.035	-.111	.054	-.069	.755	.631
Weekday	.001	-.092	.043	-.059	.094	-.168	-.147	.641	.226	-.137
Surf_Cond	.004	-.027	-.032	.882	.053	.094	-.045	.059	.027	-.008
Weather	.038	-.033	-.028	.879	.047	.084	-.063	.072	.050	.022
Rural/Urban	.259	-.469	.084	-.066	.276	-.118	-.212	.241	.043	-.073
Lt_Out_Sh_W	.289	-.336	.664	.034	-.141	.084	.352	.019	-.041	.057
Rt_Out_Sh_W	.305	-.286	.661	.041	-.130	.088	.333	.026	-.019	.059
Road Class	.331	.447	.359	.034	-.065	-.038	-.261	-.129	.004	.029
Rt_Out_Sh	.375	.593	-.464	-.009	-.063	-.063	-.054	.009	.018	-.009
Lt_Out_Sh	.414	.610	-.431	.005	-.058	-.059	-.068	.024	.031	-.022
GOV_Ctrl	.429	.038	-.531	.002	.160	.145	.535	.123	-.052	.056
Median_W	.567	.039	-.096	.012	-.551	-.055	.090	.226	-.034	.043
Lt_In_Sh_W	.676	.263	.271	-.029	.465	.040	-.133	-.135	.050	-.080
Rt_In_Sh_W	.677	.256	.277	-.033	.459	.039	-.137	-.137	.061	-.080
Speed_Limit	.772	.395	-.012	.021	-.219	.034	.143	.048	.002	-.002
Lane	.830	-.376	-.050	-.023	.055	.029	-.019	.005	-.009	.009
Median_TY	.832	-.360	-.159	-.009	-.111	-.013	-.056	-.028	-.005	-.010
AADT	.896	-.030	.028	-.026	.180	.032	.007	.044	-.006	-.001

TABLE 1B- Variable Grouping

	1	2	3	4	5	6	7	8	9
Age	Lane	Lt_In_Sh	Lt_Out_Sh_W	Weather	IRI	Light	GOV_Ctrl	Weekday	Sex
Road Class	Lt_In_Sh_W	Rt_In_Sh	Rt_Out_Sh_W	Surf_Cond		ACC_Time			
Road_Div	Rt_In_Sh_W	Lt_Out_Sh							
Intersect	AADT	Rt_Out_Sh							
Rural/Urban	Median_TY								
Median_W	Speed_Limit								

Before conducting regression analysis, we rearrange our data to have number of accidents for each road segment. The road segments are defined by the SHA in the road inventory data. Our new data set contains 5308 data points. We use the number of accidents which is a discrete variable as the dependent variable in our regression model.

Ordinal Logistic Model

As stated earlier, most of our data are ordinal and categorical data. Also, the probability distribution of our dependent variable (number of accidents for each road segment) within each road class was found to fit into logarithmic function. Therefore, we used a special case of GLM, ordinal logistic model. The model assumes that there is a continuous outcome variable and the observed ordinal outcome results from discretizing the scale into j-ordered groups. The model is as follows.

$$\ln(\frac{\gamma_i}{1-\gamma_i}) = \frac{\theta_j - [\beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k]}{\exp(\tau_1 z_1 + \tau_2 z_2 + ... + \tau_m z_m)} \tag{1}$$

where, γ_j is the cumulative probability of the jth category, θ_j is the threshold for the jth category, β_is ($i= 1-k$) are the regression coefficients, x_is are the predictor variables, k is the number of predictors, τ_ls and z_ls ($l=1-m$) are coefficients for the scale components and predictor variables for the scale component. $\ln(\frac{\gamma_i}{1-\gamma_i})$ is the Logit function. The numerator presents the location of the model and the denominator presents the scale. z_ls are chosen from the same set of variables as the x_is. The scale variables improve the model when there are variables with large variances.

The regression was conducted on all roads as well as the four road classes: Freeway, Arterial, Collector and Local to predict number of accidents. Tables 2 to 6 present the regression results.

RESULTS

All Roads Combined

Based on the regression results for 5308 data records, area type (urban/rural), median type, speed limit, AADT, IRI, and number of through lanes for 90 percent confidence interval are important factors in motorcycle crashes on all roads. The coefficients of many other variables are not significantly different from zero (Table 2); however, removing them from the model would worsen the model. There is a higher probability in motorcycle crashes in rural area compare to urban area. Roads with speed limit 25 and 60 are more likely to have more motorcycle crashes than roads with other speed limits.

The probability of event j (number of accidents $=j$) can be calculated as:

$$\text{Prob}(\text{event } j) = \frac{1}{1 + e^{\frac{-(\alpha_j - \beta x)}{\exp(\tau z)}}} \tag{2}$$

As presented in Model Fitting Information of Table 2, the difference between the two log-likelihoods has a significant level of 0.000. Therefore the null hypothesis, that both the location parameters and the scale parameters are zero, is rejected. The three pseudo R^2: Cox and Snell (0.749), Nagellkerke (0.992), and Mc Fadden (0.983) are sufficiently high to demonstrate that the dependent variable (number of crashes) is significantly related to the independent variables (Cameron and Trivedi, 1998). Over 80 percent of the observed number of accidents fit the model based on predicted values. Accident numbers not included in this data set can be found through interpolation or extrapolation.

Arterial Roads

There are 2646 arterial road segments and the regression results indicates that road division, area type, AADT, shoulder width, median type, and the interaction between IRI and AADT, and IRI and shoulder width are important factors in number of motorcycle crashes on arterial road segments in Maryland.

As presented in Model Fitting Information of Table 3, the difference between the two log-likelihoods has a significant level of 0.000. Therefore the null hypothesis, that both the location parameters and the scale parameters are zero, is rejected. The three pseudo R^2: Cox and Snell (0.670), Nagellkerke (0.984), and Mc Fadden (0.971) are sufficiently high to demonstrate that the dependent variable (number of crashes) is significantly related to the independent variables.

TABLE 2- Regression Results for All Roads Combined

Model Fitting Information				
Model	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	1517.752			
Final	.000	1517.752	43	.000
Link function: Logit.				

		Estimate	Std. Error	Wald	df	Sig.	Interval Lower	Interval Upper
Threshold	[NUMBER_OF_ACCIDENTS = 1]	1.597	.607	6.919	1	.009	.407	2.786
	[NUMBER_OF_ACCIDENTS = 2]	3.163	1.044	9.178	1	.002	1.117	5.209
	[NUMBER_OF_ACCIDENTS = 3]	4.504	1.540	8.557	1	.003	1.486	7.522
	[NUMBER_OF_ACCIDENTS = 4]	6.636	2.425	7.492	1	.006	1.884	11.388
	[NUMBER_OF_ACCIDENTS = 5]	7.446	2.795	7.098	1	.008	1.968	12.923
	[NUMBER_OF_ACCIDENTS = 6]	8.129	3.130	6.747	1	.009	1.995	14.263
	[NUMBER_OF_ACCIDENTS = 7]	9.376	3.818	6.032	1	.014	1.894	16.858
Location	[AREA_TYPE=RURAL]	.522	.260	4.034	1	.045	.013	1.031
	[AREA_TYPE=URBAN]	0ᵃ	.	.	0	.	.	.
	[GOVT_CONTR=County]	-3.935	8.969	.192	1	.661	-21.513	13.644
	[GOVT_CONTR=Other]	-.421	.570	.546	1	.460	-1.539	.697
	[GOVT_CONTR=State Highway]	0ᵃ	.	.	0	.	.	.
	[MEDIAN_TY=NA]	1.035	.518	3.998	1	.046	.020	2.049
	[MEDIAN_TY=NONE (UNDIV)]	-.029	.236	.015	1	.902	-.492	.434
	[MEDIAN_TY=POSITIVE BARRIER]	.599	.426	1.972	1	.160	-.237	1.435
	[MEDIAN_TY=ROUNDABOUT]	.258	1.275	.041	1	.839	-2.240	2.756
	[MEDIAN_TY=UNPROTECTED]	0ᵃ	.	.	0	.	.	.
	[SH_OUT_TY=surfaced with concrete or bitu]	0ᵃ	.	.	0	.	.	.
	[SPEED_LIMI_AC=25]	2.862	1.303	4.826	1	.028	.308	5.415
	[SPEED_LIMI_AC=30]	-2.170	2.267	.917	1	.338	-6.613	2.272
	[SPEED_LIMI_AC=35]	-1.869	1.903	.965	1	.326	-5.599	1.861
	[SPEED_LIMI_AC=40]	-1.838	1.413	1.691	1	.194	-4.608	.932
	[SPEED_LIMI_AC=45]	-.351	.609	.332	1	.565	-1.544	.843
	[SPEED_LIMI_AC=50]	-.293	.424	.480	1	.488	-1.124	.537
	[SPEED_LIMI_AC=55]	.158	.302	.275	1	.600	-.434	.750
	[SPEED_LIMI_AC=60]	1.035	.453	5.215	1	.022	.147	1.923
	[SPEED_LIMI_AC=65]	0ᵃ	.	.	0	.	.	.
	SH_OUT_W_AC	.016	.028	.312	1	.576	-.040	.071
	THROUGH_LA_AC	-.125	.130	.932	1	.334	-.380	.129
	AADT_AC	.000	.000	3.376	1	.066	.000	.000
	MEDIAN_WACD	-.002	.002	.740	1	.390	-.007	.003
	ROUGHNESS_AC	.004	.002	2.991	1	.084	.000	.008
	[RoadDivision=Divided no barrier]	-.124	.276	.203	1	.653	-.665	.417
	[RoadDivision=Divided with barrier]	-.338	.307	1.213	1	.271	-.940	.264
	[RoadDivision=Not Divided]	-.079	.341	.054	1	.817	-.747	.589
	[RoadDivision=Other]	0ᵃ	.	.	0	.	.	.
Scale	[AREA_TYPE=RURAL]	-.660	.190	12.056	1	.001	-1.033	-.288
	[AREA_TYPE=URBAN]	0ᵃ	.	.	0	.	.	.
	[GOVT_CONTR=County]	.888	1.032	.742	1	.389	-1.134	2.911
	[GOVT_CONTR=Other]	-.003	.289	.000	1	.991	-.570	.563
	[GOVT_CONTR=State Highway]	0ᵃ	.	.	0	.	.	.
	[MEDIAN_TY=NA]	-2.027	.773	6.875	1	.009	-3.542	-.512
	[MEDIAN_TY=NONE (UNDIV)]	-.313	.163	3.660	1	.056	-.633	.008
	[MEDIAN_TY=POSITIVE BARRIER]	-.381	.237	2.588	1	.108	-.845	.083
	[MEDIAN_TY=ROUNDABOUT]	-.054	.474	.013	1	.909	-.983	.875
	[MEDIAN_TY=UNPROTECTED]	0ᵃ	.	.	0	.	.	.
	[SH_OUT_TY=surfaced with concrete or bitu]	0ᵃ	.	.	0	.	.	.
	[SPEED_LIMI_AC=25]	-4.155	5.215	.635	1	.426	-14.375	6.066
	[SPEED_LIMI_AC=30]	.586	.582	1.013	1	.314	-.555	1.727
	[SPEED_LIMI_AC=35]	.340	.545	.390	1	.532	-.727	1.408
	[SPEED_LIMI_AC=40]	.261	.392	.444	1	.505	-.507	1.029
	[SPEED_LIMI_AC=45]	-.045	.309	.021	1	.884	-.650	.560
	[SPEED_LIMI_AC=50]	-.144	.242	.354	1	.552	-.620	.331
	[SPEED_LIMI_AC=55]	-.220	.188	1.374	1	.241	-.588	.148
	[SPEED_LIMI_AC=60]	-1.593	.573	7.718	1	.005	-2.716	-.469
	[SPEED_LIMI_AC=65]	0ᵃ	.	.	0	.	.	.
	SH_OUT_W_AC	-.017	.021	.649	1	.420	-.058	.024
	AADT_AC	.000	.000	6.330	1	.012	.000	.000
	MEDIAN_WACD	.001	.001	1.824	1	.177	-.001	.004
	THROUGH_LA_AC	.162	.069	5.591	1	.018	.028	.296
	ROUGHNESS_AC	.000	.001	.127	1	.721	-.002	.003

TABLE 3- Regression Results for Arterials

Model Fitting Information				
Model				
	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	508.550			
Final	.000	508.550	35	.000

Link function: Logit.

		Estimate	Std. Error	Wald	df	Sig.	95% Confidence Interval	
							Lower Bound	Upper Bound
Threshold	[NUMBER_OF_ACCIDENTS = 1]	11.564	19.894	.338	1	.561	-27.427	50.556
	[NUMBER_OF_ACCIDENTS = 2]	15.284	28.268	.292	1	.589	-40.119	70.688
	[NUMBER_OF_ACCIDENTS = 3]	21.943	44.991	.238	1	.626	-66.237	110.124
	[NUMBER_OF_ACCIDENTS = 4]	27.818	60.364	.212	1	.645	-90.493	146.129
	[NUMBER_OF_ACCIDENTS = 6]	31.753	70.813	.201	1	.654	-107.037	170.543
Location	[AREA_TYPE=RURAL]	4.903	13.228	.137	1	.711	-21.024	30.830
	[AREA_TYPE=URBAN]	0ᵃ	.	.	0	.	.	.
	[GOVT_CONTR=County]	-15.398	50.691	.092	1	.761	-114.751	83.955
	[GOVT_CONTR=Other]	-5.505	15.221	.131	1	.718	-35.338	24.327
	[GOVT_CONTR=State Hi]	0ᵃ	.	.	0	.	.	.
	[MEDIAN_TY=NONE (UN]	-.188	.759	.061	1	.805	-1.674	1.299
	[MEDIAN_TY=POSITIVE]	-1.282	3.539	.131	1	.717	-8.218	5.653
	[MEDIAN_TY=ROUNDABO]	-2.434	6.763	.130	1	.719	-15.688	10.821
	[MEDIAN_TY=UNPROTEC]	0ᵃ	.	.	0	.	.	.
	[THROUGH_LA_AC=2]	10.292	17.624	.341	1	.559	-24.251	44.834
	[THROUGH_LA_AC=3]	6.217	13.421	.215	1	.643	-20.087	32.521
	[THROUGH_LA_AC=4]	6.709	13.255	.256	1	.613	-19.270	32.688
	[THROUGH_LA_AC=5]	3.120	15.370	.041	1	.839	-27.005	33.246
	[THROUGH_LA_AC=6]	6.971	13.535	.265	1	.607	-19.556	33.499
	[THROUGH_LA_AC=7]	0ᵃ	.	.	0	.	.	.
	AADT_AC	.000	.000	.128	1	.720	-.001	.000
	SH_OUT_W_AC	.295	.822	.129	1	.720	-1.316	1.905
	MEDIAN_WACD	-.003	.010	.111	1	.739	-.024	.017
	ROUGHNESS_AC * SH_OUT_W_AC	-.004	.011	.134	1	.714	-.025	.017
	AADT_AC * ROUGHNESS_AC	.000	.000	.137	1	.711	.000	.000
	[RoadDivision=Divided]	-2.722	7.440	.134	1	.715	-17.305	11.861
	[RoadDivision=Not Divi]	-2.495	6.846	.133	1	.716	-15.913	10.923
	[RoadDivision=Other]	0ᵃ	.	.	0	.	.	.
Scale	[RoadDivision=Divided]	1.678	.761	4.861	1	.027	.186	3.170
	[RoadDivision=Not Divi]	1.645	.786	4.383	1	.036	.105	3.185
	[RoadDivision=Other]	0ᵃ	.	.	0	.	.	.
	[AREA_TYPE=RURAL]	-1.721	.379	20.594	1	.000	-2.464	-.978
	[AREA_TYPE=URBAN]	0ᵃ	.	.	0	.	.	.
	AADT_AC	.000	.000	4.861	1	.027	.000	.000
	SH_OUT_W_AC	-.148	.066	5.020	1	.025	-.278	-.019
	ROUGHNESS_AC * SH_OUT_W_AC	.001	.000	6.980	1	.008	.000	.002
	AADT_AC * ROUGHNESS_AC	.000	.000	10.638	1	.001	.000	.000
	[THROUGH_LA_AC=2]	-1.606	2.620	.376	1	.540	-6.741	3.529
	[THROUGH_LA_AC=3]	.106	2.646	.002	1	.968	-5.080	5.292
	[THROUGH_LA_AC=4]	.359	2.549	.020	1	.888	-4.637	5.354
	[THROUGH_LA_AC=5]	.611	2.585	.056	1	.813	-4.456	5.678
	[THROUGH_LA_AC=6]	-.223	2.556	.008	1	.930	-5.233	4.787
	[THROUGH_LA_AC=7]	0ᵃ	.	.	0	.	.	.
	[MEDIAN_TY=NONE (UN]	-.224	.215	1.084	1	.298	-.646	.198
	[MEDIAN_TY=POSITIVE]	.075	.222	.115	1	.734	-.360	.510
	[MEDIAN_TY=ROUNDABO]	1.189	.543	4.799	1	.028	.125	2.252
	[MEDIAN_TY=UNPROTEC]	0ᵃ	.	.	0	.	.	.
	[GOVT_CONTR=County]	1.144	1.031	1.231	1	.267	-.877	3.165
	[GOVT_CONTR=Other]	-.186	1.592	.014	1	.907	-3.305	2.934
	[GOVT_CONTR=State Hi]	0ᵃ	.	.	0	.	.	.

Freeway/Expressway Roads

The regression results for 678 freeway road segments are presented in Table 4. The Model Fitting Information indicates that the null hypothesis, both the location parameters and the scale parameters are zero, is rejected. The three pseudo R^2: Cox and Snell (0.740), Nagellkerke (0.927), and Mc Fadden (0.841) are sufficiently high and demonstrate that the dependent variable (number of crashes) is significantly related to the independent variables.

No variable directly affected the number of motorcycle crashes on freeways, since this type of road has a very limited range of variables. For example, over 60 percent of freeways are divided, contain barriers, and have a speed limit of 55 mph.

Collectors

There are 1038 collector road segments and the regression results indicates that road division, area type, number of through lanes, IRI and the interaction between IRI and number of through lanes, and the interaction between IRI and shoulder width are important factors in number of motorcycle crashes on collector road segments in Maryland.

As presented in Model Fitting Information of Table 5, the null hypothesis that both the location parameters and the scale parameters are zero is rejected. The three pseudo R^2: Cox and Snell (0.690), Nagellkerke (0.945), and Mc Fadden (0.895) are sufficiently high and demonstrate that the dependent variable (number of crashes) is significantly related to the independent variables.

Locals

There are 705 local road segments and the regression results indicate that speed limit is the only significant factor in the number of motorcycle crashes on local road segments in Maryland.

As presented in Model Fitting Information of Table 6, the null hypothesis that both the location parameters and the scale parameters are zero is rejected. The three pseudo R^2: Cox and Snell (0.681), Nagellkerke (0.991), and Mc Fadden (0.984) are sufficiently high and demonstrate that the dependent variable (number of crashes) is significantly related to the independent variables.

TABLE 4- Regression Results for Freeways

Model Fitting Information				
Model				
	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	947.343			
Final	1.164	946.179	14	.000

		Estimate	Std. Error	Wald	df	Sig.	95% Confidence Interval	
							Lower Bound	Upper Bound
Threshold	[NUMBER_OF_ACCIDENTS = 1]	1.090	.000	.	1	.	1.090	1.090
	[NUMBER_OF_ACCIDENTS = 2]	.655	.000	.	1	.	.655	.655
	[NUMBER_OF_ACCIDENTS = 3]	.383	.000	.	1	.	.383	.383
	[NUMBER_OF_ACCIDENTS = 4]	-.107	.000	.	1	.	-.107	-.107
	[NUMBER_OF_ACCIDENTS = 5]	-.415	7.170	.003	1	.954	-14.468	13.638
Location	IRI	.015	.213	.005	1	.943	-.402	.433
	THROUGH_LA * IRI	-.003	.035	.006	1	.939	-.071	.065
	AADT_AC * LT_OUT_W_AC	.000	.000	.005	1	.942	.000	.000
	[MEDIAN_TY=NONE (UNDIV)]	-.115	2.139	.003	1	.957	-4.307	4.077
	[MEDIAN_TY=POSITIVE BAR]	1.114	176.123	.000	1	.995	-344.080	346.308
	[MEDIAN_TY=UNPROTECTED]	0^a	.	.	0	.	.	.
	AADT_AC	.000	.000	.013	1	.911	.000	.000
	THROUGH_LA_AC	.063	2.281	.001	1	.978	-4.408	4.535
Scale	AADT_AC	.000	.000	.019	1	.890	.000	.000
	THROUGH_LA_AC	-.173	1.348	.016	1	.898	-2.815	2.470
	ROUGHNESS_AC	-.019	.077	.061	1	.805	-.170	.132
	[MEDIAN_TY=NONE (UNDIV)]	-.125	1.841	.005	1	.946	-3.734	3.484
	[MEDIAN_TY=POSITIVE BAR]	-.373	74.538	.000	1	.996	-146.465	145.719
	[MEDIAN_TY=UNPROTECTED]	0^a	.	.	0	.	.	.
	THROUGH_LA * IRI	.004	.012	.101	1	.750	-.019	.026
	AADT_AC * LT_OUT_W	.000	.000	.000	1	.984	.000	.000

Link function: Logit.

a. This parameter is set to zero because it is redundant.

TABLE- 5 Regression Results for Collectors

Model Fitting Information

Model	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	1496.735			
Final	0	1496.735	54	0

Link function: Logit.

		Estimate	Std. Error	Wald	df	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Threshold	[NUMBER_OF_ACCIDENTS_1= 1]	702.436	1926.040	.133	1	.715	-3072.534	4477.406
	[NUMBER_OF_ACCIDENTS_1= 2]	782.921	1955.488	.160	1	.689	-3049.766	4615.608
	[NUMBER_OF_ACCIDENTS_1= 3]	852.924	2071.152	.170	1	.680	-3206.460	4912.308
	[NUMBER_OF_ACCIDENTS_1= 4]	930.880	2282.516	.166	1	.683	-3542.769	5404.528
	[NUMBER_OF_ACCIDENTS_1= 5]	1009.488	2562.496	.155	1	.694	-4012.911	6031.887
	[NUMBER_OF_ACCIDENTS_1= 6]	1089.078	2894.593	.142	1	.707	-4584.220	6762.375
Location	[RoadDivision=Divided no barrier]	-39.682	235.629	.028	1	.866	-501.506	422.143
	[RoadDivision=Divided with barrier]	-47.900	283.259	.029	1	.866	-603.078	507.277
	[RoadDivision=Not Divided]	-46.575	276.028	.028	1	.866	-587.581	494.431
	[RoadDivision=Other]	0[a]			0			
	[AREA_TYPE=RURAL]	7.826	47.895	.027	1	.870	-86.046	101.697
	[AREA_TYPE=URBAN]	0[a]			0			
	[GOVT_CONTR=County]	-1.152	33.588	.001	1	.973	-66.982	64.679
	[GOVT_CONTR=Other]	-3.638	39.880	.008	1	.927	-81.802	74.525
	[GOVT_CONTR=State Highway]	0[a]			0			
	[MEDIAN_TY=CENTER TLA (UNDIV)]	744.351	1966.665	.143	1	.705	-3110.242	4598.944
	[MEDIAN_TY=NA]	-236.261	8081.046	.001	1	.977	-16074.821	15602.298
	[MEDIAN_TY=NONE (UNDIV)]	794.141	2010.844	.156	1	.693	-3147.042	4735.323
	[MEDIAN_TY=PAINTED]	-233.587	9276.963	.001	1	.980	-18416.101	17948.927
	[MEDIAN_TY=POSITIVE BARRIER]	746.696	1966.042	.144	1	.704	-3106.675	4600.067
	[MEDIAN_TY=ROUNDABOUT]	0[a]			0			
	[SH_OUT_TY=curbed]	21.148	125.192	.029	1	.866	-224.225	266.520
	[SH_OUT_TY=no shoulder or curbs]	-2.608	22.946	.013	1	.910	-47.582	42.365
	[SH_OUT_TY=surfaced with concrete or	0[a]			0			
	MEDIAN_WACD_1	-.441	2.795	.025	1	.875	-5.919	5.036
	[SPEED_LIMI_AC_1=25.0]	-25.510	197.926	.017	1	.897	-413.438	362.418
	[SPEED_LIMI_AC_1=30.0]	-20.006	172.400	.013	1	.908	-357.904	317.892
	[SPEED_LIMI_AC_1=35.0]	33.632	223.035	.023	1	.880	-403.508	470.772
	[SPEED_LIMI_AC_1=39.8]	50.922	315.828	.026	1	.872	-568.089	669.934
	[SPEED_LIMI_AC_1=40.0]	4.441	117.536	.001	1	.970	-225.926	234.807
	[SPEED_LIMI_AC_1=45.0]	-67.640	439.457	.024	1	.878	-928.960	793.680
	[SPEED_LIMI_AC_1=50.0]	-.482	118.506	.000	1	.997	-232.750	231.785
	[SPEED_LIMI_AC_1=55.0]	0[a]			0			
	AADT_AC_1	.000	.002	.027	1	.870	-.004	.005
	THROUGH_LA_AC_1	-76.242	453.436	.028	1	.866	-964.960	812.476
	ROUGHNESS_AC_1	-.402	2.393	.028	1	.867	-5.091	4.288
	SH_OUT_W_AC_1	-8.172	50.944	.026	1	.873	-108.021	91.677
	THROUGH_LA_AC_1* SH_OUT_W_AC_1	5.246	31.799	.027	1	.869	-57.079	67.571
	THROUGH_LA_AC_1* ROUGHNESS_AC_1	.284	1.674	.029	1	.865	-2.996	3.564
Scale	[RoadDivision=Divided no barrier]	1.265	.542	5.446	1	.020	.202	2.327
	[RoadDivision=Divided with barrier]	1.116	.597	3.486	1	.062	-.055	2.287
	[RoadDivision=Not Divided]	1.218	.531	5.261	1	.022	.177	2.259
	[RoadDivision=Other]	0[a]			0			
	[AREA_TYPE=RURAL]	.253	.139	3.283	1	.070	-.021	.526
	[AREA_TYPE=URBAN]	0[a]			0			
	[GOVT_CONTR=County]	.200	.273	.537	1	.463	-.335	.735
	[GOVT_CONTR=Other]	.095	.322	.086	1	.769	-.536	.725
	[GOVT_CONTR=State Highway]	0[a]			0			
	[MEDIAN_TY=CENTER TLA (UNDIV)]	-.707	5.481	.017	1	.897	-11.449	10.036
	[MEDIAN_TY=NA]	.051	7.308	.000	1	.994	-14.274	14.375
	[MEDIAN_TY=NONE (UNDIV)]	-2.301	5.518	.174	1	.677	-13.116	8.514
	[MEDIAN_TY=PAINTED]	2.539	8.686	.085	1	.770	-14.486	19.564
	[MEDIAN_TY=POSITIVE BARRIER]	-1.174	5.489	.046	1	.831	-11.933	9.584
	[MEDIAN_TY=ROUNDABOUT]	0[a]			0			
	[SH_OUT_TY=curbed]	-.115	.188	.375	1	.541	-.484	.254
	[SH_OUT_TY=no shoulder or curbs]	.213	.151	1.993	1	.158	-.083	.508
	[SH_OUT_TY=surfaced with concrete or	0[a]			0			
	[SPEED_LIMI_AC_1=25.0]	-.610	.779	.612	1	.434	-2.136	.917
	[SPEED_LIMI_AC_1=30.0]	-.463	.734	.398	1	.528	-1.902	.976
	[SPEED_LIMI_AC_1=35.0]	-.437	.740	.349	1	.555	-1.887	1.013
	[SPEED_LIMI_AC_1=39.8]	-1.024	.747	1.876	1	.171	-2.489	.441
	[SPEED_LIMI_AC_1=40.0]	-.451	.713	.400	1	.527	-1.848	.947
	[SPEED_LIMI_AC_1=45.0]	-.469	.866	.293	1	.588	-2.166	1.229
	[SPEED_LIMI_AC_1=50.0]	-.506	.713	.503	1	.478	-1.903	.892
	[SPEED_LIMI_AC_1=55.0]	0[a]			0			
	AADT_AC_1	.000	.000	.029	1	.864	.000	.000
	THROUGH_LA_AC_1	2.489	.918	7.358	1	.007	.691	4.288
	ROUGHNESS_AC_1	.016	.008	4.346	1	.037	.001	.032
	SH_OUT_W_AC_1	.368	.228	2.603	1	.107	-.079	.815
	MEDIAN_WACD_1	-.033	.033	.993	1	.319	-.096	.031
	THROUGH_LA_AC_1* SH_OUT_W_AC_1	-.207	.110	3.551	1	.060	-.422	.008
	THROUGH_LA_AC_1* ROUGHNESS_AC_1	-.009	.004	5.082	1	.024	-.016	-.001

TABLE 6- Regression Results for Locals

Model Fitting Information				
Model	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	85.648			
Final	.000	85.648	28	.000

Link function: Logit.

		Estimate	Std. Error	Wald	df	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Threshold	[NUMBER_OF_ACCIDENTS = 1]	-6.688	37.250	.032	1	.858	-79.697	66.321
	[NUMBER_OF_ACCIDENTS = 2]	29.319	82.335	.127	1	.722	-132.055	190.694
	[NUMBER_OF_ACCIDENTS = 3]	69.537	209.628	.110	1	.740	-341.326	480.400
Location	[GOVT_CONTR=County]	-219.334	1075.092	.042	1	.838	-2326.476	1887.808
	[GOVT_CONTR=Other]	88.059	298.170	.087	1	.768	-496.343	672.460
	[GOVT_CONTR=State Highway]	0^a	.	.	0	.	.	.
	[MEDIAN_TY=CENTER TLA (UNDIV)]	-14.000	50.450	.077	1	.781	-112.881	84.881
	[MEDIAN_TY=NA]	-385.509	2130.332	.033	1	.856	-4560.882	3789.864
	[MEDIAN_TY=PAINTED]	-204.529	1657.683	.015	1	.902	-3453.529	3044.470
	[MEDIAN_TY=POSITIVE BARRIER]	0^a	.	.	0	.	.	.
	ROUGHNESS_AC_1* SH_OUT_W_AC_1	.001	.027	.001	1	.976	-.052	.054
	[SPEED_LIMI_AC=25]	-98.379	331.084	.088	1	.766	-747.291	550.533
	[SPEED_LIMI_AC=30]	-95.397	319.610	.089	1	.765	-721.822	531.028
	[SPEED_LIMI_AC=35]	12.549	876.134	.000	1	.989	-1704.643	1729.741
	[SPEED_LIMI_AC=40]	-249.400	832.646	.090	1	.765	-1881.356	1382.555
	[SPEED_LIMI_AC=45]	-44.105	161.232	.075	1	.784	-360.113	271.903
	[SPEED_LIMI_AC=50]	0^a	.	.	0	.	.	.
	[SH_OUT_TY=curbed]	-20.223	64.631	.098	1	.754	-146.897	106.452
	[SH_OUT_TY=no shoulder or curbs]	8.905	34.759	.066	1	.798	-59.222	77.031
	[SH_OUT_TY=surfaced with co or bit]	0^a	.	.	0	.	.	.
	[RoadDivision=Divided no barrier]	17.740	60.345	.086	1	.769	-100.534	136.014
	[RoadDivision=Divided with barrier]	108.810	324.095	.113	1	.737	-526.404	744.024
	[RoadDivision=Not Divided]	4.205	39.182	.012	1	.915	-72.591	81.000
	[RoadDivision=Other]	0^a	.	.	0	.	.	.
	[AREA_TYPE=RURAL]	-9.982	36.864	.073	1	.787	-82.235	62.271
	[AREA_TYPE=URBAN]	0^a	.	.	0	.	.	.
Scale	IRI* SH_OUT_W	.000	.001	.163	1	.687	-.001	.001
	[SPEED_LIMI_AC=25]	1.987	1.096	3.289	1	.070	-.160	4.135
	[SPEED_LIMI_AC=30]	1.398	.824	2.879	1	.090	-.217	3.013
	[SPEED_LIMI_AC=35]	-2.278	2715.105	.000	1	.999	-5323.786	5319.230
	[SPEED_LIMI_AC=40]	2.896	1.186	5.966	1	.015	.572	5.220
	[SPEED_LIMI_AC=45]	1.496	1.197	1.562	1	.211	-.850	3.842
	[SPEED_LIMI_AC=50]	0^a	.	.	0	.	.	.
	[MEDIAN_TY=CENTER TLA (UNDIV)]	2.625	3.055	.738	1	.390	-3.362	8.612
	[MEDIAN_TY=NA]	4.726	4.659	1.029	1	.310	-4.405	13.857
	[MEDIAN_TY=PAINTED]	4.175	5.294	.622	1	.430	-6.202	14.551
	[MEDIAN_TY=POSITIVE BARRIER]	0^a	.	.	0	.	.	.
	[GOVT_CONTR=County]	1.722	2.350	.537	1	.464	-2.884	6.327
	[GOVT_CONTR=Other]	-1.672	1.110	2.271	1	.132	-3.847	.502
	[GOVT_CONTR=State Highway]	0^a	.	.	0	.	.	.

CONCLUSION

The preliminary data analysis indicates that motorcycle crashes have mostly occurred on State roads with no access control and speed limits of 40-55 mph. These roads are predominantly urban roads, non-divided, two-way roads, with two marked through lanes and no auxiliary lanes. Fifty percent of road segment crashes happened on arterials and nineteen percent on collectors.

Accidents often occurred during daytime hours under sunny/cloudy weather conditions and dry road surfaces. Most of the crashes have been single vehicle collision when the motorcyclist was moving straight in a constant speed far from an intersection.

Based on logistic ordinal regression analysis results we concluded that area type, median type, speed limit, AADT, IRI, and number of through lanes affect the number of motorcycle crashes on all road segments in Maryland. Interestingly, government control and shoulder type were shown to have no significant impact on the number of motorcycle crashes. For future road impact determinations, the number of motorcycle crashes for each road characteristic can be predicted using our developed ordinal logistic model.

In arterial roadways, road division, area type, AADT, shoulder width, median type, and the interaction between IRI and AADT, and IRI and shoulder width found to affect number of motorcycle crashes. Based on acceptance of location and scale variable coefficients, our data suggest that the number of motorcycle crashes could be reduced by widening of shoulders, implementing road divisions and improving IRI.

Using the model, we found speed limit to be the only significant factor in motorcycle crashes on local roads. A likely reason is that local roads have a wider range of posted speed limits than other road classes, causing speed to stand out among other variables. Furthermore, the incomplete status of data obtained for local roads makes it difficult to assess the impact of other variables in this analysis. No trend was evident when suggestions for speed improvement were considered. By examination of crash percentages alone, we found that over 74 percent of motorcycle crashes occur on undivided roads. This would indicate that the implementation of road divisions might decrease motorcycle crashes on local roads.

Conversely, road division, area type, number of through lanes, IRI, the interaction between IRI and number of through lanes, and IRI and shoulder width were found to be important in motorcycle crashes on collector roads. This road class is not usually controlled or maintained by SHA and has a wide range of IRI, shoulder width, and number of through lanes. Most of the accidents on this road class (80 percent) happened on undivided roads. Therefore, the implementation of road divisions would likely have the greatest impact on decreasing motorcycle crashes on collector roads. Our data also support the improvement of IRI and widening of road shoulders in pursuit of this goal.

In conclusion, we found that the implementation of road divisions, widening of shoulders, and improvement of the IRI would serve to have the greatest impact on motorcycle crash reduction for most roads.

REFERENCES

[1] Cameron, A. C. and Trivedi, P. K. Regression Analysis of Count Data. New York: Cambridge University Press, 1998.

[2] Cook, L. J. *Motorcycle Helmet Use and Head and Facial Injuries: Crash Outcomes in CODES Link Data*. Utah Crash Outcome Data Evaluation System, Report DOTHS811208, Salt Lake City, Utah, 2009.

[3] Fatality Analysis Reporting System (FARS), http://www.nhtsa.gov/people/ncsa/fars.html (Accessed June 2010).

[4] Garber, N. J., and Kassebaum, E. A. *Evaluation of Crash Rates and Causal Factors for High-Risk Locations on Rural and Urban Two-Lane Highways in Virginia*. Research Report VTRC 09-R1, Virginia Transportation Research Council, 2008, http://www.virginiadot.org/vtrc/main/online_reports/pdf/09-r1.pdf (Accessed May 2009).

[5] Green, E. R., Agent, K. R., Pigman, J. G. *Analysis of Traffic Data Crash in Kentucky (2003 – 2007)*. Research Report KTC-08-29/KSP2-08-1F, Kentucky Transportation Center, University of Kentucky, 2008.

[6] Haque, M. M., Chin, H. C., and Huang, H. L. *Examining Exposure of Motorcycles at Signalized Intersections*. Journal of the Transportation Research Board, 2008. 2048, 60-65.

[7] Kostyniuk, L. P. *Motorcycle Crashes: Michigan Experience*. TRB 2005 Annual Meeting CD-ROM, Transportation Research Board of the National Academies, Washington D.C., 2005.

[8] Preston, H., Keltner, D., Newton, R., and Albrecht, C. *Statistical Relationship between Vehicular Crashes and Highway Access*. Research Report MN-RC-1998-27, Minnesota Department of Transportation, 1998.

[9] Rinde, E.A. *Accident Rates vs. Shoulder Width: 2 Lane Roads, 2 Lane Roads with Passing Lanes*. Research Report, California Department of Transportation, Sacramento, CA, 1977.

[10] Sovolainen, P. T., and Mannering, F. L. *Additional Evidence on the Effectiveness of Motorcycle Training and Motorcyclists' Risk-Taking Behavior*. Journal of the Transportation Research Board, 2008, 2031, 52-58.

[11] SPSS User Guide. 2010. http://support.spss.com/ProductsExt/SPSS/ESD/17/Download/User%20Manuals/English/SPSS%20Categories%2017.0.pdf (Accessed May 2010).

[12] National Cooperative Highway Research Program (NCHRP). *A Guide for Addressing Collisions Involving Motorcycle*. 22 (500), 2008.

[13] National Highway Traffic Safety Administration (NHTSA). *Motorcycle Safety Program Plan.* 2006, http://www.nhtsa.dot.gov/portal/nhtsa_static_file_download.jsp?file=/staticfiles/DOT/NHTSA/TrafficInjuryControl/Articles/AssociatedFiles/MotorcycleSafety2006.pdf (Accessed July 2009).

[14] Turner, P. A., and Hagelin, C. *Motorcycle Helmet Use and Trends Before and After Florida's Helmet Law Change in 2000*. Journal of the Transportation Research Board, 2006, 1922, 183-187.

[15] U.S. General Accounting Office. *Highway Safety: Research Continues on a Variety of Factors that Contribute to Motor Vehicle Crashes.* Research Report GAO-03-436, 2003, http://www.gao.gov/new.items/d03436.pdf (accessed August 2009).

[16] Zhang, L. and Prevedouros, P.D. *Motorists Perceptions on the Impact of Rainy Conditions on Driver Behavior and Accident Risk*. TRB 2005 Annual Meeting CD-ROM, Transportation Research Board of the National Academies, Washington D.C., 2005.

DOES THE SHARES OF INSIDER OWNERS MATTER TO THE CORPORATE STRATEGIES AND PERFORMANCE?

Chi-Hsing Tseng, National Pingtung Institute of Commerce,
51 Min Sheng E. Rd., Pingtung 900, Taiwan, R.O.C.
886-8-7238700 Ext. 6215
e-mail: tseng@npic.edu.tw

ABSTRACT

The purpose of this research is to explore the relationships among shares of insider owners, corporate strategies, and performance.   The classification of insiders includes the directors and supervisors, large shareholders, and managers.   This research treats corporate strategies as intervening variables.   This paper adopted data from "2004-2008 Business Groups in Taiwan," edited by the China Credit Information Service.   After surveying secondary data from 266 core firms of the business group, this study found that the shares of insiders have significant relationships with performance.   As to the intervening effect, the shares of insiders will affect firms' performance through corporate strategic decisions.   This phenomenon indicates that insider shareholders will dedicate to improve corporate performance in order to maintain their own interests.   However, they tend to not adopt diversification strategies due to avoid investing risks.

Keywords: agency theory, insider ownership, ultimate controllers, strategy, performance

INTRODUCTION

In recent years, the agency problem has been the research focus on corporate governance, especial in emerging countries, because these developing countries still do not have well developed governance mechanism.   As an emerging market, Taiwan also possesses some distinct characteristics that the ownership structure of Taiwanese companies mostly concentrated in the hands of family members.   Because there is lack of external control in Taiwan, internal governance has been the major mechanism for Taiwanese companies [9].   Therefore, the shares of insider owners may play the key role of corporate governance.

In the literature, several studies have suggested that the ownership structure has important implications for corporate governance and performance [35] [33] [38].   Some strategic management literature have focused on agency theory and treated it as a link between ownership structure and performance (e.g., [28] [24] [15]).   To further clarify the

relationship between ownership structure and performance, some studies adopt a diversification strategy as an important intermediary variable (e.g., [4] [5] [19] [23]). The core argument of these research is that the large shareholders can pressure on managers to adopt less risk strategies to increase company performance. From the perspective of the agency theory, the major shareholders representing as agents have the influence power in corporate strategic decisions. As the variety of stockholders, the different identities of insider will tend to adopt various strategies according to their purposes or goals. The strategic choices also will further affect the firm's performance. However, there is still not come up with a consensus that whether the large-shares insider owners will enhance the company's performance or not. Therefore, it is worth exploring the influence of insider owners.

LITERATURE REVIEW AND HYPOTHESES

Insider Shareholders and Performance

According to Taiwanese Security Transaction Law, the directors and supervisors, large shareholders, and managers who own above 10% of a company's shares need to declare their stock exchanges every month. The purpose of this rule is to prevent any possible insider transaction. Therefore, this research treats directors and supervisors, large shareholders, and managers as insider owners.

There are diverse opinions regarding the relationship between insider shares and performance. The convergence of interest hypothesis suggests that the more concentrated shares the greater company value [21], because the large shareholders will be more likely to monitor the company's operation to avoid any managerial decisions violating shareholders' interest. The other empirical studies also support the convergence of interest hypothesis (e.g., [28] [20] [40] [41]). On the contrast, the entrenchment hypothesis provides the reverse idea [22]. This hypothesis suggests that when a company's directors and supervisors, and managers own larger shares, they tend to adopt decisions that benefit their own interests and harm other small shareholders'. Studying takeover market, Stulz [37] found that when the managers own higher shares, then they tend to deny the acquisition decision even it may increase the company value, because they don't want to take any possible failure risk. Fan and Wong [13] also obtained the same result that there is a negative relation between insider shares and performance. Some researchers doubt that there is a nonlinear relation between insider shares and performance, such as McConnell and Servaes [26], Hermalin and Weisbach [18], and Cho [8]. However, these studies still can't come up with a consensus.

According to agency theory, this research concludes that insiders who own higher shares will tend to choose the strategy that may gain more profit.   As long as the company has better performance, the insiders can obtain more benefits.   Therefore, insider shares will positively influence on the company's performance in the long run.   Thus, the Hypothesis 1 is proposed as below.

H1: The higher the insider shares the greater the company's performance.

The Intervening Effect of Strategy

Most studies suggest that large shareholders will adopt strategies with less risk to ensure their interests (e.g., [1] [18] [27] [36]).   Some strategic management literature focusing on agency theory have treated the strategy as a link between ownership structure and performance [28] [24] [15].   Therefore, this research will treat strategy as the intervening variable.

Reviewing literature relating to the strategy and performance, this study found that diversification and internationalization are the corporate strategies possess the most influence on performance [6] [11] [16] [30].

Some research have shown that insider shares has a negative relationship with diversification (e.g., [2] [7] [12] [25] [31]), because large shareholders would like to avert from diversification risk.   However, Shleifer and Vishny [34] proposed the reverse opinion that large shareholders would adopt a higher risk strategy to maximize business value.   Therefore, this research infers that insider shareholders tend to adopt less risk strategy.   On the other hand, insider shareholders will adopt those strategies that may create more business value.

The internationalization businesses may possess lower system risk than those domestic businesses [32].   In the 1980s, the amount of foreign direct investment by Taiwanese companies has grown enormously due to the production cost soaring.   Reviewing the internationalization experiences of Taiwanese corporations, Tseng [39] found the major reasons of internationalization are seeking for lower production cost, approaching the local market, and obtaining preempted advantages.   Some studies have also shown that a higher degree of internationalization may spread operation risk and further increase performance (e.g., [10] [14]).   Therefore, to compare with diversification strategy, internationalization may lower the system risk and create more company value.   Thus, this research proposes Hypotheses 2 and 3 as below.

H2: The higher the insider shares the greater the company's performance by adopting less diversification strategy.

H3: The higher the insider shares the greater the company's performance by adopting a higher internationalization strategy.

METHODOLOGY

Research framework

According to the research purposes and literature review, this study proposes a research framework as set out in Figure 1.
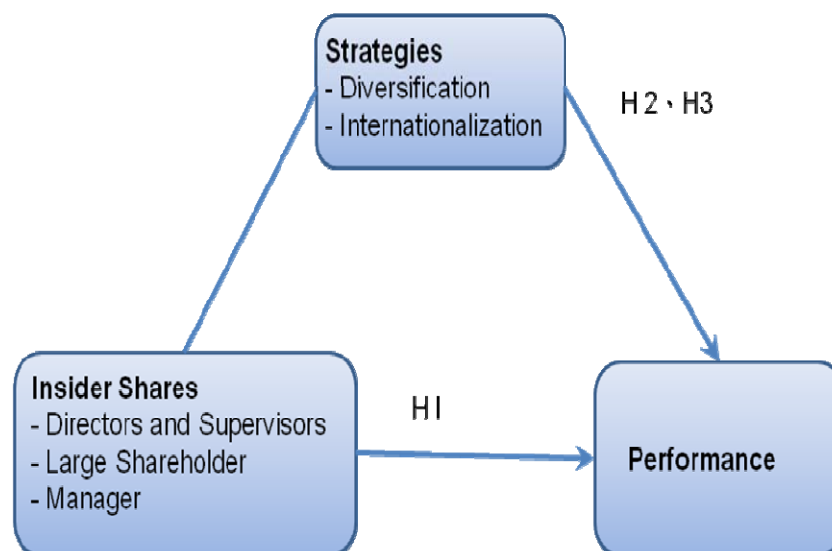

Figure 1.   Research Framework

Data

To explore the relationships among shares of insider owners, corporate strategies, and performance, this research adopted secondary data analysis.   Most of the Taiwan Stock Exchange Market （TSE） and Taiwan Over-The-Counter（OTC）listed firms are business groups, thus this research adopted the data from "2004-2008 Business Groups in Taiwan," edited by the China Credit Information Service.   But the name list of business groups is somewhat difference in every year.   This study only selected those core firms of business groups with relative complete data as research samples.   To obtain more complete data, this research also combined both database of China Credit Information Service and Taiwan Economy Journal.   After all, the number of qualified sample is 266.   Considering the

influence of insider shares may delay effect, this research adopted ownership data from 2005, strategies data from 2006, and performance data from 2007.

Independent Variable

According to Taiwanese Security Transaction Law, this research defined insider shareholders as the directors and supervisors, large shareholders, and managers who own above 10% of a company's shares. This research adopted the ratios of insider shares from 2005 database. All of the three insider shares ratios have positive skew distribution (statistic values are range from 1.40 to 3.07) and it represents that the long tail is to the right. This study adopted a logarithm transformation as a remedy for outliers and for the failures of normality. After data transformation, the skewnesses of insider shares ratios were reduced to nearly zero.

Intervening Variable

There are two kinds of strategies in this research. The diversification strategy is measured as the numbers of industry that the sample company had invested. The greater the numbers of industry represents the higher the degree of diversification. The distribution of diversification has a positive skew (statistic value is 2.74) and it should be adjusted by logarithm transformation. The degree of internationalization is measured by the numbers of foreign subsidiaries and the ratio of foreign sales. The sample company owns more foreign subsidiary, then its degree of internationalization is higher. When the ratio of foreign sales in total sales is higher, the sample company possesses a higher degree of internationalization. The distribution of foreign subsidiaries number has a positive skew (statistic value is 5.38) and this research also adopted a logarithm transformation to modified for failure of normality. After data transformation, the skewness of all intervening variables was decreased to nearly zero. As to the ratio of foreign sales, its distribution represents normality. There is no need to transform data further.

Dependent Variable

Company performance is measured by three different variables: net income ratio, return on assets, and sales growth. These financial performances are prevailed in relevant research, so this research adopted these indexes as measurement. As to the distribution of these indexes, all the skewnesses are near to zero, it represents the dependent variable is similar to normality. There is no need to transform data further.

Control Variable

As noted earlier, a firm's performance is a function of the operation of business strategy variables, such as size [17], industrial characteristics [29], and management traits.   Besides, industry has a direct effect on corporate ownership structures, which should be taken into account when measuring performance effects [38].   In order to achieve a true result with regard to the relationship among insider shares, corporate strategy, and performance, it is necessary to control these variables in the regression analysis.   As to the management traits, this study adopted family own and CEO internalized as measurements, because these two variables are available from the database and reveal the important characteristics of management teams.   Besides, these two variables reflect the distinct feature of companies in emerging market.   Therefore, this study proposes four control variables: industry, size, family own, and CEO internalized.

Differences in the intensity of competition and industry maturity may affect the level of profitability, growth, and even free cash flow.   To compare with other industries, electronic manufacturers should deal with more dynamic environment, so this research separate sample companies' industry into two parts.   A dummy variable represents the industry, with 1 for electronic manufacturers, 0 for other industries.   The Taiwan Stock Exchange Market has classified listed companies into different industries.   According to these industry codes, this research assigned 1 for those companies belong to the computer and peripheral equipment, electronic parts, and semi-conductor industries.   The total number of electronic manufacturers is 77, the others is 189.   Firm size is measured by average asset amount from 2003 to 2005.   Besides, the distribution of firm size has a positive skew (statistic value is 3.85), so this research adopted a logarithm transformation as a remedy for failures of normality.   Both the family own and the CEO internalized are dummy variables.   When the sample company belongs to family own, the family own was assigned 1 and the others were 0. The total number of family own company is 172 and the other is 94.   The Taiwan Economy Journal has defined the CEO internalized as those who possess final control power and belong to the family members.   If the CEO of the company met the requirement, then the CEO internalized was assigned 1.   The total number of CEO internalized company is 141 and the other is 125.

EMPIRICAL RESULTS

Table 1 shows the descriptive statistics and Pearson correlation coefficients for the study variables.   The table doesn't contain industry, family own, and CEO internalized, because

these control variables are dummy variables.   The correlation matrix indicates that both the shares of directors and supervisors, and managers are significantly negative related to diversification strategy.   The share of large shareholders is significantly negative related to both indexes of internationalization.   Besides, both of the shares of large shareholders and managers are significantly positive related to company performance.   As to control variable, firm size is significantly positive related to diversification, number of foreign subsidiary, net income ratio, and sales growth, while it is significantly negative related to directors and supervisors, and managers shares.

This study used a series of hierarchical regression analyses to test hypotheses.   The result is shown as Table 2.   All of the Variance inflation factors (VIF) values in regression models are below 2.5, it indicates that there is no significant sign of a multicollinearity problem.   To test Hypotheses 1, control variables were first entered, after which shares of insider were next entered as the main effect predictors of performance (see Models 1a, b, Models 2a, b, and Models 3a, b in Table 2).   When these independent variables were entered into the equation, there were significant increases in Model fit for Model 1, 2, and 3.   The share of large shareholders is significantly positive to net income ratio and return on assets; while the shares of directors and supervisors, and managers are significantly positive to sales growth.   This means that the larger the shares of insider the greater the performance of firms.   Thus, Hypothesis 1 is supported.

To test the intervening effect of strategy (Hypothesis 2 and 3), this research adopted the B-K approach [3].   This approach suggested that a series of regression models should be estimated.   One should estimate the three following regression equations: first, regressing the mediator on the independent variable; second, regressing the dependent variable on the independent variable, and third, regressing the dependent variable on both the independent variable and on the mediator.   Separate coefficients for each equation should be estimated and tested.   There is no need for hierarchical or stepwise regression or the computation of any partial or semipartial correlations [3].   Table 3 shows the first step that both shares of directors and supervisors, and managers are significantly negative on diversification strategy. However, there is no significantly relationship between shares of insiders and internationalization strategy.   In Table 2, the Models 1b, 2b, and 3b represent the results of the second step; the Models 1c, 2c, and 3c show the results of the third step.   According to the results, this research found that shares of insiders positively affect performance.   After entering intervening variables, there were significant increases in Model fit for Model 1 and 2. The standardized coefficients of diversification and number of foreign subsidiary are significantly higher than other independent variables.   To compare to the influence of insider shares, the diversification strategy possesses stronger negative influence on performance;

while the number of foreign subsidiary only positive affect net income ratio and return on assets.   Therefore, this research concludes that the higher shares of insiders will produce higher performance by adopting less diversification strategy.   Thus, Hypothesis 2 is supported.   As to the intervening effect of internationalization, this research only obtained partial effect.   Therefore, Hypothesis 3 is partial supported.

This research reviewed the influence of control variables and found that industry and firm size have more significant influence on performance and strategy.   However, the family own and CEO internalized do not possess an obvious influence on strategy.   The Table 2 shown that the family own company tends to have lower net income and CEO internalized company tends to produce lower sales growth.   This result suggested that family own and CEO internalized have a negative influence on firm performance, because these kinds of companies may lack of professional managerial skills, so they have lower performance. This study further proposes that managerial know might not prevail in emerging markets, so companies need to hire more professional managers to guarantee a long turn development.

Table 1.   Descriptive Statistics and Pearson correlation coefficients

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
Control variable										
1.Firm Size	1									
Independent Variable										
2.Directors and Supervisors	-.121*	1								
3.LargeShareholders	-.047	.007	1							
4.Managers	-.157**	-.050	-.097	1						
Intervening Variable										
5.Diversification	.577***	-.117*	-.035	-.286***	1					
6. Number of Foreign Subsidiary	.382***	-.076	-.144**	.016	.553***	1				
7.Ratio of Foreign Sales	-.006	.043	-.107*	.041	-.018	.003	1			
Dependent Variable										
8. Net Income Ratio	.184***	-.077	.131**	.032	-.013	.096	-.081	1		
9.Return on Assets	.070	.024	.110*	.142**	-.135**	.114*	-.015	-.675***	1	
10.Sales Growth	.185***	.097	-.051	.113*	-.018	.028	.030	.089	.295***	1
M	7.137	1.211	4.098	.270	.889	1.038	48.233	7.889	5.390	6.794
SD	.6477	.265	1.382	.290	.333	.465	37.571	16.010	8.291	25.230

Note. N=266

*p<.10; **p<.05; ***p<.01

Table 2. Summary of Hierarchical Regression Analysis for Performance (*N*=266)

Variables	Net Income Ratio			Return on Assets			Sales Growth		
	Model 1a	Model 1b	Model 1c	Model 2a	Model 2b	Model 2c	Model 3a	Model 3b	Model 3c
Control variable									
Industry	-.076	-.077	-.192**	.112*	.106	-.019	.043	-.049	-.093
Firm Size	.153**	.162**	.260***	.097	.135**	.218***	.243***	.302***	.391***
Family Own	-.109	-.115*	-.088	-.040	-.046	-.025	.043	.060	.071
CEO Internalized	.022	.010	.033	-.061	-.091	-.066	-.142**	-.199***	-.188***
Independent Variable									
Directors and Supervisors		-.057	-.088		.066	.032		.145**	.132**
Large Shareholders		.159**	.139**		.157**	.146**		-.044	-.056
Managers		.064	.020		.133*	.089		.251***	.217***
Intervening Variable									
Diversification			-0.354***			-.356***			-.201**
Internationalization									
Number of Foreign Subsidiary			0.220***			.262***			.044
Ratio of Foreign Sales			-0.076			-.009			.019
Model *F*	2.793**	2.752***	3.549***	1.619	2.438**	3.407***	4.223***	5.189***	4.182***
R^2	.045	.076	.134	.026	.067	.127	0.066	0.133	0.152
ΔR^2	.045*	.031*	.058***	.026	.041**	.060***	0.066***	0.067***	0.019

* p < 0.1; ** p < 0.05; *** p < 0.01

Table 3.　Standardized Regression Results for Strategy (*N*=266)

Variables	Diversification		Number of Foreign Subsidiary		Ratio of Foreign Sales	
	Model 1a	Model 1b	Model 2a	Model 2b	Model 3a	Model 3b
Control variable						
Industry	-.195***	-.166***	.250***	.252***	.010	-.030
Firm Size	.577***	.532***	.410***	.400***	-.003	.012
Family Own	.071	.067	.008	.011	.077	.096
CEO Internalized	.003	.038	-.042	-.033	.010	-.002
Independent Variable						
Directors and Supervisors		-.098**		-.002		.047
Large Shareholders		-.064		-.059		-.117*
Managers		-.150***		-.036		.060
Model *F*	40.480***	25.812***	16.918***	9.796***	.345	.836
R^2	.390	.420	.212	.216	.006	.023
$\triangle R^2$	.390***	.029***	.212***	.004	.006	.018

* p < 0.1; ** p < 0.05; *** p < 0.01

CONCLUSIONS AND IMPLICATIONS

The agency problem has been the focus in emerging markets and how to avoid internal transaction is an important issue for Taiwanese companies.　There is ample literature proposed that the ownership structure will affect the firm's performance.　However, there are still diverse suggestions on this topic and lack of literature for emerging markets.　This research explored the relationship among insider shares, strategy, and performance. Moreover, this paper considered some management traits of emerging markets as control variables, such as the family own and the CEO internalized.

After surveying secondary data from 266 core firms of the business group, this study found that the shares of insiders have significant relationships with performance; while the shares of insider have significant negative influence on diversification strategy.　However, there is no significant relationship between shares of insider owners and internationalization strategy. As to the intervening effect, the shares of insiders will affect firms' performance through corporate strategic decisions.　In summary, the higher the shares of insider, the higher the company's performance.　This study suggests that insiders tend to adopt less risk strategy, such as less degree of diversification, so that higher insider shares will lead to higher

performance.   Therefore, this research proposes that the higher shares of insider owners will not always lead to agency problem.   It depends on the goals of insider owners.   Besides, this research also found that family own and CEO internalized companies have lower firm performance.   Thus, this study suggests company should hire more non-family members as professional managers, especially in emerging market.

REFERENCE

[1]   Amihud, Y. & Lev, B. *Risk Reduction as a Managerial Mmotive for Conglomerate Merger*. Bell Journal of Economics, 1981, 12, 605-617.

[2]   Anderson, R., Bates, T., Bizjak, J. & Lemmon, M. L. *Corporate Governance and Firm Diversification*. Financial Management, 2000, 29, 5-22.

[3]   Baron, R. M. & Kenny, D. A. *The Moderator-mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Consideration*. Journal of Personality and Social Psychology, 1986, 51(6), 1173-1182.

[4]   Bergh, D.D. *Size and Relatedness of Units Sold*. Strategic Management Journal, 1995, 16(3), 221-240.

[5]   Bethel, J.E. & Liebeskind, J. *The Effects of Ownership Structure on Corporate Restructuring*. Management Journal Summer Special Issue, 1993, 14, 15-31.

[6]   Chang, S. C. & Wang, C. F. *The effect of Product Diversification Strategies on the Relationship between International Diversification and Firm Performance*. Journal of World Business, 2007, 42, 61-79.

[7]   Chen, S. S. & Ho, K. W. *Corporate Diversification, Ownership Structure, and Firm Value: The Singapore Evidence*. International Review of Financial Analysis, 2000, 9, 315-326.

[8]   Cho, M. H. *Ownership Structure, Investment, and the Corporate Value: An Empirical Analysis*. Journal of Financial Economics, 1998, 47, 103-121.

[9]   Chow, E.H., Chen, J.T., & Chen, K.H. *Family Business, Affiliated Groups, and the Value of Taiwanese Firms*. Journal of Financial Studies, 1996, 14(1), 115-139. (In Chinese)

[10] Daniel, J. D. & Jeffrey B. *Profit Performance: Do Foreign Operations Make a Difference*. Management International Review, 1989, 29(1), 46-56.

[11] Datta, D. K., Rajagopalan, N. & Rasheed, A. M. A. *Diversification and Performance: Critical Review and Future Directions*. Journal of Management Studies, 1991, 28(5), 529-558.

[12] Denis, D. J., Denis, D. K. & Sarin, A. *Ownership Structure and Top Executive Turnover*. Journal of Financial Economics, 1997, 45, 193-221.

[13] Fan, J. P. H. & Wong, T. J. *Corporate Ownership Structure and the Informativeness of Accounting Earning in East Asia*. Journal of Accounting and Economics, 2002, 401-425.

[14] Feng, C. M. (2006). *International Business Management: Constructing Global Advantage*. Chein Churn Culture.

[15] Gedajlovic, E. & Shapiro, D. *Management and Ownership Effects: Evidence from 5 Countries*. Strategic Management Journal, 1998, 19(6), 533-555.

[16] Grant, R. M., Jammine, A. & Thomas, H. *Diversity, Diversification, and Profitability among British Manufacturing Companies*, 1972-1984. Academy of Management Journal, 1988, 31(4), 771-801.

[17] Hall, M. & Weiss, L. *Firm Size and Profitability*. The Review of Economics and Statistics, 1967, 49: 319–331.

[18] Hermalin, B. E., & Weisbach, M. S. *The Effects of Board Composition and Direct Incentives on Firm Performance*. Financial Management, 1991, 20,101-112.

[19] Hoskisson, R.E., Johnson, J.R.A., & Moesel, D.D. *Corporate Divestiture Intensity in Restructuring Firms: Effects of Governance, Strategy and Performance*. Academy of Management Journal, 1994, 37(5), 1207-1238.

[20] Hudson, C., Jahera, J. & Lloyd, W. *Further Evidence on the Relationship Between Ownership and Performance*. The Financial Review, 1992, 27, 227-239.

[21] Jensen M.C. & Meckling, W. H. *Theory of the Firm: Managerial Behavior, Agency Costs, and Ownership Structure*. Journal of Financial Economics, 1976, 3, 305-360.

[22] Jensen, M. C. & Ruback, R. S. *The Market for Corporate Control: The Scientific Evidence*. Journal of Financial Economics, 1983, 11, 5-50.

[23] Lane, P.J., Cannella, Jr. A.A., & Lubatkin, M.H. *Agency Problems as Antecedents to Unrelated Mergers and Diversification: Amihud and Lev Reconsidered*. Strategic Management Journal, 1998, 19(6), 555-578.

[24] Li, M. & Simerly, R. *The Moderating Effect of Environmental Dynamism on the Ownership and Performance Relationship*. Strategic Management Journal, 1998, 19(2), 169-179.

[25] Lins, K. & Servaes, H. *International Evidence on the Value of Corporate Diversification*. Journal of Finance, 1999, 54, 2215-2239.

[26] McConnell J. J., & Servaes, H. *Additional Evidence on Equity Ownership and Corporate Value*. Journal of Financial Economics, 1990, 27(2), 595-612.

[27] Mikkelson, W. H. & Ruback, R. S. *An Empirical Analysis of the Interfirm Equity Investment Process*. Journal of Financial Economics, 1985, 14(4), 523-553.

[28] Oswald, S. L. & Jahera Jr. *The Influence of Ownership on Performance: An Empirical Study*. Strategic Management Journal, 1991, 12, 218-228.

[29] Porter, M.E. Competition Advantage. New York: The Free Press, 1985.

[30] Qian, G. *Multinationality, Product Diversification, and Profitability of Emerging US Small and Medium Sized Enterprises*. Journal of Business Venturing, 2002, 17(6), 611-633.

[31] Servaes, H. *The Value of Diversification during the Con-glomerate Merger Wave*. Journal of Finance, 1996, 51, 1201-1225.

[32] Shapiro, A. *Financial Structure and Cost of Capital in the Multinational Corporation*. Journal of Financial and Quantitative Analysis, 1978, 13, 211-266.

[33] Shleifer A. & Vishny R. W. *A Survey of Corporate Governance*. Journal of Finance, 1997, *52*(2), 737-783.

[34] Shleifer, A. & Vishny, R. W. *Large shareholders and corporate control*. Journal of Political Economy, 1986, 94, 461-488.

[35] Short, H. *Ownership, Control, Financial Structure and the Performance of Firms*. Journal of Economic Surveys, 1994, 8(3), 203-249.

[36] Smith J. C. & Stulz, R. *The determinants of firms' hedging policies*. Journal of Financial and Quantitative Analysis, 1985, 20, 391-405.

[37] Stulz, R. M. *Managerial Control of Voting Rights, Financial Policies and the Market for Corporate Control*. Journal of Financial Economics, 1988, 20, 25-54.

[38] Thomsen, S. & Pedersen, T. *Ownership Structure and Economic Performance in the Largest European Companies*. Strategic Management Journal, 2000, 21(6), 689-705.

[39] Tseng, C.H. *The Motivation of Taiwanese Firms' Direct Foreign Investment and Resources Based*. Bulletin of National Pingtung Institute of Commerce, 1999, 1, 37–55. (In Chinese)

[40] Vance, S. *Inside or Outside Directors: Is There Really a Difference*. Across The Board, 1995, 15-17.

[41] Yermack, D. *Higher Market Valuation of Companies with a Small Board of Directors*. Journal of Financial Economics, 1996, 40, 185-211.

Design of Green Supply Chains with Carbon Footprint Considerations

Navneet Vidyarthi
Department of Decision Sciences and MIS
John Molson School of Business, Concordia University
1455 De Maisonneuve Blvd. West, Montreal, QC, Canada, H3G 1M8
Email: navneetv@jmsb.concordia.ca

Satyaveer S. Chauhan
Department of Decision Sciences and MIS
John Molson School of Business, Concordia University
1455 De Maisonneuve Blvd. West, Montreal, QC, Canada, H3G 1M8
Email: sschuha@jmsb.concordia.ca

Rajesh K. Tyagi
HEC Montréal, 3000, chemin de la Côte-Sainte-Catherine
Montréal (Québec), Canada H3T 2A7
Email:rajesh-kumar.tyagi@hec.ca

ABSTRACT

This paper will discuss the impact of considering carbon emissions on the design of sustainable supply chain network. Traditionally, supply chain network design attempts to determine the location and the capacity of the suppliers, plants, and DCs as well as the allocation of the customers to DCs so as to maximize the long term economic performance of the company. In this paper, we present a supply chain network optimization model to support decisions that take into account both cost and carbon footprint. We analyze a variety of settings where the carbon footprint has major impact on the design and configuration of the supply chain. These models can be used by individual firms to reduce their individual carbon footprints (or to adhere to restrictions on carbon emissions) without compromising profitability.

APPLICATION OF ANP METHODOLOGY IN REVERSE SUPPLY CHAIN

Sharon M. ordoobadi, University of Massachusetts-Dartmouth,
285 Old Westport Road, Dartmouth, MA. 02747. sordoobadi@umassd.edu

ABSTRACT

The purpose of this study is to develop a decision tool to help managers with decisions regarding their reverse logistics policies and selection of the appropriate third party reverse logistics provider. Selection of a logistics provider is multidimensional in nature and interdependent relationships exist among various elements of the system. In addition, both quantitative and qualitative factors need to be considered in the evaluation process. The Analytic Network Process (ANP) methodology is capable of handling these requirements and thus is used in the development of our decision tool. Once potential logistics providers and selection criteria are identified by the decision maker several pair-wise comparisons are conducted to determine local priorities for the selection criteria/sub-criteria and alternatives. These priorities are then used to determine the overall priorities for rankings of the providers. The provider with the highest ranking is selected to perform the desired functions.

INTRODUCTION

Over the past several years, closed loop supply chain has gained considerable attention in industry and academia. The Original Equipment Manufacturers (OEMs) have no choice but to offer generous return policies to stay competitive and to comply with the environmental regulatory issues. Over the last few years a number of legislations, especially in European Union, have been introduced that focus on recycling (Ferguson & Browne, 2001). For example, The European Union has introduced the end of life vehicle (ELV) program that requires manufacturers to recycle auto bodies built since 2002. The European Directive on Waste of Electrical and Electronic Equipment (WEEE Directive) require that producers take care of the recovery of the products at the end of their useful lives. From these and numerous other examples it is obvious that the end-of-life product recovery is a major issue that manufacturers have to address. As a result many manufacturers have added the reverse supply chain to the traditional forward chain in order to manage the flow of products from consumers to the manufacturer. However, the question that remains to be addressed is whether the manufacturers should be involved in all activities required for reverse logistics or should seek the help of a third party reverse logistics (3PRL) provider.

Outsourcing to a third party is an option only if it proves to be both strategically and economically sound. If these conditions prevail then the issue of selecting the appropriate third party is addressed. A number of studies are conducted in the logistics field to determine whether outsourcing the reverse logistics functions is a sound decision

(Ordoobadi, 2009; Bolumole, 2000; Stank & Maltz, 1996; Skojett-Larson, 2000). The present research starts with the premise that outsourcing to a third party is justified both strategically and economically. Thus here we address the next issue namely how to select an appropriate 3PRL provider to perform the desired functions. This is accomplished by developing a decision tool for evaluation and selection of a 3PRL provider.

The rest of the paper is organized as follows. Background and research motivation is covered in section 2. Section 3 provides a detailed description of the model development. Finally, the paper concludes with summary in section 4.

II. BACKGROUND AND RESEARCH MOTIVATION

In today's competitive market manufacturers have no choice but to implement an effective policy for handling the product returns. Manufacturers and retailers have to offer a generous take-back policy to stay competitive. In addition, growing concern for the environment has led to increased interest and focus in closed loop supply chain over the past several years. As a result research on the importance of reverse logistics, closed loop supply chain, and product recovery has gained a lot of popularity in the last decade. Three major categories of research were uncovered in the literature: research on various issues related to the reverse logistics; use of a third party to perform reverse logistics activities; and development of decision models for reverse logistics decisions. The works done in each category along with corresponding author(s) are summarized below:

Reverse logistics issues	Author(s)
Benefits of reverse logistics	Wilding, 2004 Ferrer & Whybark, 2001
Reverse logistics activities and dimensions	Fleischmann *et al,* 1997
Product recovery management	Thierry *et al*, 1995 Toffel, 2004 Mollenkopf *et al.* 2007
Impact of return policies on manufacturers/retailers	Mukhopadhyay & Setoputro, 2004 Rogers & Tibben-Lembke, 2001 Blackburn *et al.* 2004
Reverse logistics and stages of product life cycle	Tibben-Lembke, 2002
Firm's position in supply chain and impact on reverse logistics decision	Tibben-Limbke & Rogers, 2002
Demand chain theory/management and their implications	Kosior & Strong, 2006

Complexity of supply chain management for
recoverable manufacturing systems

Guide, *et al.*, 2000
Inderfurth, K., 2005
Srivastava & Srivastava,
2006

Third party service providers and reverse logistics   **Author(s)**

Use of third party to perform reverse logistics   Aghazadeh, 2003
Sink & Langley, 1997

Research agenda for third party logistics   Selviaridis & Spring, 2007

Selection/evaluation of third party service provider   Meade & Sarkis, 2002

Identification and classification of criteria in   Qureshi *et al.*, 2008
evaluation of 3PL providers

Benefits of using third party logistics providers   Aktas & Ulengin, 2005

Effective methods for evaluations of third party
service providers   Boyson *et al*., 1999

Role of third party logistics provider in supply chain   Bolumole, 2001

Reverse logistics decision models   **Author(s)**

A prescriptive model based on satisficing principles   Boer *et al,* 2006

System dynamics study of closed loop supply chain   Kumar & Yamaoka, 2007

A conceptual model with five distinctive phases for
selection of third party logistics provider   Sink & Langley, 1997

A qualitative approach for identifying factors affecting
reverse logistics systems   Knemeyer, 2002

Closed-loop supply chain models   Savaskan *et al.*, 2004

Fuzzy approach for selection of the third party   Kannan, 2009
reverse logistics provider

The above literature points to the significance of reverse logistics. However, the research
on the selection of the appropriate 3PRL provider is few and far between. To address this
issue the present study is an attempt to develop a decision model that helps manufacturers
with evaluation and selection of the 3PRL provider. Such a tool is developed by first
identifying the selection criteria that are used in the evaluation of 3PRL providers. These
selection criteria are both qualitative and quantitative in nature and are interdependent. In
addition, the subjective judgment of the decision maker plays an important role for
comparing these criteria. Thus a methodology is required that can handle quantitative as

well as qualitative factors, allows for interdependencies among the selection criteria, and includes the decision maker's subjective judgments.

The ANP methodology is uniquely capable of meeting all the above requirements. As a result ANP is chosen as the methodology to be used in the development of the tool. Pair-wise comparisons are performed to determine the relative importance of the selection criteria. Local priorities are determined for each pair-wise comparison matrix and the overall priorities are calculated for ranking of the 3PRL providers. The provider with the highest ranking is then selected to perform the desired functions. The development of the proposed model is detailed in the next section.

III. DEVELOPMENT OF THE MODEL

Following steps are used in the development of the proposed model.
- Identify the selection criteria.
- Construct the ANP model
    - o Determine the overall objective.
    - o Create a hierarchy with overall objective of selecting the best provider at the highest level, the selection criteria/ sub-criteria at the second level, and the potential candidates at the lowest level.
    - o Form a hierarchical network to represent interdependent relationships between different levels of hierarchy as well as within each level of hierarchy.
- Determine rankings of the 3PRL providers.
    - o Perform pair-wise comparisons to determine relative importance of the selection criteria.
    - o Perform pair-wise comparisons to determine the relative contribution of the 3PRL providers to the selection criteria.
    - o Determine local priorities for the pair-wise comparison matrices.
    - o Determine the overall priorities for each provider.
    - o Rank the providers according to their overall priorities.
- Select the provider with the highest ranking

These steps are explained in detail in the following sections.

Identification of the selection criteria

Identifying the selection criteria for 3PRL provider is a challenging task as these criteria are often company and/or decision maker specific. However, to generate a list of general criteria that are used in the selection process we have searched the available literature. Numerous criteria for selection of 3PRL providers were uncovered in the reverse logistics literature. The selection criteria identified by the researchers (Boyson *et al.* 1999; Dowlatshahi, 2000; Lynch, 2000; Jeffery & Ramanujam, 2006; Langley *et al.* 2002; Khoo & Mitsuru, 2006; Kannan, 2009; Selviaridis & Spring, 2007) are:
- o Cost
- o Service quality
- o Remanufacturing capability
- o Shipment and tracking ability
- o Recycling capability

- Overall working relations
- Warehouse management
- Customer satisfaction
- Employee morale
- Reliability
- Financial stability
- Flexibility
- Responsiveness to requests
- Profitability
- Supply chain planning

After careful review of the literature and eliminating the duplicates we categorized the selection criteria into three major categories:

- *Operational capability criteria*: The criteria that are used to measure the ability of the 3PRL provider to perform reverse logistics functions.
- *Performance criteria*: The criteria that are used to measure the performance of the 3PRL providers.
- *Organizational impact criteria*: The criteria that are used to measure the impacts of the use of 3PRL provider on the organization.

Furthermore, we identified several sub-criteria within each category as listed below:
- *Operational capability criteria*: Collection, storage, inspection, sorting, Recycling, Remanufacturing, and after market introduction.
- *Performance criteria*: Cost, reliability, service quality, responsiveness to requests, and flexibility.
- *Organizational impact criteria*: Customer satisfaction, employee morale, and economic advantage (profitability).

This classification is used in the development of the proposed model. However, each individual organization/decision maker can customize this list to fit the specific situation under which the evaluation and selection process is performed. Once the selection criteria and sub-criteria are identified the next step is construction of a network that illustrates the hierarchies and interdependence relations among decision factors. This step is explained in the next section.

Construction of the ANP model

To construct the ANP model the following steps are completed:
1. Identify the overall objective.
2. Identify the appropriate selection criteria, and sub criteria that meet the company's needs.
3. Create a hierarchy with overall objective of selecting the best 3PRL provider at the highest level, the selection criteria at the second level, the sub-criteria at the next level, and the alternatives (potential 3PRL providers) at the lowest level.
4. Form a hierarchical network to represent relationships between different levels of hierarchy (alternatives, selection sub-criteria, and selection criteria) as well as the

interdependencies within each level of hierarchy (among selection criteria, among selection sub-criteria, and among alternatives). This network is used to determine the rankings of the 3PRL providers as explained in the next section.

Determination of the rankings of the providers

To determine the rankings of the potential providers following steps are completed:
1. Form pair-wise comparison matrices.
   - Perform pair-wise comparisons within each level of hierarchy.
   - Perform pair-wise comparisons to determine the relative contribution of the selection criteria to their immediate upper level objective.
   - Perform pair-wise comparisons to determine the relative contribution of the selection sub-criteria to their immediate upper level hierarchy.
   - Perform pair-wise comparisons to determine the relative contribution of the alternatives (3PRL providers) to their immediate upper level sub-criteria.
2. Calculate local priorities for selection criteria, selection sub-criteria and alternatives.
3. Use priorities calculated in the previous step to determine the overall priorities for each 3PRL provider.
4. Rank the providers based on their overall priorities.

IV. CONCLUSIONS

The main goal of the study was to provide decision makers with a tool to help them with their decisions regarding evaluation and selection of the third party reverse logistics providers. To start the process it was assumed that the use of third party reverse logistics provider was justified both strategically and economically. Thus the proposed model was developed to provide a systematic approach for evaluation and selection of the third party reverse logistics providers. In order to illustrate the interdependence among factors involved in the decision making process and inclusion of the subjective judgment of the decision maker the ANP methodology was applied in the development of the model. Selection criteria and sub-criteria as well as potential third party providers were identified. A network hierarchy was established to show the relationships between the selection criteria/sub-criteria and alternatives. Several pair-wise comparisons were performed at various levels of hierarchy to calculate the local priorities. These priorities were used to determine the overall weights for the potential providers. The providers were then ranked based on their overall weights and the provider with the highest ranking was selected to perform the reverse logistics functions.

References are provided upon request by contacting the author.

Game Theory Framework For B2B Channel Policy

Conway Lackman

International Consulting Group and Duquesne University

ABSTRACT

A game theory model is presented in which commissions (revenue sharing contracts) determine sales growth and profit distribution that emerge from negotiations between B2B supply chain parties, namely manufacturers (M) and distributors (D). Profit distribution and sales growth are determined simultaneously in this game and reasonably approximate B2B M-D bargaining. Profit distribution is an endogenous outcome of the bargaining process. A negative relation exists between profit inequality and sales growth. Profit distribution is the main determinant of supply chain composition (number of D's as a percent of the total supply chain members) .Key findings include identification of (1) channel efficiency conditions that drive agents to reach agreement quickly, (2) the impact of differential commission agreements on rates of sales growth, profit redistribution, and channel policy stability, (3) the impact of SBE conditions strategies in a subgame perfect Nash equilibrium including the dynamics of broken agreements in successive bargaining periods on channel policy stability and profits of the parties; (4) empirical determination of the impact of M and D target discount factors on channel policy stability and profits of the parties (5) the impact of supply chain composition on bargaining conditions for both parties and (6) both parties' rate of return under both rapid and slow sales growth situations.

Keywords: game theory in markets, marketing models, B2B marketing models

INTRODUCTION

Quantitative analysis in marketing is an important dimension of marketing knowledge as evidenced by the literature (Benidito, 1987; Binmole & Samuelson, 2001; Chaudri, 2005; Chussil and Reibstein, 1994; Jagpal, 1999; Kotler and Armstrong, 2008; Lehmann and Winer, 2008; Lilien, 1993; Lilien and Rangaswany, 1998; Palda, 1969). There is an emerging priority need for marketing managers to measure the financial consequences of marketing decisions, including channels.

Modeling of channels has been pursued in the literature in non-game theoretic frameworks. Gencturk and Aulakh (2007) found that different D relationships in the supply chain between U.S. M's and foreign D's operate in a norms-control base governance system and Brown, Dev and Lee (2000) examined three governance mechanism according to how well they mitigate opportunism in marketing channels applied to the hotel industry, but neither included a comprehensive set of input variables characterized by game-theoretic models or any B2B applications .

Dynamic game theory applications in marketing have been pursued without supply chain interactions. Ding (2007) took a step in this direction in a consumer behavior game theory model, but did not analyze channels Fruchter, (2008) presented a useful Nash game framework for price-quality competition at the final product level, but did not include channels in his model. Basu, Mazumdar and Rau (2007) deploy a Nash game to determine optimal price in multibrand choice at retail, but don't include channels.

Dynamic game theory applications that incorporate supply chain interaction include Jorgensen (1986), Eliashberg and Steinberg (1987) in a M –R (retailer) framework where M controls its production and transfer price and R controls its purchase rate and retail price. Jorgensen and Zaccour (2004) contains a chapter on differential game models of supply chains with focus on advertising and coordinating devices Kogan and Tapiereo (2007) extend a number of static supply chain games to a dynamic framework. Karray and Martin-Herran (2008) developed a three stage game theoretic model in a Stackelberg game, wherein national and private label brands compete pursuant to optimal price-advertising strategies in the channel. Amrouche, Martin-Herran, and Zaccour (2008) reframed this problem in a feedback Stackelberg game, but neither set of authors included commission structure, discount rates, and channel of distribution investment (measured by supply chain allocation) which Karray and Martin-Herran recommended as an area of future research. None of these models included B2B applications.

An important component of B2B marketing is supply chain analysis and management, In the literature, models of channels examined in game theoretic frameworks have largely ignored B2B and are missing important variables such as commission structure, discount rate, or channel allocation. Ouardighi, Jorgensoen, and Pasim (2008) , focus on M-R interaction in a game, but not in a B2B application. M controls product quality and advertising. R controls pricing where revenue sharing contracts (commissions) drive policy. Models in other frameworks that examined B2B channels (Gencturk and Aulakh, 2007, Brown, Dev and Lee, 2000) are also missing these important variables.   As a result, a rigorous framework is needed to elevate the descriptive and prescriptive power of current theory in the B2B channel application that captures

the channels strategy in a game theoretic framework as well as incorporating missing important variables such as commission structure, discount rate, and channel allocation. This study attempts to fill the gap represented by the scarcity of studies on financial consequences in B2B supply chains by incorporating the financial consequences of B2B channels policy in a Nash game theoretic framework that reflects the real world B2B bargaining among players in the supply chain The model developed here introduces several important variables previously not addressed in previous studies including commission structure, discount rates, and, channel of distribution investment (measured by supply chain investment allocation) and examines their impact on channel member bargaining. As Ding (2007) pointed out, qualitative theories are not very powerful in some complex situations and cannot provide precise insights. Moreover, they are not very effective as building blocks for quantitative theories based on agent decision making. This paper contends that game theory modeling of B2B channels can substantially add to the understanding of B2B channel behavior and improve channel decision making and policy, as well as measure the financial consequences of channel policy. To support that contention this article provides such a model.

This article will address two salient methodological issues in modeling the gaming/bargaining in a B2B marketing channel: (1) conceptualizing B2B M-D games in constructs convenient for game theoretical examination, and (2) selecting from the game theory "menu" a model that has potential for capturing the above constructs.

This paper is organized as follows: (1) the model structure is described; (2) the conditions for a sustainable bargaining equilibrium within the model are specified given the objective

functions of supply chain players (M and R); (3) an empirical estimation of the model's parameters are presented based on U.S. data; (4) limitations of the study are explained; (5) a summary and conclusions are presented; (6) managerial implications and (7) suggestions for future research are specified.

THE MODEL

The Nash Equilibrium (Nash, 1951) in infinite games with a continuous payoffs theorem is an intuitively logical starting point. The model presented here is an adaption of Chang's (1995) game construct selected and proponed as a robust explanatory model, called MM hereafter. To capture M-D bargaining in a Nash equilibrium oriented game, we deploy MM involving the D's provision of products and/or services to end user customers which affect profits and the growth of sales. This paper attempts to apply this game with a focus on Nash equilibria for two common B2B players: M's and D's. The focus of the game's objective function is profit distribution via D commissions. The traditional B2B game may be considered to be basically three games and a sub game. The focus is on the subgame consisting of two players per game each with a singular payoff/objective function.

In the MM, sales are driven by D commissions, which deter the M's' profits. The MM assumes that some or all of the commission revenues can be transferred to a class of agents (Ds) who use these revenues to increase their own profits. D's and M's have partly conflicting interests about commissions. How such conflict is resolved depends on the D structure and is the

subject of later sections.

Application of the MM includes a direct and intuitive relationship between channel policy, including commission policy, and sales growth. Time is discrete and indexed by $t = 0.1$. A closed system is specified, populated by two types of agents: M and D. Each class has a large number of identical agents. The number of D's as a fraction of the supply chain will be denoted by λ. The representative M owns amount k_t of sales called "capital." The sales M produces in the channel represents the "capital" invested by M in the channel. In period t, s/he can produce more capital according to a posited sales function:

(2.1) $$y_t = Ag_t^{\alpha}\kappa_{\tau}^{(1-\alpha)}$$

where y denotes sales, g_t denotes D's' expenditure on channel services at t and $A > 0$, $\alpha = (0,1)$. The sales function which incorporates some channel expenditures is important for sales. Think of g_t as channel infrastructure and assume that α is fairly small. The typical M takes, as given, the ratio of D's' expenditures to sales. That is $\theta_t = g_t/y_t$. In period t, the typical M must pay a proportional commission τ_t and decide how much capital to consume and to leave for the next period. The evolution of capital is given by:

(2.2) $$k_{t+1} = k_t + (1-\tau_t)\, y_t - c_t = R_t k_t - c_t,$$

where k_t denotes capital investment in the channel, τ_t denotes commission rate, y_t denotes the M's sales via (2.1) and the definition of α_t, and, c_t denotes cost,

(2.3) $$R_t = 1 + (1-\tau_t)\, A^{1/(1-\alpha)}\theta_t^{\alpha/(1-\alpha)}.$$

R_t is M's (gross) after commission rate of return on investment. The rate of return in period t is determined by the parameters θ_t and τ_t. A stochastic process $\{(\theta_t, \tau_t)\}\infty_{t=0}$ will be called a

channel policy. M' preferences are described by:

(2.4) $\qquad E[\Sigma^{\infty}_{t=0} \beta^t u(c_t)] = E[\Sigma^{\square}_{t=0} \beta^t \underline{{}^c t^{1-\theta}}]$

$$1-\sigma$$

where $0 < \beta < 1$ and $\sigma > 0$, $\sigma \neq 1$ where σ is the inverse of the elasticity of intertemporal

substitution and $E(.)$ denotes expectation. M's problem is to maximize (2.4) subject to (2.2) and

(2.3), given a channel policy and the initial quantity of capital.

The rest of the sales are specified to make the analysis as simple as possible. In each

period, the M transfers channel commissions, (the difference between revenues and channel

expenditures) to the Ds. Ds do not have another source of profit and cannot borrow. Each D's

commission and sales is given by:

(2.5) $\qquad c^w_t = (1-\lambda)(\tau_t y_t - g_t)/\lambda$

$\qquad \qquad = [1 + A^{1/(1-\alpha)}(\theta_t^{\alpha/(1-\alpha)} - \theta_t^{1/(1-\alpha)}) - R_t] (1-\lambda)k_t/\lambda,$

where the last equality is easily deduced from (2.1), (2.2) and (2.3).

Given a channel policy, Ds' lifetime sales is defined by (2.3), and the evolution of k_t

determined by the behavior of M D' preferences are linear: $E\Sigma^{\infty}_{t=0} \delta^t c^w_t$. For this sum to

converge, δ has to be small enough relative to the factor of growth of sales.

This sales evolution will depend on channel policy. D and M have partly conflicting

interests about channel policies. M benefits from channel services but are hurt by commissions.

Their most preferred channel policy involves some positive level of channel services and

positive commissions but zero increase in commissions to D. D benefits from commissions and

would like commissions to be strictly larger than the amount needed to finance channel services

but small enough not to cause too large a fall in sales and investment. This conflict is resolved by

a negotiation between two D parties, each representing an agent class. Bargaining is focused on

commission agreements. Agreement specifies a channel policy. We restrict attention to constant

agreements that specify a constant θ (D expenditures to sales) and a constant τ (commission

rate). Suppose that a constant channel policy (θ, τ) is implemented without delay, starting in

period zero. There is a perfect foresight sales equilibrium, whose main features are described by

the following.

Assumption One: Let $\theta_t = \theta$ and $\tau_t = \tau$ be such that $\beta R^{1-\sigma} < 1$ and $\delta(\beta R)^{1/\sigma} < 1$ where $R =$

$1+(1-\tau)A^{1/(1-\alpha)}\theta^{\alpha/(1-\alpha)}$ is the equilibrium M's price, and σ is the inverse of the elasticity of

intertemporal substitution. Let $k_0 = k$.

The discounted utility function of the representative M is:

(2.6a)
$$v(k,\theta,\tau) = \iota(R) \frac{k^{1-\sigma}}{1-\sigma}$$

(2.6b)
$$\iota(x) = x^{1-\sigma} \{1 - (\beta x^{1-\sigma})^{1/\sigma}\}^{-\sigma}$$

The utility function of D is:

(2.7)
$$w(k,\theta,\tau) = \frac{(1 + A^{1/(1-\alpha)}(\theta^{\alpha/(1-\alpha)} - \theta^{1/(1-\alpha)}) - R)}{1 - \delta(\beta R)^{1/\sigma}} \, [\,(1-\lambda)k/\lambda]$$

Where $\delta = $ M's discount factor and $\beta = $ D's discount factor. The growth of sales is given by

$k_{t+1}/k_t = (\beta R)^{1/\sigma}$.

First, Assumption One describes the discounted payoffs to D's and M's of an immediate

commission agreement (θ, τ) when the stock of capital is k. This sales function is recursive.

Therefore, $v((k,\theta,\tau)$ and $w(k,\theta,\tau)$ are the payoffs to D's and M's, from any period t on and discounted to period t, of an agreement to set $\theta_s = \theta$ and $\tau_s = \tau$ for $s \geq t$, if the stock of capital at t is $k_t = k$.

Second, sales display unbounded growth provided θ and τ are such that $\beta R > 1$. Long-run growth depends on channel policy.

Third, existence of perfect foresight equilibrium requires v and w to be finite.

The Reason for Requiring $\beta R^{1-\sigma} < 1$ and $\delta(\beta R)^{1/\sigma} < 1$.

The first requirement is satisfied if $R > 1$ or $\sigma > 1$. The second is satisfied if $\delta(\beta R^*)^{1/\sigma} < 1$, where R^* is defined below. The Median M Theorem tells us about sales. Let's consider a goal to pick a constant channel policy. Resulting policy would maximize the utility of the representative M if there are more M's than D's, that is, if $\lambda < 0.5$. M's most preferred policy is given by $\theta = \tau = \alpha$. A condition of channel efficiency exists when $\tau = \alpha$, which says that commissions are enough to finance channel expenditure and no resources are transferred to D's. Such policy maximizes the rate of growth in sales. The associated rate of return is given by

$$R^* = 1 + A^{1/(1-\alpha)}(\alpha^{\alpha/(1-\alpha)} - \alpha^{1/(1-\alpha)}).$$

If $\lambda < 0.5$, the winning policy would maximize the utility of D's, $w(k,\theta,\tau)$.

Ds most preferred policy requires the channel efficiency condition $\theta = \alpha$. A D's most preferred policy requires picking τ such that the return is:

$$(3.1) \qquad R_* = 1 + (1-\tau) A^{1/(1-\alpha)}\alpha^{\alpha/(1-\alpha)}$$

where R_* is the solution of: $\text{Max}_{R\varepsilon[1,R^*)} = \dfrac{(R^* - R)}{1-\delta(\beta R)^{1-\sigma}}$

The most preferred choice by D's would not maximize sales growth. In fact, $R_* = 1$

implies that D's would be willing to sacrifice sales growth for bigger commissions.

Agreements (θ, τ) that satisfy the channel efficiency condition require (1) the condition $\theta = \alpha$ and (2) the condition $R = 1 + (1-\tau) A^{1/(1-\alpha)} \alpha^{\alpha(\bar{1}-\alpha)} \varepsilon [R_*, R^*]$. Any agreement satisfying both conditions is constrained efficient in the sense that there cannot be another (constant) agreement that makes both D's and M's better off.

What if there is no immediate agreement? The parties will keep negotiating commissions until they reach an agreement, or forever. All agents have an incentive to reach agreement quickly. Absence of agreement implies:

(1) loss of potential sales: τ and θ are both zero.

(2) D's do not buy and sales do not grow (because sales possibilities are given by $k_{t+1} = k_t - c_t$, where c_t denotes M's' sales).

(3) M's must decide how much to sell and invest based on their expectations about current and future rate of return. Future rate of return may depend on what the agreement will be. Suppose that there is no agreement at $t = 0$ but that it is common knowledge agreement will be reached in the next period. Sales-investment plans of Ms in period zero will depend on their expectations about the channel policy (θ', τ').

In the market, different commission agreements imply different factors of sales growth and different degrees of profit redistribution. M's would like to choose commissions so as to maximize growth and minimize commission redistribution. D's would prefer slower growth but some commission redistribution. Both sides may benefit from a commission agreement because there are no sales in its absence. Sales growth rate and the degree of profit inequality that would

be observed in this market depend on D institutions, indicating what channel policy would be chosen by M's.

Agreement

What policy would be chosen by a vertically integrated channel? Assume that the channel is controlled by two parties called D (representing D's) and M (representing M's). These parties will be assumed to exchange offers and counter-offers. Bargaining procedure is as follows. At $t = 0$, with k_0 given, one of the parties is chosen randomly (in a way to be specified below). After the offering party has chosen an offer, say $a_0 = (\theta_0, \tau_0)$, the responding party may then accept the offer (Y) or reject it (N). If a_0 is accepted, bargaining ends, and a (constant) channel policy $\theta_t = \theta^0$, $\tau_t = \tau^0$, $t \geq 0$, is immediately implemented. Thus, the acceptance in the initial round results in channel policy stability.

If the offer is rejected, bargaining continues at $t = 1$. In the remainder of period zero, M decides how much to sell (c_0) or to invest (k_1), which determines the amount of capital k_1 at $t = 1$. One of the players is chosen randomly to make an offer $a_1 = (\theta^1, \tau^1)$. If the offer is accepted, bargaining ends and agreement a_1 is implemented. Thus, acceptance at the initial round does not preclude channel policy stability.

No Agreement

If the offer is rejected, M's decide their sales (c_1) and investment (k_2). Period two then starts with capital k_2 and one of the parties chosen randomly to make an offer.

Let's examine the process by which a party is chosen to make an offer in each period. In each period t, the D is chosen with probability λ to make an offer; negotiations take place in a

channel team meeting. . One of the D's is randomly asked to propose a commission package.

Within the channel team, there is need for consensus among the two parties in order to change the commission regime. The need for consensus grants D a degree of power not apparent from the sales goal. The D's can turndown commission proposals and cause damage (no sales in the absence of an agreement). Using this veto power, s/he can appropriate some of the benefits of sales growth. This assumes that the "voice" of each group in the channel becomes stronger the greater the group's size. Given the assumption that λ, the relative number of D's, is also equal to probability that the D makes offers. Thus, in the absence of early bargaining rounds agreement, unstable channel policy results.

Commission negotiations occur in "real time". Market activity does not wait until an agreement: sales and investment take place while the parties negotiate. The bargaining procedure is similar to Binmore's (1987) version of Rubenstein (1982). The principal difference is in the MM is that there are two strategic interacting players: the D parties and a large number of M agents. Interplay between M investment and commission negotiations is absent from models of the Rubenstein type.

Sustainable Bargaining Equilibrium (SBE)

An appropriate equilibrium concept to characterize the solution of this kind of problem is that of the SBE. We use Chang's allocation rule F to describe Ms' sales (c_t) and investment (k_{t+1}).

Let j = D,M (a strategy for player j, denoted S_j) be a description of what offer to make (if it is j's turn to offer) and which offers to accept. Given k_0, an allocation rule F and a strategy pair

$S = (S_D, S_M)$ we can determine the expected payoff to each party. A strategy pair S is a Nash equilibrium if S_D maximizes D's expected payoff given S_M, F, and k_0. A Nash equilibrium is subgame perfect if its continuation is a Nash equilibrium after any history. Now, pose natural restrictions on equilibrium allocation rules. The representative M has capital $k_t(h^{t-1})$ and has to decide how much to sell and invest in period t. A strategy pair S is given. The continuation of S after h^t, call it S/h^t, induces a distribution over future bargaining outcomes, and which agreement will be reached and when. For the allocation rule F to be consistent with optimizing behavior, the behavior that F prescribes after history h^t, $c_t(h^t)$ and $k_{t+1}(h^t)$, must maximize the expected utility of a M given the distribution over the bargaining outcomes induced by S/h^t.

Existence and Characterization of a Sustainable Bargaining Equilibrium

The sustainable bargaining equilibrium (SBE) is an allocation rule F and a strategy pair S such that F is competitive given S, and given F, S is a subgame perfect Nash equilibrium. The two parties' strategies are optimal against each other after any history.

This provides sufficient conditions for the existence of Pareto optimal (PO) stationary SBEs. These SBEs solve a pair of normal equations. The SBEs in this section are stationary in the sense that the players' strategies and the investment factor implied by the allocation rule will be independent of previous history. Focusing on PO, stationary SBEs is justified. First, one may want to restrict attention to stationary equilibria because they depend on history, in this case, especially the amount of capital. Second, a stationary equilibrium is relatively easy to compute.

To derive candidate strategies and allocation rules, let's define a PO stationary SBE. L offers $a_0 = (\theta_0, \tau_0)$ whenever D is selected to make an offer and M offers $a_1 = (\theta_1, \tau_1)$ if C is

selected. Pareto optimality implies that agreement is immediate. Note that θ_0 and θ_1 must be equal to α. Pareto optimality and stationarity imply that channel policy outcomes must be constrained efficient. There must be an agreement with an equilibrium return given by:

(4.1) $$x = 1 + (1-\tau_0)\, A^{1/(1-\alpha)}\alpha^{\alpha(1-\alpha)}$$

(4.2) $$y = 1 + (1-\tau_1)\, A^{1/(1-\alpha)}\alpha^{\alpha(1-\alpha)}$$

whenever the D or the M , respectively, makes an offer.

What allocation rules can be competitive after any history h^t? An agreement will certainly be reached in period (t+1). The agreed upon policy will be a_0 if D is chosen to make an offer and a_1 if M is chosen. Fix any h^t and k_t, and recall that the probability that the D makes an offer is equal to λ. The M's problem is then to choose c_t and $k_{t=1}$ to maximize her expected utility, that is $u(c_t) + \beta\,[\,\lambda\, v(k_{t+1}, a_0) + (1-\lambda)\, v(k_{t+1}, a_1)]$. Solution of this problem is easily shown to be:

(4.3a) $$k_{t+1} = [\,\zeta(x,y)^{1/\sigma} / (1 + \zeta(x,y)^{1/\sigma})\,]\, k_t$$

and $c_t = k_t - k_{t+1}$, where x and y are defined,

(4.3b) $$\zeta(x,y) = \beta\,[\lambda\psi x) + (a-\lambda)\,\psi(y)\,]$$

and ψ is given by (2.6b).

Strategies are a subgame perfect Nash equilibrium if in each period the proposer makes an offer that leaves the responder indifferent between accepting or rejecting the offer. When it is M's turn to answer to an offer a_0, the value to M of accepting the offer, $v(k_t, a_0)$, must equal M's value of rejecting it, $u(c_t) + \beta[\,\lambda\, v(k_{t+1}, a_0) + (1-\lambda\, v(k_{t+1}, a_1)\,]$, with c_t and k_{t+1} determined by the allocation rule and v given by Assumption One. It turns out that M is indifferent if and only if:

(4.4) $\qquad \psi(x) = [\ 1 + (\beta\{\ \lambda\psi(x) + (1-\lambda)\ \psi(y)\ \}\ )^{1-\sigma}\ ]^{\sigma}$

Similarly, D is indifferent between accepting or rejecting a_1 when it is D's turn to respond and if and only if $w(k_t, a_1) = \delta\ [\ \lambda\ w(k_{t+1}, a_0) = (1-\lambda)\ w(k_{t+1}, a_1]$ which, using (2.7), (4.1), (4.2) and (4.3), reduces to:

(4.5a) $\qquad \partial(y) = \delta\ [\ \lambda\ \partial(x) + (1-\lambda)\ \partial(y)\ ]\ \zeta(x,y)^{1/\sigma}\ /\ (1 + \zeta(x,y)^{1/\sigma})$

where ζ is given by (4.3b) and:

(4.5b) $\qquad \partial(x) = \dfrac{R^* - x}{1 - \delta(\beta x)^{1/\sigma}}$

The two equations (4.4) and (4.5) are crucial to characterize the SBE postulated here. Following Rubinstein, call (4.4) and (4.5) the normal equations. The following theorem shows that if x and y satisfy the normal equations and, are constrained efficient outcomes, there is an SBE.

Proposition 1: Suppose x and y belong to $[R_*, R^*]$ and satisfy the normal equations. Also, assume that $\beta x^{1-\sigma}$, $\beta y^{1-\sigma}$, $\delta(\beta x)^{1/\sigma}$, and $\delta(\beta y)^{1/\sigma}$ are less than one. Then the following is an SBE via Allocation rule F: defined by (4.3) above:

Strategy for D: Offer $a_0 = (\alpha, \tau_0)$ when selected to make an offer, where τ_0 satisfies (4.1); when selected to answer to an offer, accept any offer such that $w(k_t, a) \geq \delta\ [\ \lambda\ w(k_{t+1}, a_0) + (1-\lambda)\ w(k_{t+1}, a_1)\ ]$.

Strategy for M: Offer $a_1 = (\alpha, \tau_1)$ when selected to make an offer, where τ_1 satisfies (4.2); when selected to answer to an offer, accept any offer such that $v(k_t, a) \geq u(c_t) + \beta\ [\ \lambda v(k_{t+1}, a_0) + (1-\lambda)\ v(k_{t+1}, a_1)\ ]$. The proof is in Chang's Appendix (Chang 1995). Proposition 1 shows that a

solution to the normal equation exists and identifies an SBE:

Proposition 2: Assume that $R_* = 1$ and that $\sigma > 1$. Then the normal equations (4.4) and (4.5) have a solution (x,y) in $[1, R^*]^2$. The proof of Brouwer's Fixed Point Theorem is in Chang's Appendix.

Joining Proposition 1 and Proposition 2 we obtain the following:

Corollary: If $\delta(\beta R^*)^{1/\sigma} < 1$, $R_* = 1$, and $\sigma > 1$, then there is an SBE.

In the SBE of the Corollary, bargaining stops in the first period. Thus, an early round agreement leads to stable channel policy.

This Corollary suggests how one may study the set of SBEs discussed: if one chooses parameters such that $\sigma > 1$, $R_* = 1$ and $\delta(\beta R^*)^{1/\sigma} < 1$, the Corollary ensures existence but not uniqueness. In the numerical exercises of the next section, Chang was unable to find multiple solutions of the normal equations.

SBEs characterized in this section are independent of the initial quantity of capital k_0. The results of Chang's assumption about functional forms have many consequences. One can allow for renegotiations after an initial agreement without changing the results. Assume that, if an agreement is broken in period t, new commission negotiations start in period (t+1) and that commissions are set to zero again until there is a new agreement. It is intuitively clear that the parties agree on the SBE of the original model and there is never an incentive for either party to break this agreement. Thus, a stable channel policy results with both parties better off in terms of profits.

EMPIRICAL ANALYSIS OF CHANNEL POLICY FORMULATION

The model specifies a negative observable relation between sales growth and commission inequality. The set of SBEs discussed in the previous section are dependent on changes in the underlying parameters of the sales function.   SBE is obtained by solving the normal equations and analyzing how it is affected by changes in each of the parameters. The objective is to examine what one would observe in a cross-section sample of industries which may differ in their fundamental parameters by solving the normal equations numerically. An alternative procedure is to find an SBE and then to use the Implicit Function Theorem. However, since the normal equations are highly nonlinear, that approach becomes messy very quickly. Parameters of the model are the inverse of the elasticity of intertemporal substitution σ, the discount factors of D and M, δ and β, the parameters of the sales function A and α, and the relative number of D's, λ. An empirically plausible benchmark set of parameters for which the Corollary of the last section applies are chosen from the U.S. Census of Manufacturers. Σ equal to 2 seems plausible. As for the parameter α of the sales function in an SBE the efficiency condition $\alpha=\theta = g/y$ suggests that α can be chosen from the Census of Manufacturers. We chose conservative benchmark parameter values: $\alpha= 0.1$ for the benchmark and $\delta= 0.9$, $\beta= 0.95$, and A $= 0.25$. With these benchmark values, one can solve the normal equations. The corresponding solutions of x and y are 1.099 and 1.117. In equilibrium, the average rate of return on capital is around 11 percent. Expected sales growth in the benchmark case is 1.026. Sales would grow 2.6 percent per year. The maximum possible sales growth is about 4.5 percent. In equilibrium sales growth is less than 3 percent.

Profit distribution is summarized by the expected value of the ratio of D profit to total profit denoted by "D share":

$$\text{D share} = \lambda c^{w}/(1-\lambda)y = (\tau y - g)/y = (R^{*} - R)/(\alpha^{\alpha/(1-\alpha)}).$$

For the benchmark parameters, D share is equal to 0.055 (5.5%). D share is always less than λ, and hence an increase in D share implies a more egalitarian commission structure. D share is a measure of the distribution of profit after commission.

Three parameters are especially influential in the determination of SBE's: discount factor, β, for M, δ, for D and λ, the composition of the supply chain. Table 1 shows the SBEs associated with different values of the M discount factor β. As β increases sales growth increases and the D share of profit decreases. Larger β implies a larger "growth pie" to be divided between the two agent classes. A larger β implies that M become more patient in the commission negotiation. Bargaining becomes relatively more favorable to M. Hence the D share decreases. Thus, in rapid sales growth situations, Ds should expect tougher bargaining conditions and are better off settling in the early rounds.

TABLE 1 ABOUT HERE

Table 2 shows different values of the Ds' discount factor δ. A larger δ makes Ds more patient, and, favors the D in the commission negotiations. A larger δ should imply a larger transfer to D, slower growth, and greater D share. Thus, in slow sales growth situations, D face less difficult bargaining conditions and should hold out on settling.

TABLE 2 ABOUT HERE

Table 3 shows changing λ, composition of the supply chain. With larger values of λ the

number of D's increases; D share improves, but growth worsens. The probability that D makes offers is increasing in λ. As λ increases, the D gets a stronger voice in the channel and uses its increased power to obtain more commissions. D share and growth move in opposite directions as we vary parameters.

TABLE 3 ABOUT HERE

The model is consistent with the negative empirical correlation between profit inequality and sales growth. Growth and profit inequality do not depend on the initial amount of capital per M, k_0. One could have expected increases in either k_0 to make M stronger, implying less profit redistribution via commissions. The reason that such intuition fails is that changes in k_0 have ambiguous effects on relative bargaining power. It both strengthens and weakens M: strengthens because they sell more, but it weakens because each agreement becomes more valuable and the cost of waiting increases.

LIMITATIONS

The results naturally are subject to some qualifications. Some limitations of this model are: (1) sensitivity of the equilbria to which finite grid is specified and (2) Nash is a special case of this theorem. If the payoff functions are not continuous, the reaction correspondences can fail to have a closed graph and/or be nonempty. Such a game does not have pure strategy equilibrium. If payoffs are discontinuous, mixed-strategy may fail to exist as well. But, absent Nash equilibrium, predictions can be made by imposing itieratee strict dominance and

rationalizability. In addition, there are normal equations that do not depend on capital and hinge on the specific functional forms. It is possible by assuming different functional forms, the effect on policy remains unknown. Also implicitly assumed is the local uniqueness of SBEs. Stationary SBEs are, in fact, locally unique. Uniqueness of SBEs without the stationarity assumption remains a question.

SUMMARY AND CONCLUSIONS

A model is presented in which commissions and target discount rates determine sales growth and profit distribution and emerge from negotiations between B2B supply chain parties, namely Ms and Ds. An empirical negative relation exists between profit inequality and sales growth. Profit distribution (Persson and Tabellini, 1994) is the main determinant of the D structure. Profit distribution is an endogenous outcome of the bargaining process. Profit distribution and growth are determined simultaneously in this game and reasonably approximate M-D bargaining. An empirical negative relations exists between increases in D discount rate (δ) and sales growth. A positive relationship exists between M discount rate (B) and sales browth. A negative relationship exists between distribution structure (β) and sales growth. A negative relationship exists between distribution structure (λ) (number of D's as a percent of the total supply chain members), . and sales growth.

MANAGERIAL IMPLICATIONS FOR CHANNEL POLCY FORMULATION

Seven key guidelines for marketing and supply chain managers emerged from this study:*:
(1) Agreements (θ,τ) that satisfy the channel efficiency condition require $\theta=\alpha$ and the condition

$R = 1 + (1-\tau) A^{1/(1-\alpha)} \alpha^{\alpha(\bar{1}-\alpha)}\varepsilon$ [R_* , R^*]. Any agreement satisfying both conditions is constrained

efficient, i.e., there cannot be another (constant) agreement that makes both D and M better off.

All agents have an incentive to reach agreement quickly. Absence of agreement implies a loss of

potential sales: τ and θ are both zero. Absence of a commission agreement, D's don't buy and

sales do not grow.

(2) In the market, different commission agreements imply different rates of sales growth

and different degrees of profit redistribution. M's would like to choose commissions that

maximize growth and minimize profit redistribution via commissions.   D's would prefer slower

growth but some profit redistribution in their favor that increase commissions. Both sides may

benefit from a commission agreement because there are no sales in its absence. Thus, the

acceptance in the initial round results in channel policy stability. In the absence of early

bargaining rounds agreement, unstable channel policy results.

(3) Under SBE conditions, strategies are a subgame perfect Nash equilibrium. If x and y

belong to [R_*, R^*] and satisfy the normal equations, and, $\beta x^{1-\sigma}$, $\beta y^{1-\sigma}$ $\delta(\beta x)^{1/\sigma}$, and $\delta(\beta y)^{1/\sigma}$ are

less than one, then the following strategies generate a sustainable bargaining equilibrium based

on allocation rule F (4.3):

Strategy for D: Offer $a_0 =_. (\alpha,\tau_0)$ when selected to make an offer, where τ_0 satisfies

(4.1); when selected to answer to an offer, accept any offer a such that $w(k_t,a) \geq \delta$ [λ $w(k_{t+1}, a_0)$ +

$(1-\lambda)\, w(k_{t+1}, a_1)$].

Strategy for M: Offer $\alpha_1 = (\alpha, \tau_1)$ when selected to make an offer, where τ_1 satisfies (4.2); when selected to answer to an offer, accept any offer such that $v(k_t, a) \geq u(c_t) + \beta$ [$\lambda\, v(k_{t+1}, a_0) + (1-\lambda)\, v(k_{t+1}, a_1)$]. In the SBE of the Corollary, bargaining stops in the first period. Thus, an early round agreement leads to stable channel policy.

(4) If an agreement is broken in period t, new commission negotiations start in period (t+1) and commissions are set to zero again until there is a new agreement. The parties agree on the SBE of the original model and there is never an incentive for either party to break this agreement. Thus, a stable channel policy results with both parties better off in terms of profits.

(5) Benchmark values for the MM parameters chosen for all industries from the Census of Ms predict an average rate of return of 9.9% and 11.7% for M'and D, respectively, and an average sales growth factor of 2.6% per year. Profit distribution yields an average 5.5% share for D.

(6) The M and D target discount factors as well as the composition of the supply chain are important policy variables. A larger M discount factor β increases sales growth. D share of profit decreases as a larger "growth pie" is divided between the two agent classes. A larger λ implies that the M becomes more patient in the commission negotiation. Bargaining becomes relatively more favorable to M. Hence the D share decreases. In rapid sales growth situations, D should expect tougher bargaining conditions and are better off settling in the early rounds.

A larger D discount factor δ makes D more patient, favors D in the commission negotiations. A larger δ should imply a larger commission transfer to D, slower growth, and

greater D share. Thus, in slow sales growth situations, D face less difficult bargaining conditions and should hold out on settling.

(7) Finally, with a larger λ, composition of the supply chain if the number of Ds increases; D share improves, but growth worsens. The probability that the D makes offers is increasing in λ. As λ increases, D gets a stronger voice in the channel and uses its increased power to obtain more commissions. D share and sales growth move in opposite directions.

FUTURE RESEARCH

An integrated model including the Karray & Marin-Herran price and advertising variables and the Ouardighi et. al., 2008 quality improvement, national advertising and retail price integrated with the commission, channel allocation, and discount rate variables in a game theoretic framework offers the possibility of a more comprehensive model. Relaxing the assumption that Ds do not have another source of profit and cannot borrow would provide more congruence of the model to the real world . However, the considerable complexity of such a model built in a game theoretic framework and the extensive data requirements for empirical testing places this effort beyond the scope of this paper. Nonetheless, the pursuit of such a model has considerable merit.

REFERENCES

Amrouche, Nawel, Guiomar Martin-Herran, and George Zaccour (2008), "Feedback Stackelberg Equilibrium Strategies When the Private Label Competes With the National Brand," *Annuals of Operations Research*, 164(1), 79-96.

Basu, Amiya, Tridib Mazumdar and Seth P.Rau (2007), "*European Journal of Operational Research*, 182 (3), 1084-99.

Benidito, Anthony (1987), "Modeling Rationally in Marketing Direct Mail with Game Theory," *Journal of Academy of Marketing Science* 15 (4) 22-32.

Binmore, Kenneth (1987) "Modeling rational players I," *Economics and Philosophy*, 3, 179–214.

Binmore, Kenneth and Larry Samuelson (2001), "Coordinated Action in the Electronic Mail Game," *Games and Economic Behavior*, 35, 6-30.

Brown, James.R., Chekitan Dev and Dong-Gin Lee (2000), "Managing Marketing Channel Opportunism: The Efficacy of Alternative Governance Mechanisms," *Journal of Marketing*, 64 (2) 51-66.

Cashon, G (2003), Supply Chair Coordination with Contracts in Handbook in Operations Research and Management Science: Supply Chain Management (eds.) Graves, S and DeKok, T, Amsterdam: Elsevier..

Cashon, G and M.A. Lavivierre (2005), Supply Chain Coordination with Revenue Contracts: Strengths and limitations, Management Science, 51, 30-54.

Chang, Roberto (1995), "D Party Negotiations, Profit Distribution, and Endogenous Growth,"

Federal Reserve Bank of Atlanta Review, 95-3, (June), 1-33.

Chaudhri, Vivek (2005), "Game Theory & Business Applications," *Economic Record*, 18, 127-140.

Chussil, Mark and David J. Reibstein (1994), "Strategy Analysis with Valve War," San Francisco: Scientific Press.

Ding, Min (2007), "A Theory of Intraperson Games," *Journal of Marketing,* 7, 1-11.

Eliashberg, J. and R. Steinberg (1987), Marketing-Productiion Decisions in an Industrial Channel of Distgribution, Management Science, 33, 981-1000.

Fruchter, G.E. (2008), "Dual-Role Based Pricing in a Dynamic and Competitive Environment," *Journal of Optimization Theory and Applications*, 138 (1), 27-45.

Fundenberg, Drew and Jean Tirole (1991), Game Theory, Cambridge: MIT Press.

Gencturk, Ezra F. and Preet Aulakh (2007), "Norms-and Control-Based Governance of International M-D Relational Exchanges," *Journal of International Marketing*, 92-126.

Jagpal, Sharon (1999), Marketing Strategy and Uncertainty, Oxford: Oxford Press.

Jorgensen, S. (1986), Optimal Production, Purchasing, and Pricing: A Differential Games Approach, European Journal of Operations Research, 24, 64-76.

Jorgensen, S. and G. Zaccour (2004), Differential Games in Marketing, Boston: Kluwer Academic Publishers.

Karray, Salma and Guiomar Marin-Herran, (2008), "Investigating the Relationship Between Advertising and Pricing in a Channel with Private Label Offering: A Theoretic Model," *Review of Marketing Science*, 6, 1-37.

Kogen K. and Charles Tapiero (2007), Suplly Chain Games: Operations Management and Risk Valuation, New York: Springer.

Kotler, Phillip and Gary Armstrong (2008), Principles of Marketing, Upper Saddle River, N.J.: Pearson-Prentice Hall.

Lehman, Donald and Russell Winer (2008), Analysis for Marketing Planning, New York: McGraw-Hill

Lilien, Gary (1993), Marketing Management Analytical Exercises for Spread Sheets, San Francisco: Scientific Press.

Nash, John (1951), "Non-Cooperative Games," *Annals of Mathematics*, 54, 286-295.

Ouardighi, Fouad, Steffen Jorgensoen, and Fredrico Pasim (2008), Dynamic Game of Operations and Marketing Management in a Supply Chain, International Game Tehroy Review, 10 (4), 373-397.

Overgaard, Per, Ram Orzach and Yair Taubman, (2002), "Modest Advertising Signals Strength," Rand Corporation, 1-36.

Palda, K. (1969), Economic Analysis for Marketing Decisions, Englewood Cliffs: Prentice Hall.

Persson, Torsten and Guido Tabellini (1992), "Growth, Distribution, and Politics" in Economic, Growth, and Business Cycles, Cukierman, Hercowitz and Leiderman (eds.). Cambridge: MIT Press.

Rubinstein, Ariel (1982), "Perfect Equilibrium in a Bargaining Model." *Econometrica* 50, 97-109.

Weibull, Jorgen (2005), Evolutionary Game Theory, Cambridge: MIT Press.

TABLE 1

Effect of M' Discount Factor β

	β			
	0.90	0.93	0.95	0.97
x	1.091	1.096	1.099	1.103
y	1.113	1.115	1.117	1.119
Growth	0.996	1.014	1.026	1.038
L share	0.061	0.057	0.055	0.050
R^*	1.149	1.149	1.149	1.149

Fixed Parameters: $A = 0.25$, $\alpha = 0.1$, $\delta = 0.9$, $\sigma = 2$, $\lambda = 0.5$

TABLE 2

Effect of D Discount Factor δ

δ				
	0.80	0.84	0.87	0.90
x	1.105	1.102	1.101	1.099
y	1.124	1.121	1.119	1.117
Growth	1.029	1.028	1.027	1.026
L share	0.045	0.049	0.052	0.055
R^{*}	1.149	1.149	1.149	1.149

<u>Fixed Parameters</u>: A = 0.25, α = 0.1, β = 0.95, σ= 2, λ = 0.5

TABLE 3

Effect of Supply Chain Composition λ

	λ			
	0.4	0.5	0.6	0.7
x	1.108	1.099	1.089	1.077
y	1.124	1.117	1.108	1.098
Growth	1.030	1.026	1.021	1.014
L share	0.041	0.055	0.068	0.085
R^{*}	1.183	1.183	1.183	1.183

Fixed Parameters: $\alpha = 0.1$, $\delta = 0.9$, $\sigma = 2$, $\beta = 0.95$, $A = 0.3$

THE HIDDEN ELEMENT IN MANAGING THE VALUE CHAIN—
SHARING REWARDS FOR GOOD PERFORMANCE

William D. Presutti, Jr., Ph. D., C.P.M.
Associate Professor, Supply Chain Management
Palumbo-Donahue School of Business
Duquesne University
Pittsburgh, PA USA

ABSTRACT

The concept of the value chain was first introduced in a formal way in Michael Porter's 1985 book "Competitive Advantage." Over the ensuing 25 years, much has been written about managing the value chain and its component activities. That work has focused on the more technical elements of value chain management. Little attention has been paid to issues relating to building the shared commitment of the people involved in the value chain activities, commitment necessary for effective value chain management. One of the keys to building that shared commitment is through the equitable distribution of the financial rewards resulting from competitive and profitable value chains.

Introduction

A significant literature exists on the subject of the value chain, all of it born of the concept introduced by

Michael Porter (Porter 1985). The literature is replete with the how-to's involved in managing the value

chain. There is no dearth of content on issues like building supplier relationships, inventory control,

managing transportation, distribution, warehousing, customer service, managing operations, the impact

of technology and other more technical issues.

The concept of the value chain demonstrates perhaps more than any other concept in business the

importance of internal and external collaboration. The typical focus of external collaboration is between

a firm and its customers and suppliers. Equally important is internal collaboration, the lateral

relationships among value chain participants and the relationships between management and non-

management staff. Without internal collaboration, the value chain breaks and resembles more a

collection of silos making any attempts at effective external collaboration difficult. In essence, the value

chain is a metaphor for shared commitment. What ought to follow is the equitable sharing of financial rewards that emanate from successful performance.

It is through a well-managed value chain that a firm creates value and wealth for society. From the input end to the final customers, the value chain designs the products and services, acquires the raw materials and services necessary to provide the final goods and services that the market values and gets those goods and services to the customers quickly and efficiently. Although easy to express in a few words here, experienced practitioners understand the breadth and complexity of the activities and linkages necessary to satisfy markets competitively and profitably. This process requires that the talents and dedication of everyone in the organization be tapped to create sustainable competitive advantage. It should logically follow that those who create the outcomes ought to share in the resulting benefits through increased compensation. We know that business professionals need to manage the linkages in the value chain if it is to be an effective instrument in organizational success. However, a link that is almost completely ignored is the one that ties successful performance to financial benefits for all of the participants in the value chain. That link addresses an important issue in the 21st century economy. The issue is how the fruits of the success of an organization should be equitably shared to increase the sense of identity that is a necessary condition for competitive advantage. {1} Although Porter presented a comprehensive model of the value chain that included the secondary support activity of human resource management (see Figure 1), his discussion related to that activity is silent on the issue of reward sharing as is the existing literature on the value chain.

Porter's Value Chain
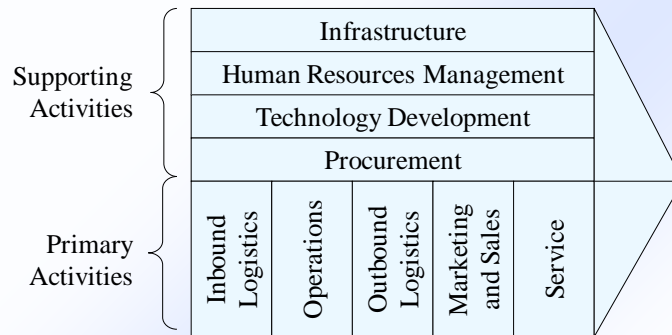
Figure 1 {10}

Another model of the value chain suggests that culture and people make up one of the foundation

pillars of the value chain. The culture is one dominated by collaboration. The people should be selected

and compensated to reinforce that cultural value. (Presutti/Mawhinney, 2009) See Figure 2.

 {11}
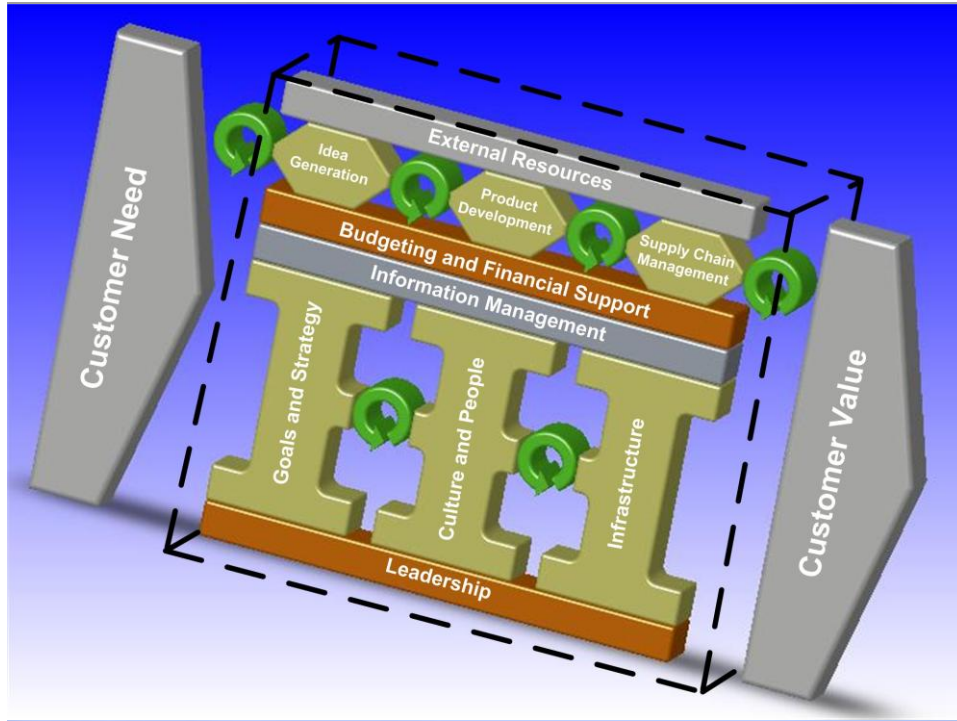
Figure 2

This article is an attempt to make a contribution to the value chain literature by addressing the reward sharing issue. It presents a rationale of why the issue is important , defines what is meant by shared rewards, presents a justification of why a shared rewards system ought to be a component of the value chain, provides examples of organizations using a shared rewards approach, and expands on the definition of a "sustainable business." There is no attempt to prescribe a specific approach to shared rewards although some basic considerations in designing such a program are presented. The existing management literature thoroughly addresses the design issues.

Justifying Rewards Sharing

Consider this observation: "An economy works well when people personally identify with it , so their self-esteem is tied up with its activities…This is especially relevant today because a relatively uninterested, insecure workforce is unlikely to bring about a rigorous (economic) recovery." {14} Although that comment addresses the current state of the U. S. economy, the observation applies

regardless of economic conditions. How is that identification achieved, the lack of interest and insecurity overcome? One of the keys is to develop a true sense of ownership in the organization. That ownership perspective builds esteem by the identification with a successful enterprise. In essence what we need to do is make every person a capitalist. Although the average value chain participant does not invest financial capital in the firm, he or she invests something equally valuable—time and talent. There needs to be a return on that investment over and above a wage or salary if that personal identity with something larger than the wage or salary is to occur.

That the identification issue needs to be developed and reinforced is demonstrated by a few revealing statistics. Worker productivity in the U. S., aided by advances in technology, has grown an inflation-adjusted 50 percent since the mid-1970's. Yet the income of the average worker, those in the value chain whose energy and talents make a difference day-by-day has stagnated. It is estimated that if the gains in productivity had been evenly shared across the workforce, the typical worker's income would be about 35 percent higher now that it was in the mid-1970's. {16} Then, perhaps, income disparity in the U. S. economy would be less of an issue than it is today. {Note 1}

Managing an effective value chain requires collaboration throughout the organization, not just laterally but from top to bottom. In essence, a sense of teamwork needs to permeate the organization. That is difficult to accomplish when those at the top are making 250-400 times what the average worker earns. In 1965, it was 24 times. (Wallis, 2010) "Twenty years ago, people would view today's ratio of CEO pay to frontline pay as a disaster for the organization, destructive to any sense of collaboration." {13}

Since effective value chain management requires collaboration among participants and since this collaboration requires a sense of identity by participants as meaningful members of the team, then

reward sharing must become part of the discussion when issues involving managing the value chain are raised.

Sharing Rewards Through Variable Compensation

"Variable pay plans are organizational systems for sharing economic benefits of improved productivity, cost reductions, quality, and overall business performance in the form of regular cash bonuses." {12} Variable pay ties a portion of a person's compensation to overall company performance and, thereby, gives the value chain participant a stake in that performance and reinforces the importance of collaboration brought about by a heightened sense of identity with the organization.

Effective collaboration will help to define the culture of successful organizations as we move through the 21st century. "It is teamwork and cooperation, some of the essential skills of emotional intelligence, that will drive competitive success." {6} Therefore, it is imperative that compensation practices evolve to reinforce the importance of those skills in the value chain. The continuing focus on individual rewards when what is needed is collaboration throughout the organization is akin to "rewarding A while hoping for B." (Kerr 1975) Unfortunately, that is what appears to be happening. A recent study of incentive plans identified a trend toward an increased focus on rewarding individual performance. {17} Why? Inertia appears to be the best explanation. Individual rewards have been the norm in practice; it is the way it has always been done. Unfortunately, the focus on individual rewards can reinforce the silo thinking that is anathema to the effective management of the value chain. It assumes that those doing an individual employee's evaluation can decide on degrees of difference in performance that are often impossible to discern. How does one explain the difference in performance if one subordinate is awarded a four percent pay increase and another a three percent increase? This may be possible at the extremes of very high and very low performers. However, it is difficult to make distinctions among most others. The results often leave the subordinates confused and disappointed. If corporations are

interested in moving beyond the silo thinking that limits value chain effectiveness, then they will need to get beyond the compensation practices of the past and present and work toward a system of compensation that reinforces the importance of collaboration and builds the individual's identity with the organization.

Overcoming the inertia to change pay practices may be a challenge but it is change justifiable by the evidence. Authorities on compensation practices make two key points in support of variable compensation systems. "The winning solution for implementing a HR program that adds proven value to the bottom line is a variable pay plan **for everyone in the company.** (emphasis added) No human resource program compares with variable pay in terms of generating a high performance place to work, creating a culture of performance and getting value to the business through ROI as variable pay. (Short-term annual) variable pay is justified for everyone in the organization because every employee should influence some key measure of short-term performance." As for the impact on ROI, "a well-designed variable pay plan returns four times the cost of incentive payments in terms of the ROI to the organization. Even the typical plan not ideally designed returns two times what it costs the organization in terms of payments to employees." {17}

The key element of a variable compensation system is that it needs to permeate the value chain and apply to everyone below the executive level (who typically have a separate system). Some firms attempt to use rewards for small teams with shared goals where "team members are interdependent and share goals." {17} There are two problems with this approach. First, it overlooks the fact that interdependencies pervade the value chain and are not limited to the small team members only. It is difficult to determine the impact of small team decisions on actions elsewhere in the value chain. Second, attempting to reward small teams based on performance measures applicable to that team may create the problem of suboptimization of organizational goals. This is a symptom of "haywire

organizational boundaries" when functional specialists begin to view their localized goals ahead of the organization's goals in order to optimize their own achievements and rewards." {2}

Variable Compensation—The Ultimate Pay for Performance

One of the benefits of introducing a variable compensation component to a firm's reward structure is that it is the ultimate pay for results system. Bonuses are paid if the company is profitable. If it is not, the bonuses go away or at least vary based on the level of profitability. Therefore, a sense of entitlement is avoided. {Note 2} Although much is made of not paying simply for effort, it is important to recognize that processes must be in place for results to occur. Therefore, an organization-wide variable compensation system must be accompanied by an employee involvement system where everyone can be included in identifying and addressing the barriers that may exist to profitable company performance. For this system to work there needs to be lots of information-sharing throughout the organization that trumps the "need to know" mentality that permeates much of corporate America. As Risher notes, it is important "to keep employees informed about new products, company success stories, financial performance, and new marketing campaigns. Employees want to know how their efforts are tied to these developments. They want to feel they are part of the story…When incentives are linked to the performance measures, it reinforces an employee's understanding that their contribution is important to the company's success." Regarding incentives, he goes on to argue that team incentives focused on a common goal are especially powerful.

Essentially, a successful variable compensation system should be designed to do two things: support process improvement that leads to better corporate performance and a pay out according to the degree of productivity improvement achieved. {7}

A variable compensation system will thrive in a corporate culture characterized by:

:flexibility and adaptability in thinking and approach

:a continuous improvement mindset

:the ability to capitalize on creativity and innovation

:initiative

:a willingness to push decision making to the lowest level

These are behaviors that exist in an empowered, adaptive work culture that is fundamental to the success of a variable compensation system and, in turn, to a value chain that delivers sustainable competitive advantage. {7}

Consider this experience by the author. On a trip to Japan to visit Japanese companies and academic institutions, a session was scheduled at a mid-size maker of night vision equipment . During the presentation by the company executives, the managing director of the facility shared the company's sales, significant cost items and profitability. When, during the question and answer session, an American colleague commented that in the U. S. this information is shared only a need to know basis, our Japanese host was perplexed. How, he asked, do you expect the people working in the company to understand how what they do impacts the company's performance if you do not share information with them? And how do you provide the focus for the ideas on improving the company's performance that the management expects? They essentially considered all of their employees managers who are in line to share in the success of the company both through the intrinsic reward of being part of a successful organization and extrinsically through the sharing of the income they create. {Note 3}

This attitude reflects the thinking on what can be done to maximize the chances of success in a system of variable compensation. People in the trenches need to know how the company measures its financial

success (ROI, etc.), the factors driving financial success and, of those factors, which can they most likely impact. {7}

Variable Compensation—The American Experience

Experience with variable compensation systems is hardly a Japanese phenomenon. Some well-known American companies include variable pay as part of the compensation package.   Their practices ought to be emulated by all companies if the issue of income disparity is to be effectively addressed.

The first significant and lasting American experience with variable compensation dates to the 1930's with the introduction of what today we call gainsharing. It was conceived and developed by Joseph Scanlon. Scanlon was employed by a troubled steelmaker. He developed a system that promoted worker-management cooperation and increased productivity to help save the employer. It was a system through which the company and employees would share in the gains (cost reductions) if productivity was improved. Employee involvement in decision making was also part of the system. Production councils were organized made up of representatives of non-management employees and managers to attack production inefficiencies and set productivity goals. Bonuses were paid if the goals were exceeded. Seventy-five percent of the bonus pool went to the non-management workers, 25 percent went to management.

 Scanlon became an official of the United Steelworkers Union and continued his work on his gainsharing plan. Early applications of the "Scanlon Plan" saw bonuses of as much as 27 percent over and above base pay distributed to non-management workers. In the years after Scanlon introduced his reward sharing plan, companies in various industries had adopted it, "in industries where profits were excellent and non-existent, where relations between management and workers were good and bad, where productivity was easy or hard to measure." {15} Although Scanlon developed his plan for application in

a union environment, it is applicable to any environment and all workers including service-oriented

organizations and the public sector. {9}

For example, a current well-known company, Progressive Insurance, includes gainsharing as part of its

compensation structure. The Company's description of the program includes the following:

> "Gainsharing, a key component of our compensation plan, allows all regular Progressive
> employees to receive a target percentage of their salary as a bonus that's based on how
> Progressive performs throughout the year. We believe in an even playing field, which is why
> every regular employee is automatically a participant in the gainsharing program…Gainsharing is
> based on how well we meet our objectives as an entire business…For an entry level job, (an
> employee) may receive a gainshare payout anywhere between 0 and 16 percent of eligible
> earnings for that year."   (Progressive.Com/Gainsharing 6/12/2010).

Several factors point to gainsharing as the most successful approach to introducing variable pay to

an organization. It offers:

> *the largest return on invested payroll because it is generally self-funded and has an expected
> net cost to the organization of "0" dollars
> *employee acceptance
> *no entitlement expectations
> *more positive work practices that provide for greater dignity and respect in the workplace
> because bonuses are directly tied to improvement in business processes over which the
> employees have some control. {7}

  Unfortunately, surveys indicate that a distinct minority of companies, approximately 13 percent, are

using gainsharing. Surveys also indicate that, of those companies using gainsharing, 81 percent reported

positive overall program performance. {9}.

Gainsharing is not the only approach to variable compensation.   Lincoln Electric Company has its

Lincoln Incentive Management philosophy that evaluates its employees on work quality, dependability,

Ideas generated, cooperation, and output. Bonuses are based on an employee's performance

evaluation . Profits are set aside for a bonus pool and distributed among the employees. The bonus

typically equates to 90% of an employee's annual wages. {3}   Although more

individually focused, the performance evaluation on which the bonuses are based includes the elements

of cooperation and idea generation, two of the same criteria important in all variable compensation

programs and characteristic of employees important in a culture of collaboration. The lesson from

Lincoln is that firms can retain an individual focus in a variable compensation system. However, it must

include significant evaluation criteria that emphasize the importance of collaboration with other

members of the value chain.

The important point is not the specific form that a reward-sharing component of the compensation

system should take. Over the years, much has been written about gainsharing, Lincoln Electric and

other forms of variable compensation. This vast literature is readily available to any company interested

in including reward-sharing as part of its compensation structure. {Note 4} The important point is that

it should be done and included more universally in compensation systems throughout the American

economy. It should be done as part of a company's efforts to effectively manage its value chain because

of the interdependencies, cooperative behavior and group performance required. (Mericle and Kim,

2004)   Study and discussion of the value chain is incomplete without addressing the issue of variable

compensation in the chain and its applicability to all value chain participants.

The Sustainable Enterprise

The value chain needs to be designed and managed to serve the needs of the sustainable enterprise.

The issue of sustainability is a critical consideration in the leadership and management of the 21st

century corporation. Byrne (2010), in describing the "sustainable executive," notes that he or she

"embraces sustainability as a business opportunity rather than a regulatory burden...They lead their organizations by three principles. The first is rational—focusing on efficiency, cost savings, and conservation. The second is natural—understanding that business and the environment are interconnected. The third is human—fostering safe, healthy work places and **allowing employees to share in the success of the business. Sustainable executives break down the walls and allow knowledge to spread throughout the company, enabling all employees to participate in the success of the business.**" (emphasis added) {5}

It is, then, an opportune time to bring the variable compensation issue to the forefront of the issues being addressed in the contemporary literature on value chain management. No organization can pursue the goal of becoming a sustainable enterprise without the complete commitment of those in the trenches of the value chain. And, as Byrne notes, that enterprise is one in which success is shared and knowledge spread to create a sense of ownership among the employees and to help them participate in the involvement systems inherent in effective approaches to variable compensation.

Blackburn {4} has developed a model sustainability policy, "A Company Commitment to Sustainability," that addresses the triple bottom line of economic success, environmental responsibility, and social responsibility. The social responsibility component of the model policy focuses on "respect for people" and says, in part, "We treat our employees in a respectful, fair, non-exploitive way especially with regard to **compensation and benefits, promotion, training, open (and) constructive dialogue with management, and involvement in decision making.**" (emphasis added) Here we have a recognized authority on the issue who is including compensation and involvement as important components of sustainable enterprise. According to a McKinsey study of chief financial officers, the move toward sustainability including compensation, information sharing, and employee involvement increased

profits and shareholder value by 12 percent demonstrating that becoming a sustainable enterprise makes solid business sense. {5}

Conclusion

The issue of variable compensation needs to become more visible in the value chain management literature. It will expand on the human resource component of the value chain that Porter included in his original design 25 years ago. An effective contemporary view does not look at variable compensation systems like gainsharing as merely pay plans. They should be viewed as a means to get employees involved in the management of the value chain by tapping into their initiative and creativity in all of the value chain's activities. It is this synergistic effort among all value chain participants that creates the gains in performance that allows the participants to share in the value of what they have created. The pay reinforces the productive behaviors needed to move the organization forward. {7}

When Scanlon's gainsharing approach began to take root in the 1950's, an executive from an Illinois company said, "As far as I am concerned, Joe has the answer to the future for American free enterprise capitalism. " {15} Today, with income disparity at record levels, with worker insecurity high, and worker identification with employers tenuous given the massive job losses that have occurred, a reward-sharing component in all corporate compensation systems is needed to at least address the income disparity and worker identification issues to help secure the future success of American free enterprise capitalism.

Notes

1. For the period 1979-2006, the increase in inflation adjusted income for the top one percent of income earners in the U. S. was 260 percent. The average increase for all others in the income distribution was 27 percent. Source: Congressional Budget Office, Historical Effective Tax Rates 1979-2006 as presented in Hacker, Jacob S. and Pierson, Paul (2010). **Winner Take All Politics**. NY: Simon and Schuster

2. The idea of variable compensation used here refers to cash bonuses paid on a regular basis throughout the year. It is not profit sharing, which is typically distributed once a year and is, in most cases, tied to the employee's retirement plan.

3. Japanese companies have typically paid bonuses on a semi-annual basis in July and December. The bonus payouts equate to about 18 percent of total pay. Source: Conrad, Harald and Heindorf, Viktoria (2006). "Recent Changes in Compensation Practices in Large Japanese Companies: Wages, Bonuses, and Corporate Pensions." **Perspectives on Work, Employment, and Society in Japan**. NY: Palgrave Macmillan.

4. On the subject of "Financial Results of Gainsharing" alone, Google identifies 19,300 works that discuss the opportunities, challenges and results for this approach to variable compensation.

References

1. Akerlof, George A. and Kranton, Rachel E. (2010). **Identity Economics**. Princeton, NJ: Princeton University Presss.

2. Ashkenas, R. et. al. (1995). **The Boundaryless Organization.** San Francisco, CA: Jossey-Bass Publishers.

3. Attinello, Joseph et. al. (2005). "Lincoln Electric: A Case Study." www.plu.edu-oliverkl/doc/lincon-electric.doc.

4. Blackburn, William R. (2008). **The Sustainability Handbook**. Washington, D. C.: Environmental Law Institute.

5. Byrne, James (2010). "The Sustainable Executive Creates a Bridge to the Future," Pittsburgh, Pa: **Pittsburgh Post-Gazette**, March 16, 2010.

6. Goleman, Daniel (1995). **Emotional Intelligence**. NY: Bantam Books

7. Hill, Brad (2000). "Assessing Variable Compensation Readiness." Fay, Charles H. (Ed.). **The Executive Handbook on Compensation.** NY: The Free Press

8. Kerr, Steven (1975). "Rewarding A While Hoping for B." **Academy of Management Journal**, December 1975.

9. Mericle, Kenneth and Kim Dong-one (2004). **Gainsharing and Goal Sharing: Aligning Pay and Strategic Goals.** Westport, CT: Praeger Publishers.

10. Porter, Michael (1985). **Competitive Advantage.** NY: The Free Press.

11. Presutti, William D., Jr. and Mawhinney, John R. (2009). "The Value Chain Revisited." **International Journal of Value Chain Management**, Volume 3, No. 2.

12. Schuster, Michael H. (2007). "Aligning Compensation Strategy." Competitive Human Resources Strategies, LLC.

13. Senge, Peter (2010). "Reshaping Business." **Biz Ed**, May/June. Tampa, Fl: AACSB International.

14. Shiller, Robert J. (2010). "Stuck in Neutral? Reset the Mood." **Sunday Money**, The New York Times, January 31, 2010.

15. Time Magazine (1955). "Management: The Scanlon Plan." September 26, 1955, Vol. LXVI, No. 13.

16. Wallis, Jim (2010). **Rediscovering Values**. NY: Howard Books Div. of Simon and Schuster.

17. Zingheim, Patricia and Schuster, Jay (2007). "What Are Key Pay Issues Right Now?" **Compensation and Benefits Review,** May/June.

Radio frequency identification (RFID) has received increased attention from practitioners and academics. Due to the mandates from Wal-Mart and the U.S. Department of Defense, January 2005 can be considered as the "big bang" for RFID. Since then, many press reports have revealed Wal-Mart's starts-stops-and-starts and thus has perceived RFID as a failing technology. Yet, empirical evidence have revealed actual benefits of RFID impacts on supply chain performance. In order to better understand the field, we conducted a survey of logistics professional in order to gauge the current deployment and widespread of RFID implementation. In this paper we extend our prior work and present our most notable findings that suggest indisputable evidence of RFID's ROI. For practitioners and academics, the empirical evidence presented can help identify implementation areas where RFID can have the greatest impact. The data can be used to build the business case for RFID and therefore better estimate ROI and the payback period.

CAPACITATED HIERARCHICAL CLUSTERING HEURISTIC
FOR MULTI DEPOT LOCATION ROUTING PROBLEMS

Marco Lam, York College of PA
441 Country Club Road, York, PA 17402, 717-815-1585, mlam@ycp.edu

John Mittenthal, The University of Alabama
350 Alston Hall, Tuscaloosa, AL 35487, 205-348-608, jmittent@cba.ua.edu

ABSTRACT

In this paper, we develop a hierarchical clustering based heuristic for the Multi Depot Location Routing Problem (MDLRP). We contribute to the literature in two ways. First, we develop two stopping rules for capacitated hierarchical clustering. Hierarchical clustering does not require *a priori* assumptions about the number of clusters. We test the performance of stopping rules that determine the number of clusters under various assumptions. We show that when these clusters are subsequently used to create vehicle routes these stopping rules can result in significant savings. Second, we compare the performance of our hierarchical clustering based heuristic approach for the MDLRP with heuristics and metaheuristics previously proposed in the literature. Our heuristic often finds lower costs than the heuristics proposed in the literature and generate solutions within 4% of various metaheuristic approaches proposed in the literature. As a result of this level of performance, we suggest that this heuristic can be used as a warm start solution for any improvement procedure.

1

A DECISION – PLANNING LABORATORY

Frenck Waage,
University of Massachusetts Boston
100 Morrissey Boulevard
Boston, Massachusetts, U.S.A.
617 287 7736
Frenck.waage@umb.edu

ABSTRACT

The management of a business, and of its value chain, is repeatedly forced to act following events that have already occurred, and also forced to act in the anticipation of coming events before they occur. Management would certainly find a decision laboratory very helpful in these cases. In the laboratory the wisdom of different policies, decisions and actions will be tested before any actions and expenditures are committed to. This paper describes how such a laboratory can be designed in practice. The paper focuses the design on the company's value chain.

Key words: value chain, laboratory,information systems

INTRODUCTION

Every business organization produces its products or services along a value chain. The value chain is the sequence of actions that is required to create the finished output that the customers want. The sequence of steps involved in realizing a product described in Figure 1 (in the appendix) is abstractly equivalent to the sequence in any real value chain. The equivalency is established as follows. Modify and particularize Figure 1 so it correctly represents building and de-building of inventory stocks and flows of materials in any real value chain. Then modify the generic equations which describe Figure 1 so they correctly represent the corrected and modified figure. You now possess an accurate model of your real value chain.

The laboratory discussed in this paper consists of the generic equations which generate the business results in Figure 1. To formulate and understand these generic equations, and Figure 1, are step one when setting up this laboratory in a real organization. The second step is to modify Figure 1 to agree with the observed and real value chain. The third step is to modify the generic equations so it correctly represents the modified and correct figure. The last step is to use the laboratory by processing any proposed decision or planning scenario through the equations to learn the impacts the decisions and plans will have on the business results should they be deployed for real. The laboratory reveals such results before any actions are taken in real life and before serious resources are committed.

This paper describes what is involved in setting up this generic laboratory, and it presents select applications.

RELATED LITERATURE

The relevant literature is huge. There are literature overviews in (Harlan et al., 2001) and (Lamming et al., 2001). The literature that relates to this paper covers three different directions.

The first direction is in papers and books that create mathematical representations (models) of supply networks – large or small, static or dynamic – and that use the representations to describe how the networks work, how best to design them, and how best to manage them? This direction has two sub-directions:

(1) original mathematical methods as presented in (Dantzig, 1963), (Johnson et al., 1973), (Chopra et al., 2003), (Ragsdale, 2004), (Simchi-Levi et al., 2003), and

(2) innovative applications of the original mathematical methods that show how supply networks can be described, designed and managed as discussed in (Dong et al., 2004), (Nagurney et al., 2002 and 2005), (Tayur et al., 1999), (Zhang et al., 2003), (Galinec et al., 2009), (Spohrer et al., 2009), (Chandra et al., 2009).

The second direction is in papers and books that develop methods with which to gauge the effectiveness and the efficiency of the performance of value chains – large or small, static or dynamic, under conditions of risk and certainty. This main direction has two sub-directions:

(1) methods for measuring performance in terms of "six sigma & Lean", QFD and customer value as discussed in (Askin et al., 2001), (Baudin, 2004), (Connor, 2001), (Gans, 2002), (Iordan et al., 2001), Feitzinger et al., 1997), (Sharma et al., 2010), (Lim, 2001), (Karl et al., 2010)

(2) methods for setting up and managing effective production – inventory systems in supply networks as discussed in (Blinder, 1986), (Caldentey et al., 2003), (Chen et al., 2002), (Eichenbaum, 1989), (Glasserman et al., 1995), (Holt et al., 1960), (Kaminsky et al., 2004), (Lawrence, 1984), (Mahajan et al., 2001), (Monden, 1993), (Kattan et al., 2010), (Hu et al., 2010), (Lu et al., 2003), (Song et al., 2002), (Torres et al., 2009).

The third direction is a heterogeneous collection of subjects and is found in papers and books that develop special approaches and comprehensive solutions to value chain problems including managing risk in supply networks, modeling multi-objective problems, developing tactical operations plans, using game theory to understanding supply networks, applying data envelopment analysis for effectiveness measurements, modeling E-business supply chains, and more. A good selection in which these readings can be found is in (Back, 2008), (Cachon et al., 2004), (Graff, 2006), (Lee et al., 2000), (Song et al., 2002), (Aas et al., 2010), (Sherer, 2010), (Sreenivas et al., 2008) (Lee, 2008), (Chen, 2009), (Vanauy et al., 2009), (Sunil, 2008), *(Simchi-Levi, 2004).

This paper describes how, consistently with the best works cited above, a manager can construct a comprehensive managerial value chain model in practice which can be used as his decision and planning laboratory.

SUMMARY OF MATHEMATICAL VARIABLES USED

$d_{k,t}$ = units sold in market k during time period t.

$I_{k,t-1}$ = units of the finished product in inventory in market k at the end of period t-1.

$J_{j,t-1}$ = units of the finished product in inventory at distribution center j at the end of period t-1.

$KC_{f,t-1}$ = units of component C in inventory at factory f at the end of period t-1(C = 1, 2, …,q).

$LC_{w,t-1}$ = units of component C held in ware- house w (w = 1, 2) in inventory at the end of period t-1

$S1_{s,t-1}$ = units of component 1 which supplier s holds in inventory at the end of period t-1

$Y_{j,k,t}$ = units shipped from distribution center j to market k during period t.

$X_{f,j,t}$ = the quantity of finished products shipped from factory f to distribution center j during period t

α = the number of components type 1 needed to produce one unit of the finished product

β = the number of components type 2 needed to produce one unit of the finished product

$WC_{w,f,t}$ = quantity of component "C" (C = 1, 2) shipped by warehouse w to factory f

$VC_{s,w,t}$ = the quantities of component C which supplier s ships to warehouse w during t

$ZC_{s,t}$ = units of component C that supplier s produces during period t

DESIGNING THE GENERIC LABORATORY

We here describe the material flows through a generic value chain in Figure 1. Figure 1 describes how three Market Outlets sell the finished product. Two distribution centers supply the three markets with units of the finished product. Two factories produce the finished product and ship the finished products to two Distribution Centers. The two factories produce the products by combining two different components. Two suppliers produce the components and then ship the components to two warehouses. The two warehouses supply the components to the two factories. The value chain described in Figure 1 is managed by an Operations Manager.

We will now create the analytical model which constitutes the Management Laboratory. The model consists of simultaneous dynamic equations.

First: Measure the demand rate at each market outlet

Measure demand for the finished product as a velocity – the number of units demanded per time unit at each of the three markets. This is done with tested forecasting techniques.

Second: Identify the most economical way of supplying the markets from the distribution centers

The dynamic material flows and inventory stock changes in market k during period t are measured by equation (1.k). Equation (1.k) assures that the quantities received from the distribution centers plus the beginning inventories in market k are sufficient to satisfy final demand at k. Equation (1.k) represents m simultaneous equations, one for each market k, and k = 1, 2, …, m.

$$I_{k,t-1} + \sum Y_{j,k,t} \; - \; d_{k,t} \; - \; I_{k,t} \; \geq \; 0 \qquad (1.k)$$

Three of these equations have been explicitly written as (1.1), (1.2) and (1.3). The total

costs incurred from these material flows and from the inventory stocks are: 1) the costs of transportation from the distribution centers to the markets measured by $\sum\sum a_{i,k,t}*Y_{i,k,t}$, plus 2) the costs of carrying the average inventories at the three market outlets $\sum b_{k,t}*(I_{k,t-1} - I_{k,t})$. The lower case letters "a" and "b" are constant unit cost coefficients.

$$I_{1,t-1} + Y_{1,1,t} + Y_{2,1,t} - d_{1,t} - I_{1,t} \geq 0 \qquad (1.1)$$
$$I_{2,t-1} + Y_{1,2,t} + Y_{2,2,t} - d_{2,t} - I_{2,t} \geq 0 \qquad (1.2)$$
$$I_{3,t-1} + Y_{1,3,t} + Y_{2,3,t} - d_{3,t} - I_{3,t} \geq 0 \qquad (1.3)$$

To determine the optimal values to assign to the decision variables in (1.k) cannot be made independently of deciding how to manage the distribution centers. That problem is discussed next.

Third: The most economical way to supply the distribution centers from factories
The material flows and inventory stock changes at distribution center j during period t are measured by equation (2.j). Equation (2.j) guarantees that the quantities of the finished products on hand at the distribution centers are sufficient to satisfy final demand at the markets. Equation (2.j) represents n simultaneous equations, one for each distribution center j, and j = 1, 2, …, n.

$$J_{j,t-1} + \sum X_{f,j,t}) - \sum Y_{j,k,t}) - J_{j,t} \geq 0 \qquad (2.j)$$

Two of these equations have been explicitly written as (2.1) and (2.2).

$$J_{1,t-1} + X_{1,1,t} + X_{2,1,t} - Y_{1,1,t} - Y_{1,2,t} - Y_{1,3,t} - J_{1,t} \geq 0 \qquad (2.1)$$
$$J_{2,t-1} + X_{1,2,t} + X_{2,2,t} - Y_{2,1,t} - Y_{2,2,t} - Y_{2,3,t} - J_{2,t} \geq 0 \qquad (2.2)$$

Total costs incurred from these material flows and from the inventory stocks are: 1) the costs of transportation from the factories to the distribution centers measured by $\sum\sum c_{f,j,,t}*X_{f,j,t}$, plus 2) the costs of carrying the average inventories at the three market outlets $\sum d_{k,t}*(J_{j,t-1} - J_{j,t})$, and 3). The costs of shipping to the markets have been accounted for in the cost equation that follows equation (1.k). The lower case letters "c" and "d" are constant unit cost coefficients.
To determine the optimal values to assign to the decision variables in (1.k) and in (2.j) cannot be made independently of deciding how to manage the material flows through the factories. That problem is discussed next.

Fourth: The most economical way to produce the quantities that the distribution centers and the market outlets require to optimally satisfy demand
One unit of the finished product is produced at factory f (f = 1, 2, …, p) by combining α units of component 1 and β units of component 2. To be able to produce the quantity of $\sum X_{f,j,t}$ of finished product, factory f needs a quantity $\{\alpha*(X_{i1,t} + X_{i2,t})\}$ of components type 1 and a quantity $\{\beta*(X_{i,1,t} + X_{i2,t})\}$ of component 2. The two components are shipped to the factory from two warehouses.
The left side of equation [3.C.w.j] measures the total quantity of component C that factory f receives from the two warehouses. The right hand side of (3.C.w.j) measures the

quantity that factory f needs for producing the quantity of finished products that must be sent to the two distribution centers. Equation (3.C.w.j) requires the two sides to be equal. This also guarantees that the quantity which is shipped to the distribution centers is also the quantity that is required to satisfy final demand in the markets.

$$\sum WC_{w,f,t} = \sum \{\alpha*(X_{f,j,t})\} \qquad (3.C.w.f)$$

$$W1_{1,1,t} + W1_{2,1,t} = \{\alpha*(X_{1,1,t} + X_{1,2,t})\} \qquad (3.1.1)$$
$$W2_{1,1,t} + W2_{2,1,t} = \{\beta*(X_{1,1,t} + X_{1,2,t})\} \qquad (3.2.2)$$
$$W1_{1,2,t} + W1_{2,2,t} = \{\alpha*(X_{2,1,t} + X_{2,2,t})\} \qquad (3.1.3)$$
$$W2_{1,2,t} + W2_{2,2,t} = \{\beta*(X_{2,1,t} + X_{2,2,t})\} \qquad (3.2.4)$$

Factory f starts period t with $KC_{f,t-1}$ units of component C in inventory and ends the period with $KC_{f,t}$ units in inventory. The dynamic stock – flow equations for the factories are (4.C.f):

$$KC_{f,t-1} + \sum WC_{w,f,t} - \alpha*(\sum X_{f,1,t}) - KC_{f,t} \geq 0 \qquad (4.C.w.f,t)$$

Four of these equations have been explicitly written as (4.1.1) through (4.2.2).

$$K_{1,t-1} + W1_{1,1,t} + W1_{2,1,t} - \alpha*(X_{1,1,t} + X_{1,2,t}) - K1_{1,t} \geq 0 \qquad (4.1.1)$$
$$K2_{1,t-1} + W2_{1,1,t} + W2_{2,1,t} - \beta*(X_{1,1,t} + X_{1,2,t}) - K2_{1,t} \geq 0 \qquad (4.2.1)$$
$$K1_{2,t-1} + W1_{1,2,t} + W1_{2,2,t} - \alpha*(X_{2,1,t} + X_{2,2,t}) - K1_{2,t} \geq 0 \qquad (4.2.1)$$
$$K2_{2,t-1} + W2_{1,2,t} + W2_{2,2,t} - \beta*(X_{2,1,t} + X_{2,2,t}) - K2_{2,t} \geq 0 \qquad (4.2.2)$$

Costs of shipping from the warehouses to the factories are: $\sum\sum\sum e_{w,f,t}*WC_{w,f,t}$. Costs of production are $\sum\sum g_{j,f,t}*X_{j,f,t}$. The costs of carrying the average inventories at the factories are $\sum\sum hC_{f,t-1}*(KC_{f,t-1} - KC_{f,t})$. The lowercase letters "e" , "g", and "h" are constant cost coefficients.

To determine the optimal values to assign to the decision variables in (1.k) and in (2.j) and in (4.1.1) through (4.2.2) cannot be made independently of deciding how to manage the material flows through the warehouses. That problem is discussed next.

Fifth: The most economical way to supply the warehouses with component parts
The material flows and inventory stock changes at warehouse w during period t are measured by equation (5.w).

$$LC_{w,t-1} + VC_{1,w,t} + VC_{2,w,t} - WC_{w,1,t} - WC_{w,2,t} - L1_{w,t} \leq 0 \qquad (5.C.w)$$

Ware House w (w = 1, 2) holds $LC_{w,t-1}$ units of component C in inventory at the beginning of period t, receives the quantities $VC_{1,w,t}$ from supplier 1 and $VC_{2,w,t}$ units of component C from Supplier 2 at during period t. Warehouse w ships the quantities $WC_{w,1,t}$ and $WC_{w,2,t}$ of component 1 to Factories 1, 2, and ends period t with $LC_{w,t}$ units of component C in inventory. Four of these equations have been explicitly written as (5.1) through (5.5).

$$L1_{1,t-1} + V1_{1,1,t} + V1_{2,1,t} - W1_{1,1,t} - W1_{1,2,t} - L1_{1,t} \leq 0 \qquad (5.1)$$
$$L2_{1,t-1} + V2_{1,1,t} + V2_{2,1,t} - W2_{1,1,t} - W2_{1,2,t} - L2_{1,t} \leq 0 \qquad (5.2)$$
$$L1_{2,t-1} + V1_{1,2,t} + V1_{2,2,t} - W1_{2,1,t} - W1_{2,2,t} - L1_{2,t} \leq 0 \qquad (5.3)$$
$$L2_{2,t-1} + V2_{1,2,t} + V2_{2,2,t} - W2_{2,1,t} - W2_{2,2,t} - L2_{2,t} \leq 0 \qquad (5.4)$$

The costs of transportation (not yet counted) are $r_{11t}*V1_{11t} + rM_{21t}*V1_{21t} + M_{12t}*V1_{12t} + M_{22t}*V1_{22t}$ and $M_{11t}*V2_{11t} + M_{21t}*V2_{21t} + M_{12t}*V2_{12t} + M_{22t}*V2_{22t}$ and the costs of carrying inventories are $sB_{1t-1}*L1_{1t-1} + s_{2t-1}*L1_{2t-1} + B_{1t-1}*L2_{1t-1} + B_{2t-1}*L2_{2t-1}$ and $B_{1t}*L1_{1t} + B_{2t}*L1_{2t} + B_{1t}*L2_{1t} + B_{2t}*L2_{2t}$. The lower case "r" and "s" are constant unit cost coefficients.

To determine the optimal values to assign to the decision variables in (1.k) and in (2.j) and in (4.1.1) through (4.2.2) and in (5.1) through (5.2) cannot be made independently of deciding how to manage the material flows through the suppliers. That problem is discussed next.

Fifth: The most economical way to produce the component parts that are needed

Supplier s (s = 1, 2) (in Figure 1) begins period t with $S1_{s,t-1}$ units of component 1 in inventory, produces the quantities $Z1_{s,t}$ units of component 1 at t, ships the quantities $V1_{w,1,t}$ and $V1_{w,2,t}$ of component 1 to Ware Houses 1, 2, and ends period t with $S1_{w,t}$ units of component 1 in safety inventory.

$$S1_{1,t-1} + Z1_{1,t} - V1_{1,1,t} - V1_{1,2,t} - S1_{1,t} \leq 0 \qquad (6.1)$$
$$S2_{1,t-1} + Z2_{1,t} - V2_{1,1,t} - V2_{1,2,t} - S2_{1,t} \leq 0 \qquad (6.2)$$
$$S1_{2,t-1} + Z1_{2,t} - V1_{2,1,t} - V1_{2,2,t} - S1_{2,t} \leq 0 \qquad (6.3)$$
$$S2_{2,t-1} + Z2_{2,t} - V2_{2,1,t} - V2_{2,2,t} - S2_{2,t} \leq 0 \qquad (6.4)$$

The costs of production are: $N1_{1t}* Z1_{1,t} + N21_{1t}* Z2_{1,t} + N1_{2t}* Z1_{2,t} + N2_{2t}* Z2_{2,t}$. The with N being constant unit cost coefficients. The costs of transportation have been accounted for in equations (5.j)

To determine the optimal values to assign to the decision variables in (1.k) and in (2.j) and in (4.1.1) through (4.2.2) and in (5.1) through (5.2), and in (6.1) through (6.2) can now be made by finding the values for the decision variables which optimizes the objective function, and satisfies all the equations simultaneously. We shall next discuss how this solution is obtained.

The simultaneous equations we have discussed are now displayed together in the appendix.

THE MANAGERIAL LABORATORY

The generic equations that we have so far developed are displayed for the period t as a simultaneous system of equations in the first appendix. A two-period model is generated by advancing the time index to t+1 for the 17 equations, and by adding these new 17 equations to the system of equations as 17 additional simultaneous equations. The two-period system will have 17 + 17 = 34 equations. A three-period model is generated by advancing the time index to t+2 for the 17 equations, and by adding these new 17 equations to the system of 34 equations as 17 additional simultaneous equations. The

three-period system will have 17 + 17 + 17 = 51 equations. In this manner we can create a model for any number of time periods.

The appendix shows the structure of the system to be.

The structure will always consist of a dynamic (time dependent) objective function. The objective function in this case measures total cost.

The structure will always consist of the equations which describe the material flows and the inventory stocks in the value chain. In this case there are 17 simultaneous equations describing period t. These are shown in the appendix.

The solution requires that the values for the decision variables that optimize the objective function while satisfying the stock-flow equations be identified.

The structure of the model is that of a linear program or of a linear network program. The optimal solution is obtained from software that solves linear programs. The easiest way to the solution for moderate sized problems is to place the data in an EXCEL spread sheet. The optimal solution is then obtained by using the software SOLVER's Simplex Method, or the GRG, or the Evolutionary Method. Other recognized solution methods exists: Revankar's Algorithm, Dynamic Programming, and commercially available linear programming software. We shall now discuss some of the very many management problems which can be dissected in this laboratory.

Six of the many problems the laboratory can handle are listed next. We shall discuss these in what follows.

Problem 1: Identifying the optimal "demand-Pull" policy
- The optimal Value Chain material flows
- The optimal information system

Problems 2, 3, 4: Optimal reactions to three Value Chain disturbances

Problem 5: Optimal design of a Value Chain

Problem 6: Optimal Facility, Capacity, and Location Decisions

Problem 7: Managing truly large Value Chains

A DISCUSSION OF THE LISTED PROBLEMS

Problem 1: Identifying the optimal "demand-pull" policy for a Value Chain

In this application the manager wishes to find the optimal policy for operating the value chain for two consecutive future time periods, t and t+1.

The data used are displayed in the second appendix.

For period t, the optimal solution for material flows and stocks are:

Supplier 1 produces 800 units of component 1 and 1600 units of component 2. Supplier 2 will not produce in period t

Supplier 1 ships 800 units of component 1 to warehouse 1 and 1600 units of component 2 to warehouse 1. Warehouse 1 ships 800 units of component 1 and 1600 units of component 2 to factory 1.

Factory 1 produces a total of 400 units of the finished product. Factory 1 ships 300 units of the finished product to distribution center 1, and ships 100 units of the finished product to distribution center 2

Distribution center 1 ships 120 units to market 1 and 180 units to market 2. Distribution center 2 ships 100 units to market 3. The three markets make good on all their orders, exactly, in period t. Inventories are all zero.

For period t+1, the optimal solution for material flows and stocks are:
Supplier 1 produces 900 units of component 1 and 1800 units of component 2. Supplier 2 will not produce in period t

Supplier 1 ships 900 units of component 1 to warehouse 1 and 1800 units of component 2 to warehouse 1. Warehouse 1 ships 900 units of component 1 and 1800 units of component 2 to factory 1.

Factory 1 produces a total of 450 units of the finished product. Factory 1 ships 300 units of the finished product to distribution center 1, and ships 150 units of the finished product to distribution center 2

Distribution center 1 ships 140 units to market 1 and 160 units to market 2. Distribution center 2 ships 150 units to market 3.The three markets make good on all their orders, exactly, in period t. Inventories are all zero.

Total cost for the two period t and t+1 equals $49,040.

The optimal information system
The manager's ideal information system should contain the information he needs to plan for, and to make decisions, that lead to the value chain's optimal performance. The mathematical model contains this information and it yields the optimal network performance. Therefore, the mathematical model is itself the optimal information system.

Problem 2, 3, and 4.
Identifying optimal reactions to Value Chain flow disturbances
In supply networks, interruptions, disturbances and changes will occur. When they do occur, many actions and reactions are possible. Among all the possible actions, which is the optimal one? Here is how to find out!

Problem 2: A fire at Factory 2 at time t reduces the capacity to produce there by 50% for both period t and t+1. 100% capacity will be restored for periods t+2, t+3 and on. To learn the optimal reaction to this disturbance, the supply network manager will include this new condition in the three-period model in the following manner. Assume that the capacity of Factory 2 before the fire is K_t units. The manager will constrain this to 50% of K_t after the fire in period t and in period (t+1). The constraint for period t and t+1 are written as equations (8.1) and (8.2). They are added to the original 17 equations in (7.1). For the future periods t+2, t+3, ... these constraints are removed and the capacity is returned to 100% of K.

$X_{21} + X_{22} \leq 0.5*K_t$ for period t $\qquad$ (8.1)
$X_{21} + X_{22} \leq 0.5*K_t$ for period t+1 $\qquad$ (8.2)

Problem 3: The quality and reliability of the components that supplier 1 has produced have increased noticeably over and above that of Supplier 2. Which is the optimal

operations policy for the whole network after this change? The cost of producing the two components are measured by $c27*Z1_{1,t} + c28*Z1_{2,t} + c29*Z2_{1,t} + c30*Z2_{2,t}$. To find the answer to the question, first write the cost coefficients as follows:

$C27$ = cost per unit produced / yield with yield in the range 0% to 100%. If unit cost is $16, and 20% of all units produced and sold are returned by customers, and 20% new units are produced and given to the customers for free, then the yield is $(100 - 20)\% = 80\%$. Cost is ($16.00/0.8) = $20.00. If quality and reliability improves, the yield may be 90%. The cost would then be ($16/0.9) = $17.77.

Write the cost coefficients in $c27*Z1_{1,t} + c28*Z1_{2,t} + c29*Z2_{1,t} + c30*Z2_{2,t}$ as shown using yield measures. As the yields vary, so will the coefficients. As the coefficients vary, so will the optimal solution. After the coefficients change, the Manager will solve the program to find the optimal solution. This optimal solution is the optimal operations policy after the changes. That policy is the one which is optimal in the presence of changes in the quality and the reliability measures.

Problem 4: A new road has been completed that makes the link from Ware House 1 to Factory 2 noticeably cheaper than it has been. The quantity of component 1 transported from Ware House 1 to Factory 2 is $W1_{1,2,t}$ and for component 2 is $W2_{1,2,t}$ units. They are elements in the vector ξ_t. The unit costs are c11 through c18 in the cost vector c_t (bolded in the vector). $c_t = (c1, c2, c3, c4, c5, c6, c7, c8, c9, c10, c11, c12, c13, c14, c15, c16, c17, c18, c19, c20, c21, c22, c23, c24, c25, c26, c27, c28, c29, c30)$.

Calculate how the new road will change these coefficients. Use the changed and correct coefficients in the model. The optimal solution will be optimal given the true values of these (and all the other) coefficients.

Situation D: Multiple disturbances take place simultaneously (e.g., fire in a plant, a distribution center shuts down and quits, a new market outlet is added or shut down, a new supplier is added or quits, etc), then the Supply Network Manager will modify the linear program by adding or deleting constraints as required to correctly capture the changes and by changing the coefficients that are affected. The program is then solved to identify the material flows that are optimal given the simultaneous disturbances. The new optimal solution is the optimal policy in the face of the simultaneous disturbances.

Problem 5: Arriving at the optimal design of Value Chains
Testing if it is optimal to invest in and deploy a recommended change the first step is entering the proposed new connections into Figure 1. The second step is to modify the current mathematical model to correctly accommodate the changes made in Figure 1. The last step is to solve the correct model.

When the correct linear model is solved for its optimal solution, it will provide the answer to the design inquiry. Here is how: The optimal design of a supply network consists of the nodes and arches which are not zero in the optimal solution. The optimal design consists of the nodes and the arches which have positive values in the optimal solution.

Problem 6: Deciding on the optimal Value Chain facility, its capacity, and its location
If demand varies over time then questions of capacity and location arise in supply

networks. Seeking answers to design-of-supply-networks questions always begin with a flow chart such as Figure 1. Into the flow chart one draws all the existing suppliers, ware houses, factories, distribution centers, market outlets, and all those that are contemplated but do not yet exist (all real and potential arcs and nodes). This flow chart is then translated into equations by methods we demonstrated above. The equations will represent a network of existing and phantom but potential arches and nodes. The objective function will have to be formulated using correct cost coefficients. The result is a prototype model which represents all its potential configurations. The problem is to identify from all the potential configurations which is the optimal one.

The prototype dynamic linear programming model is solved for its optimal solution. The arches and nodes that are included in the optimal solution represent the configuration that is the optimal design of the supply network. E.g., if a ware house in some given location is never used by the optimal linear program solution, then do not establish it in the supply network. If, however, a new supplier or factory in some given location always appears in the optimal linear program solution, then build it and invest in it. It is part of the optimal supply network infra structure in the optimal location. This shows how uses of the dynamic linear programming model will reveal optimal answers to design questions.

Problem 7: How to optimally manage very large Value Chains
The dynamic linear supply network prototype model discussed here will have to be "scaled up" to sizes that make use of more than one million variables and several hundred thousand constraints in order to provide optimal flow solutions through very large and complex supply networks. Examples of such large networks are Mobil Oil's global operations, AT&T's network operations, and large airlines. Models involving more than one million variables will take a while to formulate. Once formulated, however, they can be solved.

CONCLUSIONS AND NEW INSIGHTS

This paper has presented the following ideas in response to strongly felt managerial needs:
It explains how, in practice, management can develop a tractable dynamic analytical model of the material and information flows in any specific value chain, no matter how large and complex that chain may be. The model is dynamic and is used to identify the optimal solutions to very many practical planning and decision problems.
The paper explains how, in practice, management can test the wisdom of deploying alternative decisions and plans before it commits "serious money" to costly deployments. The model makes a "Decision and Planning Laboratory" available to managers for these purposes.

REFERENCES

[1] Aas. B, and B. Wallace, (2010), "Management of Logistic Planning", International Journal of Information Systems and Supply Chain Management, 3(3), 2010, (pp 1 - 17)

[2] APICS 2008 APICS Dictionary, Twelfth Edition (2008), (ed) John H Blackstone, Jr.,

APICS – The Association for Management

[3] Arntzen, B. C., Brown G. G., Harrison T. P., and Trafton, L. L. (1995). Global Supply Chain Management at Digital Equipment Corporation. Interfaces, 25 (1), 69 – 93

[4] Askin, R., and I. Goldberg, 2001, Design and Analysis of Lean Production Systems, New York, Wiley Publishing Company

[5] Back, S., and V. Prabhu, "An Algorithm for solving multiobjective problems with multiple decision makers", International Journal of Information and Decision Sciences, 1(2), 2008, (pp 178 - 193)

[6] Baudin, M., (2004), Lean Logistics; the nuts and bolts of delivering materials and goods, Productivity Press, New York, new York

[7] Beyer, D. And Ward, I. (2002). Network Server Supply Chain at HP: A Case Study. In Song, I and Yao, D (eds) Supply Chain Structures: Coordination, Information and Optimization. Kluwer Academic Publishers

[8] Blinder, A., (1986). Can the production smoothing model of inventory behavior be saved? The Quarterly Journal of Economics, CI (August), 431 – 453

[9] Cachon, G.P., and S. Netessine, (2004), "Game theory in supply chain analysis", Handbook of Quantitative Supply Chain Analysis: Modeling in the E-Business Era, (pp 13 – 66), Kluwer Academic Publishers, Boston, Massachusetts

[10] Caldentey, R., and L. M. Wein (2003), "Analysis of a decentralized production-inventory system", Manufacturing and Service Operations Management, 5(1), (pp 1 – 17)

[11] Chandra, C., and J. Grabis, "A goal model driven supply chain design", International Journal of Data Analysis, Technology and Strategy, 1(3), 2009, (pp 224 – 241)

[12] Chen, C., (2009), "A Mathematical Model of Tactical Operations Planning for Response Base Supply Chains", International Journal of Information Systems and Supply Chain Management, 2(3), 2009, (pp 77 - 89)

[13] Chen, F., Ettl, M., Lin, G. Y., and Yao, D. D. (2002), "Inventory-Service Optimization in Configure-to-Order Systems", Manufacturing and Service Operations management 4, 114 – 132

[14] Chopra, Sunil and Meindl, Peter, Supply Chain Management, Pearson Prentice Hall, Upper Saddle River, New Jersey, 2003

[15] Connor, G., 2001, Lean manufacturing for a Small Shop, Dearborn, MI., Society of Manufacturing Engineering

[16] Dantzig, George B, Linear Programming and Extensions, Princeton University Press,

Princeton, New Jersey 1963

[17] Dong, J., D. Zhang, A. Nagurney, (2004), "A supply chain network equilibrium model with random demands", European Journal of Operations Research, 156, (pp. 194 – 212)

[18] Eichenbaum, M., (1989) Some empirical evidence on the production level and production cost smoothing models of inventory investment, American Economic Review 79 (4), 853 – 864

[19] Feitzinger, E. And Lee, H. (1997), Mass Customization at Hewlett Packard: The Power of Postponement., Harvard Business Review, 75, 1, 116 – 121

[20] Galinec, D., "Procurement business services modeling in service based architecture of equipping systems", International Journal of Information Systems in the Service Sector, 1(4), 2009, (pp 50 - 60)

[21] Gans, N (2002), "Customer loyalty and supplier quality competition", Management Science, 48(2), (pp. 207 – 221)

[22] Glasserman, P. And Tayur, S. (1995). "Sensitivity Analysis for Base-Stock Levels in Multi-Echelon Production-Inventory Systems," Management Science, 42(5), 263 – 281

[23] Graff, T. O., (2006), "Unequal Competition Among Chains of Supercenters: Kmart, Target, an dWal-Mart", The Professional Geographer, 58(1), (pp. 54 – 64)

[24] Harland, C.M., R.C.Lamming, J. Zheng, and T.E.Johnson (2001), "A Taxonomy of supply networks", The Journal of Supply Chain Management, Fall, (pp. 21 – 27)

[25] Holt, C.C., Modigliani, F., Muth, I.F., Simon, H.A., (1960), Planning Production, Inventories, and Work Force, Prentice Hall, New Jersey.

[26] Hu, X., and J. Xie, "Assessing the value of early order commitment for supply chains with (s,S) policies and lost sales", International Journal of Applied Management Science, 2(3), 2010, (pp 205 - 223)

[27] Iohnson, Richard, William Newell, Roger Vergin, Operations Management A Systems Concept, Houghton-Mifflin, Boston, Massachusetts, 1973

[28] Iordan I., and F. Michael, 2001, The Lean Company: Making the Right Choices, Dearborn, MI., Society of Manufacturing Engineers

[29] Kaminsky, P. And Swaminathan, I. M. (2004). "Effective Heuristics for Capacitated Production Planning with Multiperiod Production and Demand with Forecast Band Refinement", Manufacturing and Service Operations Management 6, 184 – 194

[30] Karl, F., Moser, L., and P. Melliar-Smith, (2010), "An automated Supply Chain

Management System and its Performance Evaluation", International Journal of Information Systems and Supply Chain Management, 3(2), 2010, (pp 84 - 107)

[31] Kattan, I. A., and T. A. Khudairi, (2010), "Simulation of Inventory Systems in a Supply Chain using RFID", International Journal of Information Systems and Supply Chain Management, 3(1), 2010, (pp 68 – 86)

[32] Lamming, R.C., T.E.Johnson, J. Zheng, and C.M., Harland, (2001), "An Initial Classification of Supply Networks", International Journal of Production and Operations Management, 20(6), (pp. 675 - 691)

[33] Lawrence, K and Stelios Zanakis, Production Planning and Scheduling (eds), (1984) Industrial Engineering and Management Press, Institute of Industrial Engineers, Norcross, Georgia

[34] Lee, H.L, K. C. So, and C.S.Tang, (2000), "The value of information sharing in two-level supply chain", Management Science, 46(5), (pp 626-643)

[35] Lee, T. Y. S., "Supply chain risk management", International Journal of Information and Decision Sciences, 1(1), 2008, (pp 98 - 114)

[36] Lim, W.S., (2001), "Producer-supplier contracts with incomplete information", Management Science, 47(5), (pp. 709 – 715)

[37] Lu, Y., Song, I. S. And Yao, D. D. (2003). Order Fill Rate, Lead Time Variability, and Advance Demand Information in an Assemble-to-order System". Operations Research 51, 292 – 308

[38] Mahajan, S., and G. vanRyzin (2001), "Inventory competition under dynamic consumer choice", Operations Research, 49(5), (pp. 645 – 657)

[39] Monden, Y., (1993) Toyota production System, IIE, ISBN 0-89806-129-6

[40] Nagurney, A, D. Matsypura, (2005) "Global supply chain network dynamics with multicriteria decision-making under risk and uncertainty", Transportation Research, E 41, (pp. 585 – 612)

[41] Nagurney, A., J. Dong, and D. Zhang (2002), "A supply chain network equilibrium model", Transportation Research, E38, (pp281 – 303)

[42] Pull Production for the Shop Floor, (2002), The Production Press Development Team, Shop Floor Series, Productivity Press, ISBN 1-56327-274-1

[43] Ragsdale, Cliff T, Spreadsheet Modeling & Decision Analysis, Thompson South Western, Mason, Ohio, 2004

[44] Sharma, A. K., Mehta, I. C., and J. R. Sharma, "Analyzing Programming Tools for the Development of Quality Function Deployment", International Journal of Information and Decision Sciences, 2(2), 2010, (pp 132 - 146)

[45] Sherer, S, (2010), "Enterprise Applications for Supply Chain Management", International Journal of Information Systems and Supply Chain Management, 3(3), 2010, (pp 18 – 28)

[46] Simchi-Levi D., P. Kaminski, and E. Simchi-Levi, 2003, Designing and Managing the Supply Chain, (2nd ed.) New York, McGraw-Hill

[47] Simchi-Levi, D., S. D. Wu, and Z. M. Shen, (2004), Handbook of Quantitative Supply Chain Analysis: Modeling in the E-Business Era, Springer

[48] Song, I. S. And Yao, D. D. (2002). Performance Analysis and Optimization of Assemble-to-Order Systems with Random lead Times. Operations Research, 50, 889 – 903

[49] Song, I. S. And Yao, D. D. (eds) (2002) Supply Chain Structures: Coordination, Information, and Optimization, Kluwer

[50] Spohrer, J., and S. Kwan "Service Science, Management, Engineering, and Design (SSMED): An Emerging Discipline-Outlines References", International Journal of Information Systems in the Service Sector, 1(3), 2009, (pp 1 - 31)

[51] Sreenivas, M., and T. Srinivas, (2008), "Effectiveness of Distribution Network", International Journal of Information Systems and Supply Chain Management, 1(1), 2008, (pp 80 - 86)

[52] Sunil Dharmapala, P., "Adding value in supply chains by improving operational efficiency using data envelopment analysis: a case from published data", International Journal of Applied Management Science, 1(2), 2008, (pp 160 - 175)

[53] Tayur, S, Ram Ganeshan, and Michael Magazine (eds), (1999) Quantitative Models for Supply Chain Management, Kluwer Academic Publishers, Boston, Massachusetts

[54] Torres, F., and G. Mejia, (2009), "Optimal Solutions for a Multi-Product Multi-Echelon Production and Replenishment System", International Journal of Information Systems and Supply Chain Management, 2(2), 2009, (pp 62 - 72)

[55] Vanauy, I., Zailani, S., and N. Pujawan, (2009), "Supply Chain Risk Management Literature review and Future Reseqarch", International Journal of Information Systems and Supply Chain Management, 2(1), 2009, (pp 16 - 33)

[56] Zhang, D., J. Dong, and A. Nagurney, (2003), "A supply chain network economy: Modeling and quantitative analysis", Innovations in Financial and Economic Networks, Edward Elgar Publishers

APPENDIX A:
Figure 1: The making of a given product

The abbreviations used in the boxes in the graph have the following meanings:

Cj = Component Producer j = 1, 2, 3, 4

SAi = Subassembly maker i = 1, 2. Subassembly i = 1 is made from components 1 and 2.

Subassembly i = 2 is made from the components identified by the arrows.

Subassemblies are shipped to and then held in a warehouse.

From the warehouse the subassemblies are shipped to one of two factories, both of which produce units of a given finished product.

Each factory ships finished products to the distribution centers

From the distribution centers the finished products are shipped to the market outlets where customers buy them.
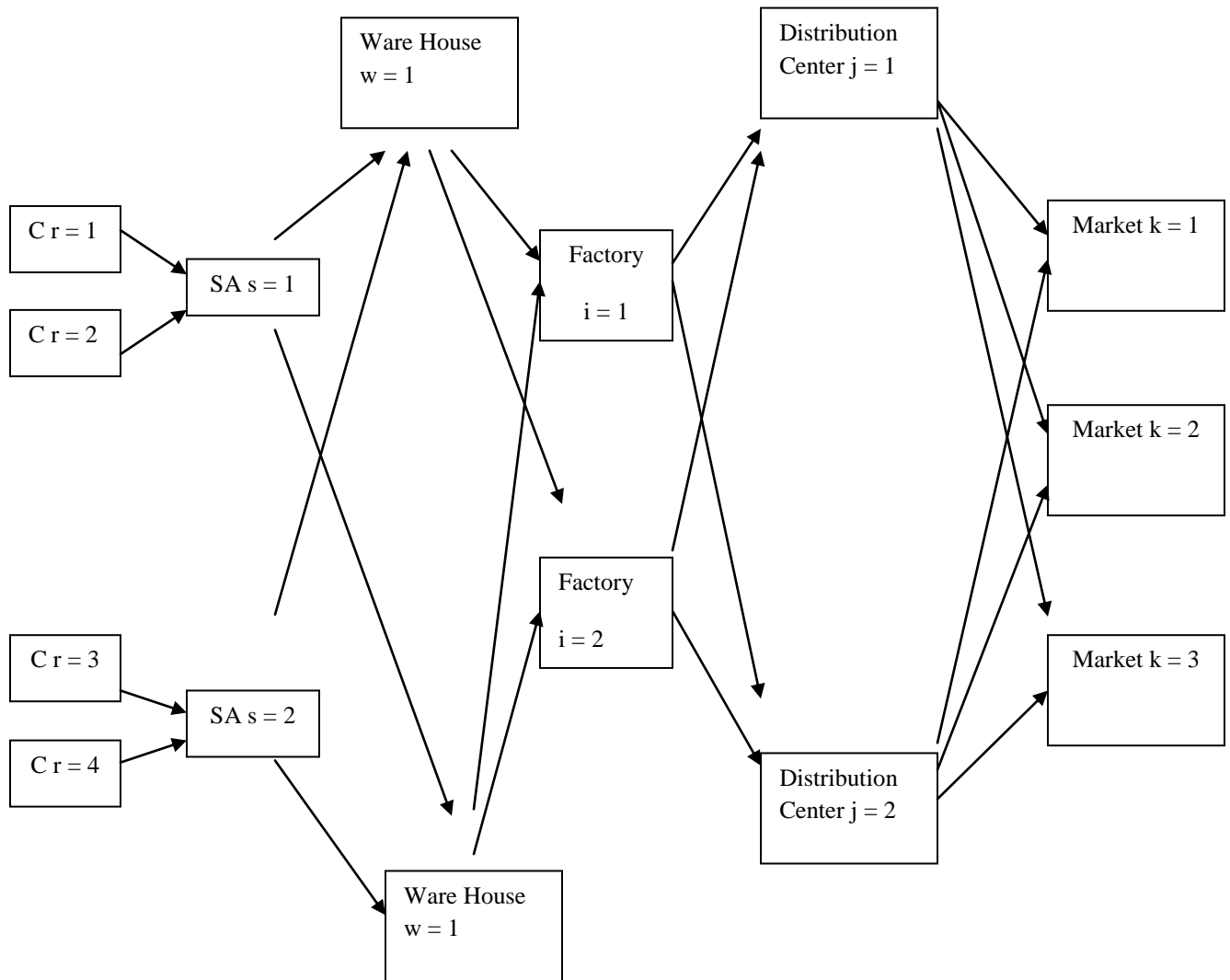
APPENDIX B:
The 17 equations we have explicitly developed and discussed in the text

$I_{1,t-1} + Y_{1,1,t} + Y_{2,1,t} - d_{1,t} - I_{1,t} \geq 0$ Market k=1, period t   (1.1)

$I_{2,t-1} + Y_{1,2,t} + Y_{2,2,t} - d_{2,t} - I_{2,t} \geq 0$ Market k=2, period t   (1.2)

$I_{3,t-1} + Y_{1,3,t} + Y_{2,3,t} - d_{3,t} - I_{3,t} \geq 0$ Market k=3, period t   (1.3)

$J_{1,t-1} + X_{1,1,t} + X_{2,1,t} - Y_{1,1,t} - Y_{1,2,t} - Y_{1,3,t} - J_{1,t} \geq 0$ Distrib. Center j=1, period t, (2.1)

$J_{2,t-1} + X_{1,2,t} + X_{2,2,t} - Y_{2,1,t} - Y_{2,2,t} - Y_{2,3,t} - J_{2,t} \geq 0$ Distrib. Center j=2, period t, (2.2)

$K1_{1,t-1} + W1_{1,1,t} + W1_{2,1,t} - \alpha*(X_{1,1,t} + X_{1,2,t}) - K1_{1,t} \geq 0$ Factory 1, period t   (4.1.1)

$K2_{1,t-1} + W2_{1,1,t} + W2_{2,1,t} - \beta*(X_{1,1,t} + X_{1,2,t}) - K2_{1,t} \geq 0$ Factory 1, period t   (4.2.1)

$K1_{2,t-1} + W1_{1,2,t} + W1_{2,2,t} - \alpha*(X_{2,1,t} + X_{2,2,t}) - K1_{2,t} \geq 0$ Factory 2, period t   (4.2.1)

$K2_{2,t-1} + W2_{1,2,t} + W2_{2,2,t} - \beta*(X_{2,1,t} + X_{2,2,t}) - K2_{2,t} \geq 0$ Factory 2, period t   (4.2.2)

$L1_{1,t-1} + V1_{1,1,t} + V1_{2,1,t} - W1_{1,1,t} - W1_{1,2,t} - L1_{1,t} \leq 0$ Ware House 1, period t   (5.1)

$L2_{1,t-1} + V2_{1,1,t} + V2_{2,1,t} - W2_{1,1,t} - W2_{1,2,t} - L2_{1,t} \leq 0$ Ware House 1, period t   (5.2)

$L1_{2,t-1} + V1_{1,2,t} + V1_{2,2,t} - W1_{2,1,t} - W1_{2,2,t} - L1_{2,t} \leq 0$ Ware House 2, period t   (5.3)

$L2_{2,t-1} + V2_{1,2,t} + V2_{2,2,t} - W2_{2,1,t} - W2_{2,2,t} - L2_{2,t} \leq 0$ Ware House 2, period t   (5.4)

$S1_{1,t-1} + Z1_{1,t} - V1_{1,1,t} - V1_{1,2,t} - S1_{1,t} \leq 0$ Supplier 1, period t   (6.1)

$S2_{1,t-1} + Z2_{1,t} - V2_{1,1,t} - V2_{1,2,t} - S2_{1,t} \leq 0$ Supplier 1, period t   (6.2)

$S1_{2,t-1} + Z1_{2,t} - V1_{2,1,t} - V1_{2,2,t} - S1_{2,t} \leq 0$ Supplier 2, period t   (6.3)

$S2_{2,t-1} + Z2_{2,t} - V2_{2,1,t} - V2_{2,2,t} - S2_{2,t} \leq 0$ Supplier 2, period t   (6.4)

APPLICATION OF SIMULATION IN INVENTORY MANAGEMENT OF EOL PRODUCTS IN A DISASSEMBLY LINE

Badr O. Johar, Northeastern University, (617) 373-7635, johar.b@husky.neu.edu
Surendra M. Gupta, Northeastern University, (617) 373-4846, gupta@neu.edu

ABSTRACT

This paper presents a simulation approach using System Dynamics (SD) technique to model the inventory buildup and behavior of disassembled parts accumulation and or consumption in a disassembly line setup. The approach is considered as a model for decision making process in a reverse-logistics environment. Given the descriptive nature of the approach, the problem is investigated under different scenarios of demand and supply of end-of-life (EOL) products to draw conclusions of the system reaction to external factors such as demand changes. This will allow the disassembly facility to examine different strategies of handling inventory of end-of-life (EOL) products and disassembled parts. System dynamics (SD) techniques are used to simulate the process and show its complexity. Numerical example is given to illustrate the approach and draw conclusions.

Keywords: Reverse supply chain, Disassembly, System dynamic, End-of-life, Inventory management

INTRODUCTION

The in-bound stream of incoming and returned end-of-life (EOL) products results in many challenges to the Original Equipment Manufacturers (OEMs) that if handled properly can represent a profitable business, as some studies showed that companies that engage in "active value recovery" have an advantage over others with "no value recovery" policy over the long term [1]. Interest in the reverse logistics has emerged during the past two decades, shifting the thinking from traditional fast introduction of new products in the market to full accountability of the products manufactured, or in other words the interest has shift from "materials recovery" to "value recovery" [1]. Now the process does not stop at the end customers. The manufacturer is obligated to bear any costs also that may arise at the end of the product life cycle to fully acquire and recover any value remaining in these items. Activities like return, inspection, disassembly, and recovery were hardly considered before and now became part of the important aspects of doing business. Companies now consider these end-of-life (EOL) products as an alternative source for secondary resources instead of virgin resources that are depleting very fast. This approach will allow the reuse of materials and parts more than once before they are finally discarded. This approach has environmental as well as economical benefits. Major challenge original equipment manufacturers (OEMs) face now is how to implement an effective reverse supply chain network that is both cost effective and efficient. The rapid increase of return of products from end customers back to the origin is a main reason behind that interest [2]. System Dynamics (SD) approach is introduced as a decision making tool to help the planner take corrective action, and simulate different scenarios in the reverse supply chain network with a disassembly facility to anticipate the behavior of the system under a

variety of situations. The system dynamics approach will allow examining the relationship between the different elements of inventory management in a disassembly context, and thus taking corrective action to minimize the overall negative impact on inventory and cost [3].

BACKGROUND AND LITERARTURE REVIEW

A review of the state of the art research in the area of Reverse Logistics (RL) and Disassembly Line was published in 2010 [4]. The system dynamics was first introduced in the 1960's by Jay Forrester in his book Industrial Dynamics and mainly were applied in forward supply chain applications. It is a solution approach based on qualitative and quantitative logical thinking [5]. First models of system dynamics application in closed-loop supply chain was targeted at the automobile recycling and motivated by the recent government regulations and environmental concerns [6]. He studied the interactions between original equipment manufacturers (OEMs), consumers, and recycling companies and modeled the effect of legal and technological changes on the industry. Vlachos presented a system dynamics model for strategic remanufacturing and collection capacity planning of a single product reverse supply chain for product recovery. The model presents the analysis of system operations such as product flows and stocks, considering capacity constraints, alternative environmental protection policies involving a take-back obligation [7]. Kumar and Yamaoka apply the system dynamics modeling approach into the Japanese automobile industry to examine the relationship between reduce, reuse, recycle, and dispose of car parts. They performed qualitative and quantitative analyses on the basis of the stock flow diagram for the closed-loop supply chain [3]. Lehr and Milling examined different strategies for collection and value recovery of end-of-life (EOL) products in electronics industry through dynamic simulation technique using system dynamics methodology [1]. Poles and Cheong uses system dynamics simulation for studying and managing complex feedback systems, particularly business and social system, to model an inventory control system in a remanufacturing process where production is integral with remanufacturing. They analyzed the total inventory cost influenced by returns rate affected by the external factors [8]. Georgiadis and Vlachos examined the effect of environmental legislations in customer demand in a remanufacturing environment on long term decision making. The model is later extended to include the impact on capacity planning and collection when managing product returns at the end of its life cycle [9].

## SYSTEM DYANMICS SIMULATION STRUCTURE AND RESULTS

The philosophy behind system dynamics (SD) is founded on the concept that industrial systems are like input-output systems. The system state changes according to the change in the rate of inflow and outflow. Hence, any system can be viewed as an input-output system with input-output rates. The system dynamics (SD) model is characterized by the feedback loop that triggers the corrective, control, action. The purpose of the model using VenSim 5.9 software is to simulate the buildup of the inventory of disassembled parts under a stochastic demand and supply setting. The model describes the different activities that take place in the disassembly facilities such as arrival of end-of-life (EOL) products, disassembly, inspection, addition to

inventory (serviceable or recyclable), and ultimately dispose of. Figure 1 is a graphical representation of the inventory management model structure using a System Dynamics (SD) modeling technique developed using VenSim PLE 5.9.

Methodology of System Dynamics (SD) Simulation Model
The first step in the model is to inspect the end-of-life (EOL) products and those pass inspections are sent to the failed EOL inventory. After that, products are sent to be disassembled and others are disposed of or recycled. It is assumed that certain as product age increase its probability to qualify for disassembly and further processing decreases. As parts become older it gets expensive to recover its value by performing disassembly operations, and they become more and more suitable for recycling rather than value recovery, repair, or remanufacturing. After the decision is made to perform the disassembly operation, the end-of-life (EOL) product will be routed through the designated workstations, in this work the assumption is that each part is disassembled separately in a different workstation, meaning no coupling of tasks at any workstation. After the part is disassembled, it will be added to the inventory, and if rejected after disassembly it is either disposed of or recycled. Once the demand for these items occur, a withdraw against the current inventory of parts will occur and the balance is usually carried to the next period. The simulation model is capable of describing number of complex entities in the inventory system such as queue size, work-in-process, inventory level, as the entities are routed in the system and inventory is generated and added. The output file will provide a daily inventory size of the different parts disassembled and the demand occurrences.

Figure (1) Model Structure of Inventory Management of EOL using SD Approach
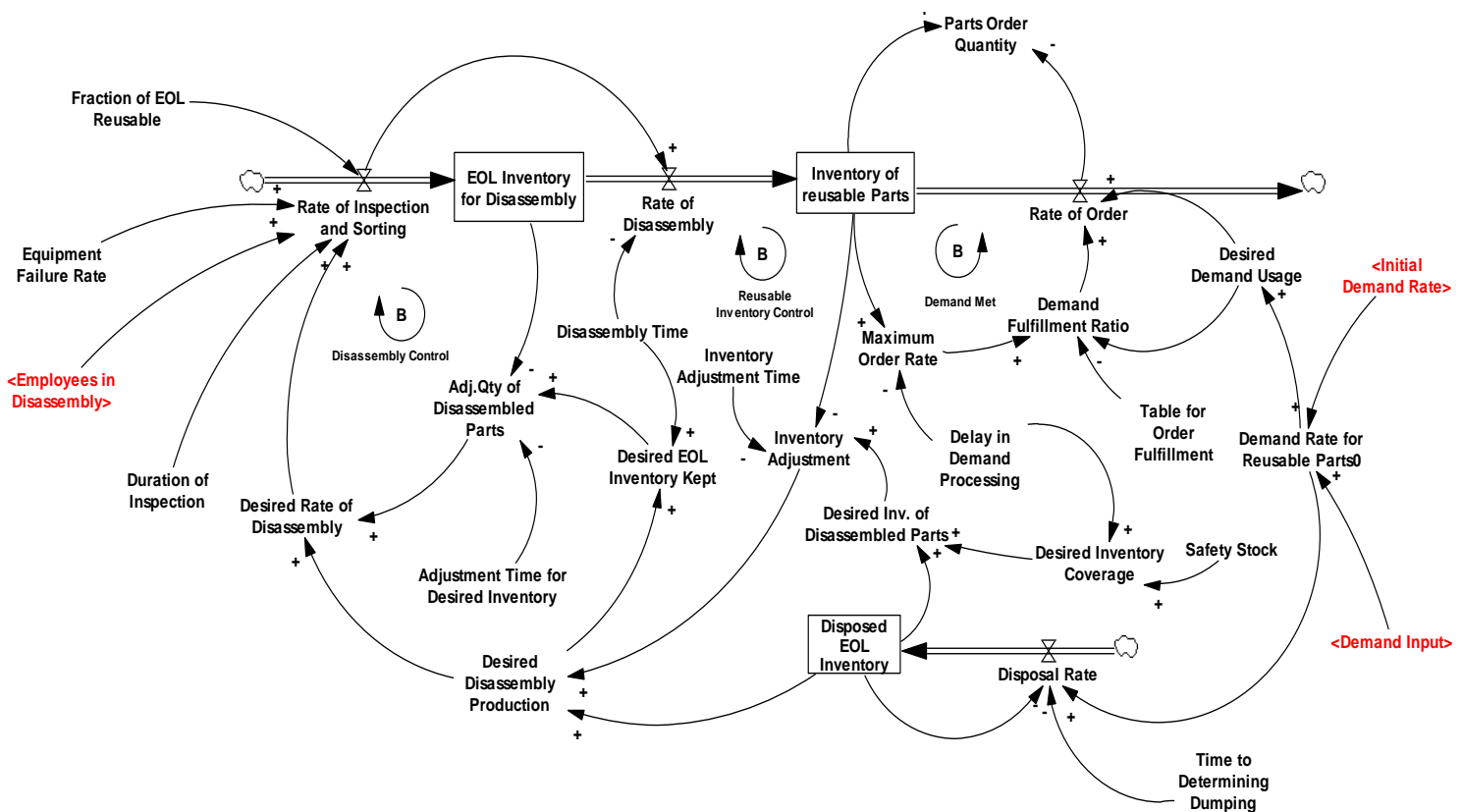
Simulation Outlook and Results

The system dynamics model for the inventory management of end-of-life (EOL) products was developed using the components strategy approach and VenSim PLE5.9 simulation software. The first step is the model setting for time and units. Initial time and final time were set at 0 and 300 respectively and with a time step of 0.125. The units of time selected as weeks and type of integration used is Euler's method. To illustrate the interdependencies between the model elements and the effects of external forces, a sudden hike in demand known as *step input* demand scenario is examined using the model. The initial demand was set at 100 units/week, and a sudden demand increase by 20%. This sudden increase in demand caused a gradual increase of the current end-of-life (EOL) products inventory of 38.90 % from 1,800 units to 2,500 units within 28 weeks span. The inventory of end-of-life (EOL) products drops after that for 15 consecutive weeks. The system reaches equilibrium state at week 60 with a steady level of end-of-life (EOL) products of 2,300 core products on hand, or 31.4% of the initial level at week 0. As a result of a sudden increase in demand, the current inventory of "reusable" parts are depleted to cope with the demand before it is started to build up inventory as the disassembly operations starts. The inventory reaches its lowest level of 300 disassembled parts in 15 weeks span, before it stabilize again at 375 parts level in week 60. . The inventory policy under sudden hike in demand can be built on the premises of the inventory system behavior. Thus, it is realistic to permit shortages in the early stages of the hike in demand. However, the equilibrium state maybe reached sooner if sufficient buffer stock is kept to take back the impact t of such demands. The figures clearly show that the inventory system reacts slowly to the sudden positive change in demand, destabilizing the system before regains control.

Figure (2) Levels Associated with Inventory System Under *Step Input*
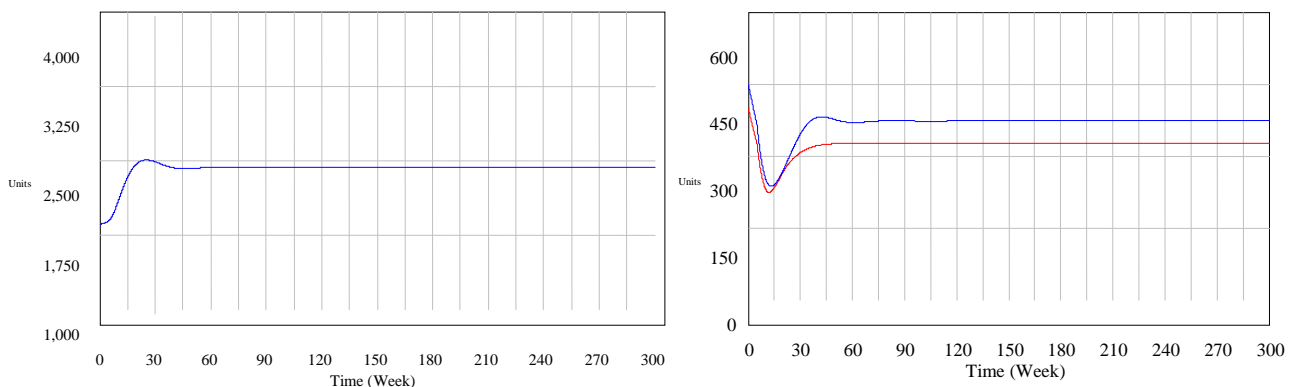
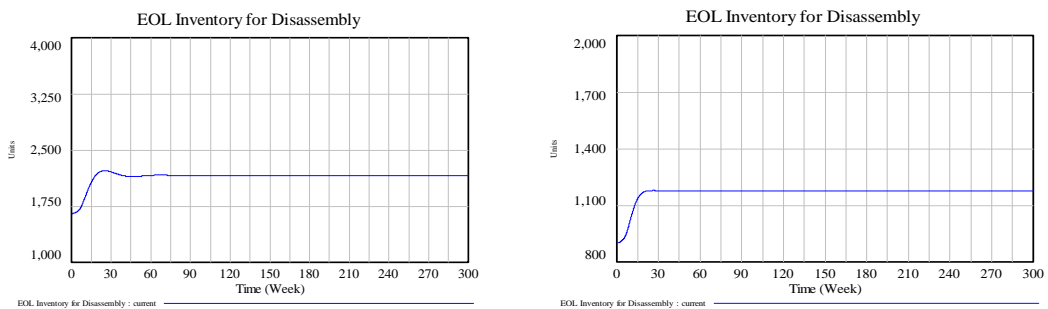EOL Inventory   Parts Inventory

Thus, the sudden hike in demand can not only be satisfied by increasing the rate of inspection, sorting, and disassembly. It is important to realize that the increase in end-of-life (EOL) products inventory is because of the increase of the collection of products from the end customers back to recovery to satisfy the higher demand for the disassembled parts. A collection and transportation network has to exist to ensure smooth availability of end-of-life (EOL) products.

**System Dynamics (SD) Original Model Sensitivity Analysis**

The original model of the end-of-life (EOL) products inventory system assumes the demand is 100 units/week. To test the model behavior when some changes occur, the following two scenarios are presented and tested, when all other parameters remain unchanged: i) when the demand changes from 100 to 112 units/week , and ii) when the disassembly time changes from 8 to 16 minutes/ product. The changes on the inventory of end-of-life (EOL) products and inventory of reusable parts inventory are shown in the following figures (3) and (4).

Figure (3) Changes in EOL Inventory Level Scenario (i) and (ii)

An increase in demand of 12 units a week, equivalent to 12%, resulted the initial on hand inventory of end-of-life (EOL) core products to be increased to 2,300 unit's levels to cope with the demand of disassembled parts. Similarly, the increase in disassembly time triggers the increase of the inventory level of end-of-life (EOL) products to 1,200 units, equivalent to 50%.

Figure (4) Changes in Parts Inventory Level Scenario (i) and (ii)
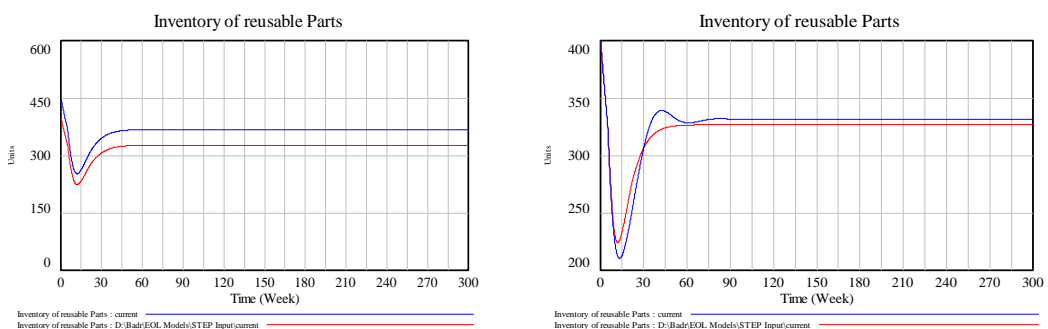
Increase in the demand and/or the increase of the disassembly time both have a negative effect on the inventory of reusable parts, causing a drop in the inventory level, before disassembly rate increases. It is clear that the system reacts better and twice as faster when the demand changes, compared to changes in the system parameter itself such as the increase of the disassembly time. In the first case, where the demand increases by 12% to 112 units/week, the drops to near 200 parts inventory and recover by week 45. In case of disassembly time changes, there is a sharp decrease in reusable parts inventory due to continuous demand and low line yield of parts, yet above the 200 parts level, and it takes longer time to stabilizes again.

CONCLUSIONS

In this paper, a System Dynamics (SD) model of inventory management problem in a disassembly line was developed and tested under demand fluctuation scenario. The probabilistic nature of the problem is assumed to provide more accurate results compared to deterministic models, hence a better understanding of the system long term behavior under external factors uncertainty. Problem methodology and model is presented using commercial software VenSim PLE 5.9. In future research, other factors such as price fluctuation or seasonal effect can be added. The, sensitivity analysis is applied to examine the impact on the inventory of disassembled parts and core products when : i) demand increases and ii) disassembly time doubles. In conclusion, the System Dynamics (SD) model is a decision making tool that helps planner forecast the long term system behavior under certain market condition and changes.

REFERENCES

[1]   Lehr, C. and Milling, P. M., "From Waste to Value - A System Dynamics Model for Strategic Decision Making in Closed-Loop Supply Chains", Proceedings of the 27th International Conference of the System Dynamics Society, Albuquerque, New Mexico, pp. 1-15, July 26-30, 2009.

[2]   Gungor, A. and Gupta, S. M., "Issues in Environmentally Conscious Manufacturing and Product Recovery: A Survey", *Computers and Industrial Engineering*, Vol. 36, No. 4, 811-853, 1999.

[3]   Kumar, S., Yamoka, T., "System Dynamics Study of the Japanese Automotive Industry Closed Loop Supply Chain", *Journal of Manufacturing Technology Management*, Vol. 18, No. 2, 115-138, 2007.

[4]   Ilgin, M. A. and Gupta, S. M., "Environmentally Conscious Manufacturing and Product Recovery (ECMPRO): A Review of the State of the Art", *Journal of Environmental Management*, Vol. 91, No. 3, 563-591, 2010.

[5]   Forrester, Jay W. "Industrial Dynamics", Cambridge, 1961.

[6]   Zamudio-Ramirez, Pavel "Economics of Automobile Recycling", Cambridge 1996.

[7]   Vlachos, D., Georgiadis, P., and Lakovou, E." A System Dynamics Model for Dynamic Capacity Planning of Remanufacturing in Closed-Loop Supply Chains", *Computers and Operations Research*, Vol. 34, 367-394, 2007.

[8]   Poles, R. and Cheong, F., "Inventory Control in Closed-Loop Supply Chain Using System Dynamics", *Proceedings of the 2009 the 27th international System Dynamics Conference*, 2009.

[9]   Georgiadis, P., Vlachos, D., and Tagaras, G. " The Impact of Product Lifecycle on Capacity Planning of Closed-Loop Supply Chains with Remanufacturing", *Production and Operations Management*, Vol. 15, No. 4, 514-527, 2006.

[10] Pochampally, K. K., Nukala, S. and Gupta, S. M., *Strategic Planning Models for Reverse and Closed-Loop Supply Chains*, CRC Press, Boca Raton, Florida, ISBN: 9781420054781, 2009.

[11] Svensson, G., "The Theoretical Foundation of Supply Chain Management: A Functionalist Theory of Marketing", *International Journal of Physical Distribution and Logistics Management*, Vol. 32, No. 9, 734-754, 2002.

Going Green Together? An Empirical Investigation of

Management Contracts in the U.S. Hotel Industry

Jie Zhang[a], Nitin Joglekar[a], and Rohit Verma[b]

[a] School of Management, Boston University
[b] School of Hotel Administration, Cornell University

We examine a two-stage service supply chain consisted of an asset owner and an operator facing investment decisions during sustainable development. In so far as the asset ownership and operations responsibilities reside in two firms, the structural and infrastructural aspects of sustainability investments involving this asset are split across firm boundary. Under what conditions will either party invest in sustainability is an important question for both theory and practice. We take a principal-agent approach to analyze this two-stage supply chain and generate propositions on the strategies of rational players driven by the investment payoffs. We show that the key decision factors include the contract parameters (length, fee structure, etc.), the cost structure of the investments (upfront fixed cost as well as ongoing operational cost impact), supply chain characteristics (locations, market segment, scale), and the operator's sustainability performance.

We test the propositions using a novel dataset constructed from multiple sources: the operating statements of hotel properties under management contracts across the U.S., extensive interviews of hotel owners, and public records of real estate valuation and transaction history. We show that the contract parameters (length, fee structure, etc.), the relationship between the infrastructural and structural aspects of the sustainability investments, the cost structure of the investments (upfront fixed cost as well as ongoing operational cost impact), supply chain characteristics (locations, market segment, scale), and operator's sustainability performance contribute to the payoff variations, and subsequently the cooperation decisions.

Our research is the first in examining the contract design issues at the intersection of service supply chains and sustainable operations management. There have been parallel developments (e.g. the stakeholder theory) to applications of sustainability diffusion, but these developments have been lacking in exploration of the operational details during the contracting process. Managerially, the findings are likely to inform asset owners and hotel operators on designing and negotiating innovative contracts that are conducive to sustainable development. Finally, the findings suggest ways that public policy may supplement contracts in promoting sustainability investment.

LEAN MANAGEMENT AND SUPPLY MANAGEMENT: THEIR ROLE IN GREEN PRACTICES AND PERFORMANCE

Sara Hajmohammad
HEC Montreal
Montréal, Québec H3T 2A7, Canada
sara.hajmohammad@hec.ca

Stephan Vachon
HEC Montreal and CIRANO
Montréal, Québec H3T 2A7, Canada
stephane.vachon@hec.ca

Robert D. Klassen
Richard Ivey School of Business, University of Western Ontario
London, Ontario N6A 3K7, Canada
rklassen@ivey.uwo.ca

Iuri Gavronski
Universidade do Vale do Rio dos Sinos
Programa de Pós-Graduação em Administração
Sao Leopoldo, RS, Brazil
iuri@proxima.adm.br

ABSTRACT

Organizations are faced with an increasing pressure to engage in sustainable development and to integrate environmental and social dimensions into their traditional performance metrics. To date, a number of studies have linked a firm's environmental practices to its economic and operational performance. On the other hand, prior research suggests that lean management and supply management are potentially important determinants of environmental performance and can be seen as capabilities that ease the adoption of environmental practices. To help understand the role of lean and supply management in improving the firm's environmental performance, a conceptual model proposes that the extent of environmental investments (i.e., environmental practices) mediates the relationship between lean and supply management with environmental performance. We tested our model using plant-level survey data from a sample of Canadian manufacturing plants. The results indicate support for our conjecture that supply management, as well as lean activities, provide means by which environmental investments are encouraged. However, only lean activities are correlated with environmental performance through environmental investments.

Keywords: Environmental Performance, Environmental Investment, Supply Management, Lean Management, Mediated Regression Analysis

ROADMAP TO CHARTING A GREEN SUPPLY CHAIN

Rebecca Angeles, Faculty of Business Administration, University of New Brunswick, 7 Macauley Lane, Fredericton, New Brunswick E3B 5A3, Canada, rangeles@unb.ca, Tel. (506) 453-4962

Abstract

Firms have been pressured to implement green supply chains in response to a web of problems like global warming, environmental degradation, natural resource depletion, rising energy costs, interdependent global supply chains, and outcry from consumers and governments, among other drivers. This paper suggests a roadmap to answer the research question: "what are the logical steps a firm could take in pursuing a green supply chain strategy?" Recent developments in sustainability collaborative and knowledge sharing are explored as well.

Keywords: green supply chain; sustainability; supply chain management

INTRODUCTION

The next frontier for gaining competitive advantage is responding to sustainability and corporate social responsibility (CSR) call for "green" initiatives from a wide range of affected stakeholders. Sustainability has been defined as the capability of a firm "…to ensure that it meets the needs of the present without compromising the ability of future generations to meet their own needs." [53]. This definition is the offshoot of the commission appointed by the United Nations to study environmental problems and to suggest long-term strategies for sustainability. Subsequent articulation of the "sustainability" concept resulted in the following three frameworks: (1) triple bottom line, (2) the natural step, and (3) ecological footprint [48].

The "triple bottom line" approach calls for firms to balance its pursuit of economic, environmental, and social performance in their initiatives. The "natural step" believes that "…in a sustainable society, nature is not subject to systematically increasing concentrations of substances, extracted from the Earth's crust, concentrations of substances produced by society, or degradation by physical means and [where] people are not subject to conditions that systematically undermine their capacity to meet their needs…." [47]. The "ecological footprint" approach compares the environmental impacts of production activities with the known limits of the earth's natural resources and the capacity of its ecosystem functionality [29].

Simchi-Levi, Kaminsky, and Simchi-Levi [44, p. 1] define supply chain management as: "…a set of approaches utilized to efficiently integrate suppliers, manufacturers, warehouses, and stores, so that merchandise is produced and distributed at the right quantities, to the right locations, and at the right time, in order to minimize systemwide costs while satisfying service level requirements." This paper integrates the concepts of sustainability and supply chain

management. The Green Clean Institute defines "green strategy" as an approach that results in the manufacture of green products and use of business practices that cause no immediate, residual, or long-term harm to the biosphere [3]. Other academics have defined a "green supply chain" somewhat similarly.   Green et al. [17] refer to it as the approach taken when environmental considerations are taken in pursuing innovation in supply chain management and industrial purchasing. Narasimhan and Carter [35] focus on the role of the procurement function in the supply chain and view environmental supply chain management as the inclusion of reduction, recycling, reuse and the substitution of materials. Godfrey [13] succinctly refers to green supply chain management as the practice of monitoring and improving environmental performance in the supply chain.

Powerful supply chain hubs could leverage their market power to persuade their trading partners to make significant revisions in their business processes in a world where carbon emissions are no longer free, for instance [14]. Supply chain management professionals are also enjoined as an important stakeholder group to lobby government and our legislators to promote incentives to support environmentalism.

RESEARCH QUESTION

The key research question addressed in this paper is: " what are the logical steps a firm could take in pursuing a green supply chain strategy?" Thus, this paper showcases a roadmap for the planning and implementation of green supply chain initiatives.   Furthermore, recent developments in sustainability collaborative and knowledge sharing are explored as well.

DRIVERS FOR SUSTAINABILITY AND GREEN SUPPLY CHAINS

Important societal developments serve as drivers for sustainability [42]. First, industrialization has led to unsurpassed levels of physical good production and consumption, and global pollution and waste generation [42] [45] [4].

Second, "civil society" stakeholders in the form of non-governmental organizations (NGOs) and citizen social activists have raised the pitch of their demands for corporate transparency in the use of the earth's resources in producing goods and services [42] [24] [38]. Consumers demand to know a supplier's energy efficiency profile, approaches to health and safety its factories, methods of production used, sources of raw materials, impacts of discarding the product, percentage of the supplier's products made with recycled material, use of renewable versus non-renewable resources, etc. [3] [6]. Ultimately, the business case for pursuing green initiatives will rest upon consumer dollars, and it is only a matter of time before the currently laggard and indifferent segment of the consumer marketplace ultimately catches up. This, in turn, will significantly affect B2B commerce and affiliated supply chains in that firms will require their suppliers to be respondent to these consumer demands [3] [14]. There are, too, pressures exerted by domestic and foreign governments through the use of legislation and regulation to ensure compliance with green and sustainability mandates.

Third, disruptive technologies such as genomics, biomimicry, nanotechnology, information technologies, and renewable energy technologies stand to render energy-intensive industries obsolete [42].

Fourth, global economic and social problems that relate with resource depletion, deteriorating ecosystems, climate change and global warming, poverty and inequity in developing countries have raised awareness for the need for worldwide resource conservation and renewal [42] [49] [45] [18] [8].

RESEARCH METHODOLOGY

Using literature review, paying close attention to leading-edge firms that have blazed trails in pursuing sustainability, data was gathered to draw patterns of ways in which these firms organized their green supply chains. The results of two studies by CAPS Research form the framework of this paper: (1) study on the green issues on supply chain management systems as reported by executives who participated in a recent CAPS Research Critical Issues Partnership conference [3] and (2) a series of interviews and in-depth discussion with executives from CAPS Research sponsoring firms and other thought leaders [19]. CAPS Research is a major nonprofit research organization that focuses on conducting both academic and practical studies on supply chain management issues since 1986. Additional material from other academic studies was used to supplement the guidelines indicated by these two studies.

Green supply chain management is in a fairly nascent stage of research [26] [33] [45] [25] [51]. At this time, it seems appropriate to sketch the early beginnings of a roadmap that could guide firms that are considering sustainability initiatives.

ROADMAP TO IMPLEMENTING A FIRM'S GREEN SUPPLY CHAIN

The following are the major steps suggested to firms intending to implement a green supply chain.

Step 1: establish the leadership for sustainability

The chief executive officer (CEO) typically deputizes a sustainability overseer, who may, then, convene a cross-functional steering committee, council, or team of some kind to support the firm's initiative [20]. Chief Sustainability Officer (CSO) is a typical title used among a variety of equivalent ones for the person who needs to move the firm through the stages of progression as the firm seeks more challenging targets [26]. Key leadership capabilities for the early stages of promoting sustainability are competencies in collaboration and influencing and change leadership [27]. Supply chain management executives must coordinate with top management in undertaking the following: (a) rolling out CEO sustainability statements to suppliers; (b) establishing rules of engagement for commodity manager interactions with suppliers; (c) deciding if sustainability criteria should be included in supplier certification; and (d) deciding the rollout of a supplier code of conduct for supporting sustainable purchasing practices [19].

Step 2: identify your firm's stage of growth in the sustainability path and act accordingly

Senge et al. [42] articulate the five stages in pursuing GSCM initiatives in a firm: (1) stage one: noncompliance with governmental sustainability directives and/or market pressures to become "green"; (2) stage two: compliance with these directives and demands; (3) stage three: going beyond compliance by proactively seeking ecological efficiencies, anticipating and neutralizing regulatory threats, and savings and financial payoffs start to outweigh the costs of initial investments; (4) stage four: pursuing an integrated sustainability strategy: actively exploiting business opportunities presented by the need for sustainability and enabling this initiative to directly impact internal capital and budget allocations, supply chains, exploitation of new markets, development of core operations, and research and development efforts; and (5) stage five: integrating sustainability in the firm's mission and purpose: cultivating sustainability as a core corporate value and pursuing it both internally and externally. Most firms probably find themselves in stages one and two.

A multifaceted challenge that needs to be addressed is to demonstrate improved economic and firm performance involves a number of issues: (1) the lack of a firm theory that governs the relationship between sustainability and economic performance [23] [50]; (2) the paucity of practical experience among firms that are leading sustainability and compliant with regulatory requirements in the current marketplace; and (3) methodological challenge of comparing different firm experiences in pursuing sustainability [23].

Firms in stage three make their supply chains sustainable by taking measures like reducing their consumption of both nonrenewable resources like coal, petroleum, and natural gas and renewable resources like water and wood [37]. They also work with their suppliers to procure eco-friendly raw materials and components and with their retailers in meeting "green" products and services that customers look for.

Firms in stage four are proactive in using sustainability in designing new products and services and/or developing new business models as well. They actively respond to consumers' demand for ecological products and offer redesigned old products or brand new products in the marketplace [37].

In developing new business models, firms in stage four presented alternative ways of doing business and offering new delivery mechanisms for sustainable products and services [37]. Waste Management, a garbage disposal market leader, created "Green Squad" to generate revenue from the waste it collected. Green Squad collects recyclable electronic waste products deposited in landfills and sells them to Sony in the U.S.

Firms in stage five make a pioneering declaration to pursue a corporate mission to contribute to society and be regenerative [42]. Interface, for instance, defines itself as being a "restorative enterprise" that does not use raw materials that harm the biosphere and does not extract any raw material from the earth that could not be recycled or rapidly regenerated.   While firms in stage five could skip all other four stages, it is possible for a firm to graduate to stage five from a lower level stage.

Step 3: explore and apply sustainability methodologies to help establish performance metrics

The firm needs tools and methodologies it can use in pursuing sustainability to assist in measuring results and performance. One such tool that has a fairly broad following is "life cycle assessment" or LCA, which is used to assess a physical products' environmental impact throughout its life from "cradle to grave", or "cradle to cradle," in which case the product is recycled and used to produce new products [42] [40]. The International Organization for Standardization or ISO has standardized the methodology now used by firms to reduce the impacts and maximize benefits of physical products from the time resource inputs are obtained through to product manufacturing, product use, and its end of life. The ISO 14000 certification requires firms to identify all environmental impacts of their operations, safe handling and disposal procedures for hazardous materials and waste, and their level of compliance with the relevant environmental regulations [11]. Supplier selection has already been impacted with firm compliance with ISO 14000 standards [7] [32].

LCA is used to analyze the profile of the raw materials (e.g., metals, energy carriers, plastics, bio-materials, etc.) used in the production of the physical good. LCA is also used to examine all the stages involved in the product life cycle and evaluate environmental trade-offs that could be made (e.g., a change in a production step could lead to lower air emissions but this could also increase water emissions) [42]. Procter & Gamble (P&G) used LCA and discovered that the worst impacts of the Tide laundry product was not during the extraction of raw materials to produce it nor the resources needed to transport it to distribution centers or retail outlets [15]. Rather, it was the necessity of using heated water when consumers washed their clothing that was the problem. As a result of this finding, P&G developed "Tide Cold Water," a product that no longer requires the use of hot water.

Another methodology is the LEED rating system used to adjudge the green characteristics of [42]. LEED uses a point rating system --- the more points are gained, the "greener" the building is. A five-part framework governs the LEED rating system: (1) sustainability of the site (i.e., erosion control, transportation, and pollution reduction); (2) water efficiency; (3) energy and atmospheric pollution; (4) materials and resource use; and (5) indoor environmental quality (i.e., indoor chemical pollution and thermal comfort).

Step 4: make a business case for the green supply chain

Economic performance can be interpreted as reduced operations costs or increased revenues, market share, sales growth, etc. Economic performance and benefits manifest very differently in various industries since their value chains would understandably be specific and unique to the nature of the extractive and transformation processes involved in each of them.

New ways of using what used to be waste products can bring important financial benefits. General Mills, for instance, used to trash the oat hulls that were a by-product of manufacturing the renowned Cheerios household cereal brand. Now, the firm recycles 86 percent of this solid waste and burn it as fuel, thus, enabling it to earn more from it than it spent on its disposal [42].

Higher-order benefits are reaped by firms that introduce new products and services and place sustainability within the core of their design concept. Plastic bottled water, which has led to the clogging of landfills, has received its share of criticisms from the WorldWide Fund for Nature and Corporate Accountability International. This, in turn, prodded Brita to replace its bottled water with the reconceptualized Brita purifying systems and water filters, which has led to a jump in water pitcher sales to 23 percent as against 2 percent for the category overall [42].

Clorox, was moved by the research finding that use of household cleaning products was a major environmental concern, next only to the use of cars [37]. Consequently, Clorox introduced its line of nonsynthetic cleaning products in its Green Works product line, which was the result of three years of development and more than $20 million in product innovation costs. This was quickly recouped by the end of 2008 when Green Works had grown the U.S. natural cleaners market by 100 percent and Clorox garnered 40 percent of the $200 million market.

The most dramatic benefits are derived when a firm changes its business model altogether to address sustainability issues. The accumulation of used and discarded large office machines and technology equipment has been a major blight in landfills so much so that it has constituted a significant environmental problem.

Fuji Xerox resolved this issue by reframing its identity from a firm manufacturing and selling photocopy machines to one that leases them [12]. Cognizant of their contribution to this environmental problem, Fuji Xerox decided to lease its machines instead and thus, gain control over the disposal and recycling of its machines. As a result, it recovered 99 percent of the materials it could still use from used equipment in Asia in 2006, thus, reducing raw material consumption in its Chinese factories by 2,000 tons.

Waste Management, a $14 billion market leader in garbage disposal, changed its business model from waste disposal to waste recovery [37]. It estimated that it was hauling about $9 billion worth of reusable materials to landfills that could be reused to generate income. The firm, thus, set up a business unit called Green Squad whose mission was to generate income from waste products. In a major coup, Green Squad partnered with Sony Corporation of America to gather and recover salvageable electronic waste.

Evidence also exists in academic research studies that sustainability initiatives could lead to the firm's economic performance. Montabon et al. [34] performed content analysis on 45 corporate social responsibility reports and through the application of canonical correlation, found a positive relationship between environmental management practices (EMPs) and firm performance

King and Lenox [21] analyzed a sample of publicly traded U.S. manufacturing firms both listed in the Compustat database and have at least one facility meeting the reporting requirements of the U.S. Environmental Protection Agency's Toxic Release Inventory (TRI). They found a positive and significant relationship between the various means of reducing pollution and profitability as measured in terms of return on assets and a statistic called "Tobin's q" which they found to better reflect the inherent value of the firm and reflects expected future gains in accordance with the more recent "pays to be green" studies.

Russo and Fouts [41] studied data on 243 firms that had received environmental ratings of the Franklin Research and Development Corporation (FRDC) over a two-year period. FRDC ratings were based on the following criteria: compliance records, expenditures, and other initiatives used to meet new demands to reduce waste reduction and support environmental protection organizations. They found that high FRDC environmental ratings were positively associated with a firm's return on assets.

In a survey of procurement managers associated with the National Association of Purchasing Managers (NAPM), Carter et al. [5] found that firms that undertake green procurement practices had better economic performance. Green procurement was represented by an index called "EPINDEX," which is a composite score of all environmental practices used by the firm. Green procurement was positively associated with net income and negatively correlated with cost of goods sold.

Step 5: implement the firm's green supply chain plan

5a: implement an internal green supply chain strategy

Although the organizational, cultural, and business process changes required to comply could be daunting and costly, the path of "least resistance" is taken by firms that simultaneously address operational efficiency and cost reduction measures that are long due.

One firm that was forced to adopt an internal GSCM strategy is Sachsenmilch AG, a German firm based in Saxony, East Germany, that manufactures 25 different kinds of cheese and milk products [52]. One important driver for the firm's green initiative was compliance with the European Union (E.U.) regulation that requires food manufacturers to document the manufacturing process for transparency should the firm need to be audited by health officials when problems arise. Thus using RFID, Sachsenmilch AG upgraded its cheese production processes to make them more efficient. RFID has the greatest impact on ripening and ageing processes where a production machine is used to monitor cheese racks as they find their way into the ripening queue line and in and out of "sweating" and "ripening" rooms. The RFID system in this part of the production process uses 16 interrogation points that record each rack's ID number and arrival time in each room to provide information visibility.

5b: implement an external green supply chain strategy

Implement external sustainability strategies is needed if firms are to optimize their overall performance by rationalizing the green operations of their supply chain trading partners, both on the downstream and upstream segments of the value chain. The Gap put into effect green sourcing guidelines for their suppliers, with an accompanying code of conduct in 1996. The Gap also publicly announces the suppliers it continually buys from and those with whom they ceased relationships due to noncompliance with these green policies.

Wal-Mart announced the implementation of its "Sustainable Product Index" that has unfolded in three phases [39]. In the first phase, Wal-Mart's suppliers will be asked to respond to a 15-question form covering the topics of energy and climate, natural resource use, material

efficiency, and people and community. In the second phase, Wal-Mart has created a "Sustainability Consortium" of universities administered by Arizona State University and the University of Arkansas to assist firms in working with their suppliers and retailers and with government and non-governmental organization in obtaining data for a global database of information on the lifecycle of products from raw material acquisition to disposal. In the third phase, a sustainability index will be developed to create a rating system needed to produce credible and accessible information about the sustainability of the products. Compliant and high-scoring suppliers will be given preferential treatment in things like shelf space access and visibility, among others, at Wal-Mart retail stores.

Working with members of one's supply base (i.e., the upstream segment of the value chain) in meeting a firm's sustainability requirements should be handled with caution. Sustainability supplier performance scorecards created from sustainability indexing methodologies should be implemented consistently so as not to alienate suppliers [9]. Preventive measures would include consulting with suppliers to ensure that required data inputs for the scorecard are actually measurable. Also, suppliers need to know how scorecard feedback will be delivered and what the consequences of the reports will be way ahead of time.

Cementing trading partner participation in such green initiatives will need to be backed by incentives and to a minimal extent, by penalties [16]. Formal corporate social responsibility (CSR) reports have been used as a means of recognizing valued complying suppliers and flagging noncompliant ones. "More enlightened" firms have used incentives such as bonuses and contract extensions to reward compliant suppliers. While contract termination is an extreme measure to use for noncompliant suppliers, there are intermediate consequences that may be put into effect such as a "design hold" from engineering a supplier's parts into new products until defects indicated by CSR reports are corrected, or graduated purchase volume step-downs.

Earthster.com (external upstream participation)

The upstream component of the a collaborative green supply chain is currently exemplified by the pioneering website, Earthster (www.earthster.com), which uses a combination of information technologies to undertake life cycle assessment (LCA) in tracking a physical product's environmental footprint and sustainability impacts from cradle-to-cradle. Life cycle assessment refers to method of analyzing a physical product's impact on the environment, tracing its life all the way from the extraction of raw materials through to the manufacturing process, use of the product by consumers, and the final disposal of the product [1]. Among other things, this initiative was motivated by the need of firms to make credible and valid claims about its environmental impacts especially when its brand image is at stake, during these times when buyer firms and consumers are increasingly green conscious.

An important contribution that the LCA methodology offers is to clarify the difference between "absolute" and "relative" environmental product impacts. An example of an absolute impact is the claim that a firm makes that the manufacture of its product releases 30 pounds of carbon dioxide to the atmosphere. In making a relative impact claim, on the other hand, a firm would say that the manufacture of its product releases 25 points less carbon dioxide than its nearest competitor product. One important web portal where this initiative has been launched and made

the LCA methodology available worldwide is an entity called Earthster, founded by Greg Norris, who is with Sylvatica, a consulting firm and is also a lecturer on LCA at Harvard University. Through the Earthster portal, participating firms can share LCA findings such as carbon emissions into water, soil, and air, without comprising their identities or those of their trading partners such as suppliers. Earthster offers a web-based software product that allows firms to conduct LCA on a specific product's raw materials extraction and manufacturing stages in order to compare this product's performance with those of similar products based on published data [1]. LCA analysis results are "certifiable" in that the data submitted to Earthster cannot enter the portal's system without being subjected to independent audits.

Firms may choose to publicize the results of the LCA analysis so that it is shared with other portal members, while stipulating that the proprietary data underlying the findings remain confidential. A hub firms (i.e., usually a buyer firm) can also be assisted by Earthster to encourage its key suppliers to conduct their own impact assessments in order to generate hard data needed to establish their products' environmental impacts. Suppliers may choose to report their products' current year's performance, yearly incremental improvements in performance, or both. This service is especially valuable to small- and medium-sized suppliers that do not have the expertise or wherewithal to conduct this type of analysis and usually have products that are at a baseline of marginal, if not outright inferior environmental or social performance. Earthster offers software tools freely and are open source that take advantage of the Semantic Web technologies and among the tools offered are those that address: (1) environmentally preferable product purchasing (EPP); (2) business process and/or product optimization; (3) corporate benchmarking, metrics for continuous improvement; (4) publishing cradle to gate Life Cycle Inventory (LCI) data for downstream purchasers; and (5) publishing cradle to grave Life Cycle Impact (LCIA) data for downstream purchasers [30]. The open-source foundation of Earthster brings with it specific benefits: (1) these software projects deliver the products at no cost; (2) the software products are fully transparent and may be modified by an open community of programmers or developers; and (3) the developed software modules can be used with other open source libraries. The data shared in this digital "green commons" allows the dissemination and sharing of sector averages for a particular business process or physical product and enables small- and medium-sized firms with no resources to evaluate the environmental impacts of their products.

GoodGuide.com (external downstream participation

GoodGuide.com is Earthster.com's counterpart in the downstream segment that provides information to final consumers about 75,000 food, toys, personal care, and household products culled from both public and private databases in order to score each on a 10-point scale that captures the products' health, environmental, and social impacts [36]. Products earn both individual and collective ratings.

Bolstered by a collaborative platform supported by a wide range of non-profit organizations, GoodGuide emerged from Berkeley's Sustainability Information Lab and has been financially supported by the Wallace Global Fund, the Overbrook Foundation, the New Place Fund, the National Collegiate Inventors Association, and other major venture capital firms. Other well-known firms and universities that have supported GoodGuide.com include Google, Amazon,

eBay, PayPal, Intuit, Massachusetts Institute of Technology (MIT), and the University of California.

GoodGuide.com evaluates products primarily on three criteria --- health hazard assessment, environmental impact assessment, and social impact assessment using a number of scientific methods (www.goodguide.com). Much more granular analysis is conducted for each of the three criteria for specific issue areas, such as climate change policies, labor concerns, and product toxicity. The website currently uses a database that stores information on approximately 600 base criteria in evaluating products and firms.

 Under "health hazard assessment," GoodGuide considers both the impacts of products on a person's health (i.e., cancer risks, reproductive health hazards, mutagenicity (i.e., possessing an agent, such as a chemical, ultraviolet light, or a radioactive element, that can induce or increase the frequency of mutation in an organism), endocrine disruption, respiratory hazards, skin and eye irritation, etc.) and the firm's operations and management policies as they affect their employees and local communities. GoodGuide health hazard assessment ratings are not, per se, hard assessments of the products or chemicals used in them but they are indicators of potential hazards associated with these products.

Under "environmental impact assessment," GoodGuide compiles data on the life cycle impacts of products from the time they are manufactured through to the time they are distributed, sold, and ready for final disposal and generates an overall rating. The data points cover issues like environmental emissions and the impacts on air, water, land, and the climate; natural resource impacts; and environmental management programs.

Under "social impact assessment," GoodGuide aggregates data on the firm's corporate policies as they affect employees' compensation, labor and human rights practices, diversity policies, and working conditions.

GoodGuide ratings are shown as data points in a scale of 1 (worst score) through to 10 (best score). These scores represent a product or firm's performance compared with their nearest competitors.

GoodGuide makes savvy use of up-to-date information technologies to promote its mission. In addition to using open source technologies that allows the portal's data to be exposed to other business applications, GoodGuide also makes its information available to consumers on the go via an iPhone app. Consumers could receive text messages about products they are interested in based on their universal product code (UPC) and the iPhone app response will follow within weeks [28]. More than 100,000 consumers have already downloaded this app as of September 2009 [36]. GoodGuide also employs high-level data mining techniques using the services of Unica's NetInsights system to constantly monitor how visitors use their portal, what information is most valuable to these visitors, and possibilities for search customization and personalized product ratings based on factors most important to a particular consumer [36].

GoodGuide has also explored multiple avenues to disseminate its findings. GoodGuide ratings are also found in TheFind.com website and links from GoodGuide to purchase at Amazon.com

have five to ten times the conversion rates of the giant retailer's norms [36]. As of September 2009, GoodGuide was negotiating with three offline retailers to enable the incorporation of its ratings into these three stores to enhance their customer buying experience.

At the moment, information retrieved from GoodGuide from the consumer side is free. There's a possible plan to introduce a business model that will allow the site to generate income by licensing data to manufacturers, retailers, and firm procurement managers [36].

SUSTAINABILITY COLLABORATIVE NETWORKS: BRIDGING STAKEHOLDER INTERESTS

The other avenue to promote sustainability efforts is through the use of sustainability collaborative networks. Because of the extensive range of coordination of efforts needed from both the public and private sectors, a collaborative network presents itself as a logical alternative for synchronizing the activities of all affected stakeholders. A prime example of this type of network is the Green Suppliers Network (GSN), a partnership involving firms in the private sector, the U.S. Department of Commerce, Manufacturing Extension Partnership (MEP), and the U.S. Environmental Protection Agency (EPA), created to assist large manufacturers improve or radically change their business processes in order to achieve green results [10]. This type of network is especially useful to small- and medium-sized firms that do not have the resources or expertise to deal with sustainability issues. To ease the transition to green practices, the GSN has incorporated lean practices with green initiatives. Traditional lean manufacturing practices already include materials and energy improvements in their templates as these activities lead to lower costs and the reuse of materials and recycling of waste materials.

Steelcase has used the GSN program to target suppliers proposing price increases for the goods they sell. Steelcase finds that the GSN assessment procedure is a logical way of investigating the reasons behind the proposed price increase. The GSN procedure also results in recommendations based on hard data that could lead to cost reductions and higher efficiencies for the suppliers --- which the buyer firm can use to mitigate the planned price increase.

There have been positive results arising from participation in GSN. Participating manufacturers have reported an average cost savings of US$585,000; on the other hand, about 15 suppliers that have used their services reported a total increase in sales of about $3.8 million.

THE NEXT LEVEL: EMERGING GREEN VALUE WEBS FOR KNOWLEDGE MANAGEMENT

The next level of B2B portals supporting sustainability focuses on collaborative knowledge management initiatives and one of the pioneering models of this is The GreenXchange initiated by Nike which is a Web-based B2B exchange with a licensing structure where firms could share intellectual property related to sustainability methodologies, products, and services and collaborate with each other and issue licenses to allow other firms to access patented research [46] [31].

Nike's key contribution is its ability to model the confidence required of a leader firm that could inspire confidence in going beyond the barriers posed by fear of losing intellectual property assets. Nike, instead, perceived its patents as transferrable assets which could, in fact, not only be revenue generating but also more innovative when shared even with its nearest competitors. Nike expects to have access to new discoveries of firms that purchase the rights to gain access to its patents as well.

GreenXchange has a database that stores titles of intellectual property information, which could also be represented as visual maps that would allow interested firms to view the patent resources currently in use and those indicative of breakthrough technologies.

Patent holders reserve the right to specify the terms of use and cost of access and create the contract that other interested firms would have to accept before being allowed to view patent information. Patent holders also protect their proprietary information by screening the types of firms that could gain access to this information. As the leader of this initiative, Nike has committed and placed 400 patents for sustainability related research to stimulate innovation among other firms in various industries. Cross-industry intellectual pollination is highly encouraged. For instance, Nike has made available its know-how for using environmentally preferred rubber which uses 96 percent fewer toxins than its original formulation. By making this expertise available on GreenXchange, even a competitor firm like Mountain Equipment Co-op could use the information for its bicycle inner tubes.

ISSUES AND CHALLENGES

Vantage Partners and Integrity Interactive surveyed supply chain and procurement managers in December 2008 and found that while more than 70 percent of them considered corporate social responsibility (CSR) "pretty important" or "extremely important" in their firms, barely 20 percent of them claim that they had made CSR compliance a prerequisite in choosing their suppliers [16]. And then, once a supplier is selected, only 56 percent of the firms required their suppliers to meet CSR standards. Less than 33 percent of the firms failed to meet their CSR standards and they did not even consider noncompliance as grounds for terminating the contract with the suppliers involved. Another 15 percent of the firms do not even look at their suppliers' CSR capabilities in making supplier selection. Another 52 percent considered supplier CSR preparedness as only one of the many factors in their supplier assessment procedures. Only 37 percent of the firms used metrics to find out if suppliers meet their CSR goals.

One effective way to overcome hurdles is to collect successes and expose the difference these make for the firm. Skeptics in the firm are far better convinced by quantitative (i.e., bottomline) rather than qualitative (i.e., brand equity) evidence [27]. PricewaterhouseCoopers in their report called "Achieving Superior Financial Performance in a Challenging Economy, 2008," found that firms engaged in sustainability initiatives in the food, beverage, and consumer products industries outperformed their nearest competitors by earning a higher return on assets, stronger, cash flow, enhanced shareholder returns, higher gross margins, and higher returns on sales [2]. This study also found that the performance of 60 big firms that reported sustainability metrics to concerned stakeholders consistently outperformed their competitors.

Firms that appreciate the potential competitive advantage of working ahead of the curve and being forward looking, stand significant chances of winning in the long term.

REFERENCES

[1] Ace, J. *Environmental Decisions Driven by Data*. 2010. Retrieved July 10, 2010, from http://www.awarenessintoaction.com/whitepapers/Life-Cycle-Analysis-Data-Environmental-Business-sustainability.html

[2] Albinson, T. *How Green Is Your Supply Chain?* Inside Supply Management, 2009, May, 20 (5), 32-33.

[3] Ashenbaum, B. *Critical Issues Report: Green" Corporate Strategies: Issues and Implementation From The Supply Management Perspective.* Tempe, Arizona: CAPS Research, 2008, November. Retrieved June 1, 2010, from http://capsresearch.org/Research/Critical_Issues.aspx

[4] Beamon, B.M. *Designing the Green Supply Chain*. Logistics Information Management, 1999, 12 (4), 332-342.

[5] Carter, C.R., Kale, R., & Grimm, C.M. *Environmental purchasing and firm performance: an empirical investigation*, Transportation Research, 2000, 36, 219-228.

[6] Case, S. *Socially Responsible Purchasing*. Inside Supply Management, 2007, January, 18 (1), 8.

[7] Chen, C. *Incorporating green purchasing into the frame of ISO 14000*. Journal of Cleaner Production, 2005, 13 (9), 927-933.

[8] Commoner, B. Making Peace with the Planet. New York, NY: Pantheon, 1990.

[9] Cooling, L. *Scorecards for Sustainability*. Inside Supply Management, 2008, October, 19 (10), 23-25.

[10] Cooling, L. *One Green Turn Deserves Another*. Inside Supply Management, 2008, August, 19 (8), 30-33.

[11] Corbett, C.J., & Kirsch, D.A. *International diffusion of ISO 14000 certification*. Production and Operations Management, 2001, 10 (3), 327-342.

[12] Fuji Xerox. *Press Release: Fuji Xerox Innovation Makes Business and Environmental Sense*. 2007, October 17. Retrieved June 25, 2010, from http://www.sustained.com.au/index.php/Profit-case-studies/Fuji-Xerox-innovation-makes-business-and-environmental-sense.html

[13] Godfrey, R. *Ethical purchasing: developing the supply chain beyond the environment*. In: Russel, T. (Ed.) Greener Purchasing: Opportunities and Innovations, Greenleaf Publishing, Sheffield, England, 1998, 244-251.

[14] Goldenhersh, L. *Supply Chain Environmentalism*. Inside Supply Management, 2009, March, 20 (3), 30-31.

[15] Goleman, D. Ecological Intelligence: How Knowing The Hidden Impacts of What We Buy Can Change Everything. New York, NY: Broadway Books, 2009.

[16] Gordon, M., & Levin, M. *Working with Suppliers to Achieve CSR Goals*. Inside Supply Management, 2009, June, 20 (6), 30-33.

[17] Green, K., Morton, B., & New, S. *Purchasing and environmental management: interaction, policies and opportunities*. Business Strategy and Environment, 1996, 5, 188-197.

[18] Hart, S.L. *A Natural-Resource-Based View of the Firm*. Academy of Management Review, 1995, 20 (4), 986-1014.

[19] Hershauer, J. *Critical Issues Report: Process Guide for Supply Management Environmental Sustainability*. 2008, October. Tempe, AZ: CAPS Research. Retrieved June 1, 2010, from http://capsresearch.org/Research/Critical_Issues.aspx

[20] Hollender, J., & Breen, B. The Responsibility Revolution: How The Next Generation of Businesses Will Win. San Francisco, CA: Jossey-Bass, 2010.

[21] King, A., & Lenox, M. *Exploring the Locus of Profitable Pollution Reduction*. Management Science, 2002, February, 48 (2), 289-299.

[22] Klassen, R.D., & Whybark, D.C. *Environmental management in operations: the selection of environmental technologies*. Decision Sciences, 1999, 30 (3), 601-631.

[23] Lankowski, L. *Corporate Responsibility Activities and Economic Performance: a Theory of Why and How They Are Connected*. Business Strategy and the Environment, 2008, 17, 536-547.

[24] Lash, J., & Wellington, F. *Competitive on a Warming Planet*. Harvard Business Review, 2007, 85 (3), 94-102.

[25] Linton, J.D., Klassen, R., & Jayaraman, V. *Sustainable supply chains: An introduction*. Journal of Operations Management, 2007, 25, 1075-1082.

[26] Lubin, D.A., & Esty, D.C. *The Sustainability Imperative*. Harvard Business Review, 2010, May, 88 (5), 42-50.

[27] Lueneburger, C., & Goleman, D. *The Change Leadership Sustainability Demands*. MIT Sloan Management Review, 2010, 51(4), 49-55.

[28] Madrigal, A. *Startup Fights Greenwashing With Data Sent to Your Phone*. Wired Science. 2008. September 10. Retrieved July 5, 2010, from http://www.wired.com/wiredscience/2008/09/website-launche/

[29] Marshall, J.D., & Toffel, M.W. *Framing the Elusive Concept of Sustainability: A Sustainability Hierarchy*. Environmental Science and Technology, 2005, 39 (3), 673-682.

[30] Maxfield, J.P. *Earthster --- Making Life Cycle Assessment More Accessible*. 2007, February 21. Retrieved June 15, 2010, from http://inspiredeconomist.com/2007/02/21/earthster-making-life-cycle-assessment-more-accessible/

[31] Mazur, A. *GreenXchange: Creating a Meta-Map of Sustainability*. 2009, May 5. Retrieved May 25, 2010, from http://www.worldchanging.com/archives/009822.html

[32] Miles, M.P., Munilla, L.S., & Mcclurg, T. *The impact of ISO 14000 environment management standards on small and medium sized enterprises*. Journal of Quality Management, 1999, 4 (1), 111-122.

[33] Mollenkopf, D., Stolze, H., Tate, W.L., & Ueltschy, M. *Green, lean, and global supply chains*. International Journal of Physical Distribution & Logistics Management, 2010, 40 (1/2), 14-41.

[34] Montabon, F., Sroufe, R., & Narasimhan, R. *An examination of corporate reporting, environmental management practices and firm performance*. Journal of Operations Management, 2007, 25, 998-1014.

[35] Narasimhan, R., & Carter, J.R. Environmental Supply Chain Management. Tempe, AZ: The Center for Advanced Purchasing Studies and Arizona State University, 1998.

[36] Neff, J. *GoodGuide puts brands' ethical claims to the test*. Advertising Age, 2009, September 28, 80 (32), 1-26.

[37] Nidumolu, R., Pralahad, C.K., & Rangaswami, M.R. *Why Sustainability Is Now The Key Driver of Innovation*. Harvard Business Review, 2009, September, 87 (9), 56-64.

[38] Porter, M.E., & Kramer, M.R. *Strategy & Society: The Link Between Competitive Advantage and Corporate Social Responsibility*. Harvard Business Review, 2006, 84 (12), 78-92.

[39] Quinn, B. *Walmart's Sustainable Supply Chain*. 2009, September. Retrieved July 4, 2010, www.pollutionengineering.com

[40] Rex, E., & Baumann, H. *Implications of an Interpretive Understanding of LCA Practice*. Business Strategy and the Environment, 2008, 17, 420-430.

[41] Russo, M.V., & Fouts, P.A. *A Resource-Based Perspective On Corporate Environmental Performance and Profitability*. Academy of Management Science, 1997, 40 (3), 534-559.

[42] Senge, P., Smith, B., Kruschwitz, N., Laur, J., & Schley, S. The Necessary Revolution: Working Together to Create a Sustainable World. New York, NY: Broadway Books, 2010.

[43] Shrivastava, P. *Environmental Technologies and Competitive Advantage*. Strategic Management Journal, 1995, 28 (6), 183-200.

[44] Simchi-Levi, D., Kaminsky, P., & Simchi-Levi, E. Designing and Managing the Supply Chain: Concepts, Strategies and Case Studies. New York, NY: McGraw-Hill/Irwin, 2008.

[45] Srivastava, S.K. *Green supply-chain management: A state-of-the-art literature review*. International Journal of Management Reviews, 2007, 9 (1), 53-80.

[46] Tapscott, D. Davos: *Nike and Partners Launch the GreenXchange. BusinessWeek.* 2010, January 27. Retrieved June 5, 2010, from http://www.businessweek.com/the_thread/techbeat/archives/2010/01/davos_nike_and.html

[47] The Natural Step International. *"What is Sustainability?"* The Natural Step website, 2008. Retrieved June 20, 2008, from http://thenaturalstep.org/com/What_is_sustainability/

[48] Toffel, M.W., & Lee, K. Sustainability at Millipore. Harvard Business School Case Study No. 9-610-012. Cambridge, MA: Harvard Business School Press, 2009, December 10.

[49] Tsoulfas, G.T., & Pappis, C.P. *Environmental Principles Applicable to Supply Chain Design and Operation*. Journal of Cleaner Production, 2006, 14 (18), 1593-1602.

[50] Ullmann, A. *Data in search of a theory: a critical examination of the relationships among social performance, social disclosure, and economic performance of U.S. firms*. Academy of Management Review, 1985, 10, 540-557.

[51] Vachon, S., & Klassen, R.D. *Extending green practices across the supply chain*. International Journal of Operations & Production Management, 2006, 26 (7), 795-821.

[52] Wessel, R. *RFID Ripens Cheese*. RFID Journal, 2007, January 8. Retrieved July 25, 2010, from http://www.rfidjournal.com/article/purchase/2906

[53] World Commission on Environment and Development (WCED). Our Common Future. New York, NY: Oxford University Press, 1987.

BUYING AN ALTERNATIVE FUEL VEHICLE

FARIBORZ PARTOVI, Drexel University, Philadelphia, PA 19104, partovi@drexel.edu
REBECCA KIM. Drexel University, Philadelphia, PA 19104, beckim@deloitte.com

ABSTRACT

Selecting an alternative fuel vehicle is a complicated decision-making process that requires considering knowledge of various characteristics of a vehicle and ability to compare these characteristics with each other. This may be not an easy task for average buyers. In this paper we purpose a modified DEA model that focuses on this shortcoming and provides an alternative model that does not require subjective comparison of criteria by the decision makers Finally, the paper provides a numerical example.

Keywords: Alternative fuel vehicle, Environment sustainable products, Data envelopment analysis

INTRODUCTION

According to the monthly energy report published by the US Energy Information Administration (EIA) 99,380 trillion BTU of energy was consumed in 2008 (US Energy Information Administration Monthly Review (December 2010).Of the total energy consumption, the EIA reports four main sectors (residential, commercial, industrial, and transportation) and their respective total energy consumptions. As seen in Figure 1, transportation sectors account for 28% of the total US energy consumption. Hence, green projects to decrease one's carbon footprint in transportation sector can make a big impact in the total energy consumption. In fact, new businesses have emerged that offer "carbon footprint consulting" services. These green firms provide clients with consulting services that help them in all aspects of their home/business in regards to their carbon output. Traditionally, environmental consulting ensures that clients maintain an appropriate measure of compliance with environmental regulations. Whereas environmental consulting assists their clients in reacting to environmental regulations, carbon consulting assists their clients in taking proactive measures to measure and to decrease total energy consumption.

Figure 1: 2008 Energy
Consumption
by Sector

- 22% Residential
- 19% Commercial
- 31% Industrial
- 28% Transportation
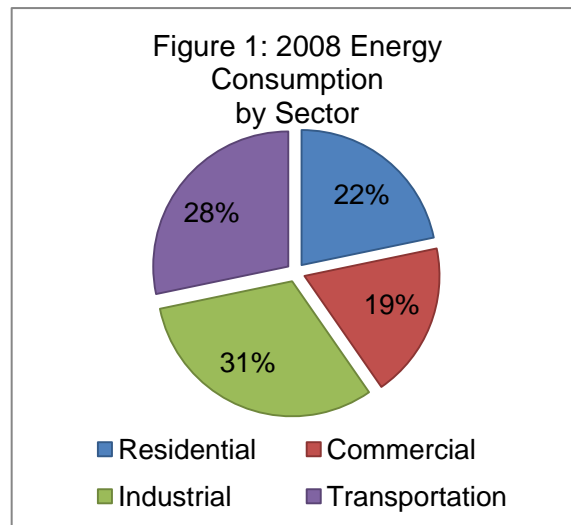
In the one's local home environment, the largest source of energy consumption is automobiles. In 2008, the largest source of energy for transportation was petroleum (oil) which originates from fossil fuels, a finite energy source. In fact, the US consumes about 22.2 million barrels/day – about 24% of the total global petroleum consumption (US Energy Information Administration International Petroleum Consumption 2010). Over the next ten years, the EIA projects that the US and global demand for petroleum will continually grow. However, many experts project that oil reserves will be depleted in the next 40 years and fossil fuels will be depleted in the next 250 years. Hence, it is imperative that fossil fuels are conserved for the next generation and that alternative energy sources are thoroughly researched. Currently, many consumers have already begun implementing alternative energy sources in their commodity selections. When deciding which types of alternative energy sourced to choose from, there are many criteria to consider. In addition, the weight of these criteria can differ with respect the decision makers' perspective. When decision- makers cannot reach agreement on assigning weights to the criteria, because of newness of the product, lack of knowledge, and personal preferences then one has to rely on decision making tools that are data driven. Data envelopment Analysis (DEA) is one of these decision making tools which is based on linear programming and can facilitate decision processes by reducing the number alternatives to a small group of efficient choices and with some modification select the most efficient alternative among various choices. In this manuscript we will demonstrate how DEA is used to fulfill a decision to purchase an alternative fuel vehicle.

The use of consumer durables such as automobiles generates multiple economic benefits and costs to consumers. In addition to the economic impacts, consumer durables also cause pressures on the ecosystem. Eco-efficiency of a consumer durable refers to the capability to produce net economic benefit

by not polluting the environment and using natural resources as little as possible. The more economic benefits a certain product can produce for given economic and environmental costs, the more eco-efficient it is and the more value it provides. Environmental pressures occur throughout the product's life-cycle, including the production, use, and disposal phases. For most consumer durables, the use phase creates the main environmental burden. Moreover, in practice, it is very difficult to attribute the environmental burden from production and disposal stages to a specific product. To assess the performance or eco-efficiency of consumer durables, it is natural to consider a consumer durable as a production unit that demands inputs (such as energy, water, and other resources) to produce outputs (desirable services and undesirable environmental effects). Adopting this perspective enables us to apply the modified DEA models to select the best alternative fuel vehicle. In particular, we draw insights from the Data Envelopment Analysis (DEA) [1] approaches, which are widely used nonparametric methods, particularly developed for comparative performance assessment. These approaches do not require arbitrary prior specification of weights and can also take different kinds of economic impacts into consideration. DEA have been applied to the measurement of environmental efficiency or eco-efficiency in numerous studies (see for example [2], [3] and [4]. However, these studies focus exclusively on the production phase, assessing eco-efficiency at the level of firms and production units. Indeed, eco-efficiency of the final products (i.e., outputs of the production process) has been neglected in this literature so far. On the other hand, a number of authors have used DEA for the evaluation of consumer durables. These studies tend to assess product characteristics from an engineering or marketing perspective, paying little, if any, attention to environmental sustainability[5].[6].[7] and [8]. In these studies, products are usually regarded as production units that use some inputs (costs) to produce some outputs (services), while their burden on the ecosystem has been left aside. In this paper, we consider a general method for analyzing eco-efficiency of consumer durables during their use phase. While we acknowledge the importance of the production and disposal phases on environmental sustainability, we believe these phases are best addressed in separate eco-efficiency assessments at the firm or plant level, Therefore, we hereafter limit attention on eco-efficiency in the use phase.

DATA ENVELOPMENT ANALYSIS (DEA)

Charnes et al. [1] introduced DEA in 1978. DEA has since been successfully employed for assessing the relative performances of a set of firms, usually called the Decision Making Units (DMUs), which use a variety of common inputs to produce a variety of common outputs. Assume that there are N DMUs producing s outputs using m inputs. Let the eth DMU produces outputs y_{re}, r =1,2,…..s using x_{ie}, i=1,2,…..m as inputs. The resulting output–input structure of DMUs is shown below. The objective of the DEA is to identify the DMU that produces the largest values of outputs by consuming the least amounts of inputs. This DMU (or DMUs, if there is a tie) is considered to have an efficiency score equal to one.

The efficiencies of the remaining relatively less efficient DMUs are obtained relative to the efficient DMUs, and are assigned efficiency scores between zero and one. The efficiency scores are computed using mathematical programming.

As DEA is now a widely recognized technique, it is not described in this paper. Interested readers are referred to [9],[10],[11], and [12]. There are different basic DEA models (see for example [9]).

One model for calculating the efficiency of a DMU is shown below:

$$\text{Max } E_e = \frac{\sum_{r=1}^{s} u_{re} y_{re}}{\sum_{i=1}^{m} v_{ie} x_{ie}}$$

Subject to:

$$\frac{\sum_{r=1}^{s} u_{re} y_{rj}}{\sum_{i=1}^{m} v_{ie} x_{ij}} \leq 1, \quad j=1,2,3\ldots\ldots n$$

$$u_{re}, v_{ie} \geq 0$$
$$r=1,2,3\ldots\ldots s, \quad i=1,2,3\ldots\ldots m.$$

In the model above, the index set i, r, and j refer to specific set of inputs, outputs, and the DMUs, respectively. The variables u_{re} and v_{ie} are multipliers to be determined by the above mathematical program, and the second subscript e indicates the eth. DMU, and E_e is the efficiency of the eth DMU. The above fractional program can be easily solved by setting the denominator of the ratio equal to unity. Thus, the problem reduces to the following *output maximization* linear programming problem.

$$\text{Max } E_e = \sum_{r=1}^{s} u_{re} y_{re}$$

Subject to:

$$\sum_{r=1}^{s} u_{re} y_{rj} - \sum_{i=1}^{m} v_{ie} x_{ij} \leq 0, \quad j=1,2,3\ldots\ldots n$$

$$\sum_{i=1}^{m} v_{ie} x_{ie} = 1$$

$$u_{re}, v_{ie} \geq 0$$
$$r=1,2,3\ldots\ldots s, \quad i=1,2,3\ldots\ldots m$$

This is the basic DEA model; the solution divides the DMUs into two groups, efficient and inefficient. In practice there is often a need to fully rank the DMUs, identifying the best-performing overall DMU among many efficient ones. This method is useful when decision-makers cannot reach agreement on assigning weights to the criteria (inputs and outputs) of DMUs. Many authors have

proposed various ranking methods in DEA; the most popular ones are: (1) The cross-efficiency model [13],[14], (2) the super efficiency model [15], (3) the benchmarking [16], and the virtual method [17]. Each of these ranking methods is subsequently explained in brief. The reader interested in further detail is referred to [13], [15], [16], and [17].

THE MODEL

This section implements the proposed ranking methods all based on DEA. As we have mentioned the strength of this DEA model is that the eventual choice is made without the subjective determination of criteria weights by decision-makers. We will not only select the right car with the best carbon footprint for the cheapest price (Manufacturer suggested Retail Price [MSRP]), but take into consideration other features such as space of the car, average miles per tank, horsepower, acceleration and average fuel costs per year as shown in table 1.

Table 1. Alternative Fuel Vehicles and Their Characteristics

Raw Data			Inputs		Outputs				
Brand	Model	Energy Source	MSRP ($)	Annual Fuel Cost ($)	Annual Carbon Footprint (tons/yr)	Miles Per Tank/ Charge	BHP	0-60 Time (s)	Passenger Volume (ft³)
1. Audi	A3 TDI	Diesel	$ 29,950	$ 1,261.00	6.2	614.46	140	8.9	89.0
2. Volkswagen	Jetta TDI	Diesel	$ 22,830	$ 1,261.00	6.2	609.00	140	8.5	91.0
3. Ford	Fusion	E85 (Ethanol)	$ 23,175	$ 2,271.00	7.5	472.50	240	7.0	100.3
4. Mercury	Milan	E85 (Ethanol)	$ 26,630	$ 2,271.00	7.4	472.50	240	7.0	100.3
5. Nissan	Leaf	Electric	$ 25,000	$ 680.00	1.6	100.00	110	9.5	90.0
6. Tesla	Model S	Electric	$ 49,900	$ 680.00	1.6	160.00	300	5.6	95.0
7. Mini	Cooper	Gasoline	$ 19,500	$ 1,366.00	5.7	488.00	118	9.7	80.0
8. Toyota	Yaris	Gasoline	$ 14,165	$ 1,308.00	5.9	399.60	106	9.8	84.6
9. Chevrolet	Volt	Hybrid	$ 40,000	$ 1,021.82	4.4	340.00	150	8.8	90.0
10. Honda	Civic	Hybrid	$ 23,800	$ 964.00	4.4	465.00	110	13.5	90.9
11. Honda	Insight	Hybrid	$ 19,800	$ 988.00	4.5	435.00	98	10.6	85.0
12. Toyota	Prius	Hybrid	$ 22,800	$ 810.00	3.7	550.00	98	9.8	93.7
13. Honda	FCX Clarity	Hydrogen	$ 21,300	$ 604.00	2.4	270.00	134	10	100.8
14. Mercedes Benz	F-Cell B-Class	Hydrogen	$ 45,000	$ 630.00	2.8	258.00	136	10.9	93.0
15. Virtual	n/a	n/a	$ 13,000	$ 500.00	1.4	700.00	350.00	5.00	105.00

As an example of usage of our methodologies, we considered fourteen alternative fuel vehicles. These fourteen vehicles (Table 1) represent six different types of vehicle energy systems – diesel fuel, electric, E85 (ethanol) flex-fuel, gasoline fuel, hybrid, and hydrogen fuel cell. Six of the models above were listed in trade articles as the most fuel-efficient cars of 2009/2010. Cars marked with one asterisk were listed in Edmunds.com as one of the "Top 10 Fuel Efficient Cars of 2009". Cars marked with two asterisks were listed in TheDailyGreen.com as one of "The Most Fuel-Efficient 2010 Cars and SUVs". Cars marked with three asterisks are concept vehicles whose production is planned for 2011 (See Table 2).

Table 2 : Sources of efficiency rating of Alternative Fuel Vehicles

1. Diesel - 2010 Audi A3 TDI [10]**
2. Diesel – 2010 Volkswagen Jetta TDI [5]*
3. E85 (Ethanol) – 2010 Ford Fusion
4. E85 (Ethanol) – 2010 Mercury Milan
5. Electric – 2011 Nissan Leaf***
6. Electric – 2011 Tesla Model S***
7. Gasoline – 2010 Mini Cooper [15]**
8. Gasoline – 2010 Toyota Yaris [8]*
9. Hybrid – 2011 Chevrolet Volt***
10. Hybrid – 2010 Honda Civic [3]*
11. Hybrid – 2010 Honda Insight [2]*
12. Hybrid – 2010 Toyota Prius [1]*
13. Hydrogen – 2011 Honda FCX Clarity***
14. Hydrogen – 2011 Mercedes Benz F-Cell B-Class***

Using Data envelopment Analysis rating method we came up with the following table:

Table3: Results of Comparing Rankings of Alternative Fuel Vehicle Using Various Approaches

Vehicle	Rank
1. Diesel - 2010 Audi A3 TDI	13
2. Diesel – 2010 Volkswagen Jetta TDI	3
3. E85 (Ethanol) – 2010 Ford Fusion	10
4. E85 (Ethanol) – 2010 Mercury Milan	12
5. Electric – 2011 Nissan Leaf	7
6. Electric – 2011 Tesla Model S	5
7. Gasoline – 2010 Mini Cooper	7
8. Gasoline – 2010 Toyota Yaris	3
9. Hybrid – 2011 Chevrolet Volt	14
10. Hybrid – 2010 Honda Civic	7
11. Hybrid – 2010 Honda Insight	6
12. Hybrid – 2010 Toyota Prius	2
13. Hydrogen – 2011 Honda FCX Clarity	1
14. Hydrogen – 2011 Mercedes Benz F-Cell B-Class	10

The results show that the best alternative car is Hydrogen – 2011 Honda FCX Clarity with Hybrid – 2010

Toyota Prius close second.

SUMMARY AND CONCLUSION

This article has presented four analytic techniques for ranking and selecting alternative Fuel vehicle. The models are all within the DEA framework and minimize the subjective nature of the decision making process. Today, because of data availability, using decision tools that are data driven are becoming more important in facilitating decision making process. This study examines a field application of DEA beyond selection of a consumer durable from engineering and marketing perspective, but also considers environmental of alternative fuel vehicles during their use phase.

The proposed models offer several unique advantages. First, the models provide a powerful analytical data based tool for selecting consumer durable goods such as automobiles, where traditional models lack data driven framework. Second, the models reduce the subjectivity of the selection process to a minimum because a major advantage of the DEA based models presented is that they do not require that the relative importance or weights of criteria for selection alternative fuel vehicle be known a *priori.* All four models considered, enable the decision maker to select the most efficient, in their top three choices without specifying criteria weights in advance. The paper both innovates, and through a real example implements, DEA-based models, offering a powerful tool for decision-making.

REFERENCES

References are available upon request from authors

Proposal for Development of a Sustainable Smart Grid Laboratory

Kourosh Rahnamai, Julie Drzymalski

Western New England College
Springfield, MA 01119, USA

krahnama@wnec.edu

In this paper, we discuss how to educate our students in a classroom environment to develop not only an in-depth knowledge but also practical experience in achieving sustainability in different environments. This study explains the development of a Sustainable Smart Grid Laboratory. This innovative smart grid laboratory will consist of ten workstations that will model the functionality of the Smart Grid in hardware and software. The interconnected workstations will emulate a small community consisting of residential, commercial and industrial units. All equipment selected for the laboratory is state-of-the-art to address the technological requirements specific to the needs of the Smart Grid.

The courses developed in this proposal addresses topics such as: Smart and efficient energy use, smart production lines, wired/wireless sensors, PMUs, grid security, cybersecurity, and privacy protection for the consumer's information.

THE EFFECT OF CONSUMER ENVIRONMENTAL RESPONSIBILITY ON THE RANKING OF SUSTAINABLE BUILDING ATTRIBUTES

Sharmin Attaran, PhD
Bryant University
1150 Douglas Pike
Smithfield, RI 02917
sattaran@bryant.edu
(401) 232-6964

Bilge Gokhan Celik, PhD, LEED AP
Roger Williams University
One Old Ferry Rd
Bristol, RI 02809
bgcelik@rwu.edu
(401) 254-3648

ABSTRACT

There is a need for the proliferation of sustainable buildings and for enhanced strategies to attract customers to such an initiative. The authors present a methodology and the preliminary findings of a study conducted on the issue of marketing sustainable buildings. The results of a survey to 300 university students compares environmental sustainability scores with a ranking of sustainable building attributes in order to determine best practices for the marketing of sustainable buildings.

Key Words: LEED, Sustainable Buildings, Customer Value

INTRODUCTION

Sustainable development efforts within many different industries have been developing faster than ever. One of these is the building industry in which owners, designers, and builders as well as public officials are continuously demanding, performing, and developing new and higher standards to achieve sustainability in the built environment.

During this process in the US, the Leadership in Energy and Environmental Design (LEED®) rating system was developed by the United States Green Building Council (USGBC) in order to rate buildings on their level of sustainability [2]. LEED New Construction (NC) 2009 aims to achieve this rating process by assessing building design and the construction process under five main categories (Table 1). If a project team seeks a stamp of sustainability for their proposed building they can apply for a LEED certification. The USGBC offers various levels of LEED certification based on the number of points the building can collect under the categories given in Table 1.

Table 1. LEED-NC 2009: Categories [6]

Category
Energy and Atmosphere (EA)
Sustainable Sites (SS)
Indoor Environmental Quality (IEQ)
Materials and Resources (MR)
Water Efficiency (WE)

According to the LEED 2009 certification system, Energy and Atmosphere is allocated the highest number of points followed by the Sustainable Sites, Indoor Environmental Quality, and Materials and Resources categories. The category of Water Efficiency, with only 10% of the total points available, is ranked as the least influential within the certification process of a building. LEED determines these values of each category from a mostly environmental perspective [2].

One of the challenges of certified green buildings is communicating the high value of these buildings to the building owners and occupants. Developing marketing strategies for green buildings can be a step towards communicating the benefits of sustainable buildings and providing the opportunity for the occupants' preferences to be taken into consideration [2].

This study focuses on the five categories of the LEED rating system and explores how willing building occupants are to pay extra for such initiatives, based on their level of environmental responsibility. This will allow the authors (by a multi-phased research plan) to develop a set of marketing strategies for sustainable buildings. In order to achieve this, this study is utilizing campus environments as its sample, as more and more academic institutions are finding the need to become more sustainable. Therefore, the authors selected university students as the initial target group to represent the scheme of building occupants. Upcoming sections discuss the conceptualization and the methodology of this research while presenting and discussing the results of the authors' first attempt to analyze the subjects' responses toward sustainable buildings.

CONCEPTUALIZATION

Environmental and ethical benefits aside, sustainable buildings also have proven long-term monetary benefits. These benefits, from a financial perspective, may include increases in worker productivity and decreases in health care and litigation related costs. Celik [1] states that it is not only the building owners and managers, but also the builders and designers, who can benefit from sustainable buildings such as LEED certified ones due to the minimization of risks such as negative health effects, legal liability, and the number of remediation cases.

Sustainability concerns are growing rapidly in the United States and Europe as more consumers are becoming more environmentally responsible based on attitudes and behavior [4]. A growing concern for the environment has increased the number of consumers joining environmental groups and contributing funds to support environmental causes. Such increased interest in environmental behavior creates viable opportunities for the sustainable building industry. If companies position themselves as being environmentally responsible, such as operating sustainable buildings, there can be positive effects on key stakeholder groups, including customers. Since environmental responsibility is becoming increasingly important for society, consumers value the way organizations manage and operate their production processes and supporting activities [3].

This study takes into account that consumers' environmental responsibility may have effects on environmental behavior and intentions to pay for LEED initiatives. This research goes beyond merely measuring general environmental consumer values, and instead, asks respondents questions regarding their level of environmental responsibility on several categories, to evaluate the specific LEED categories by determining whether they would be willing to pay more in tuition for each characteristic, and to explain their rankings. Such questions can help uncover specific reasons why consumers value one sustainable characteristic over another and how such rankings may relate to their level of environmental responsibility. Using consumers' built environment also establishes relevance to respondents in order to increase their involvement with the topic.

METHODOLOGY

The methodology used for this research project was both quantitative and qualitative. A survey was given to 296 undergraduate business students from a higher education institution. First, the authors wanted to determine how respondents scored on several dimensions of consumer environmental responsibility using the Environmentally Responsible Consumers: Ecoscale [4]. This scale measures (a) consumer knowledge awareness, (b) consumer desire and willingness to act, (c) consumer ability to act, (d) consumer opinions and attitudes concerning the environment, and (e) consumer behavior toward the environment. All items were scored on a 5-point scale ranging from *strongly disagree* to *strongly agree*, or ranging from *never* to *always*. Item scores were summed within each dimension to form dimension indices, and all 31-item scores were summed to form one overall ECOSCALE composite score.

Next, subjects were provided with brief information and scope of five main LEED categories. Respondents were asked to rank, on a scale of 1 to 5 (one being the most important and 5 being the least important), each LEED category based on their willingness to pay an extra 1% in tuition if their university were to develop such initiatives in their campus buildings. Once respondents ranked the LEED categories, they were asked to describe what factors were considered when ranking the categories in an open-ended section.

The raw results for the distribution of respondents based on the ECOSCALE composite score is given in Figure 1. Figure 2 illustrates the ranking results as extracted from the same survey. Authors are in the process of identifying critical correlations between the two scales.
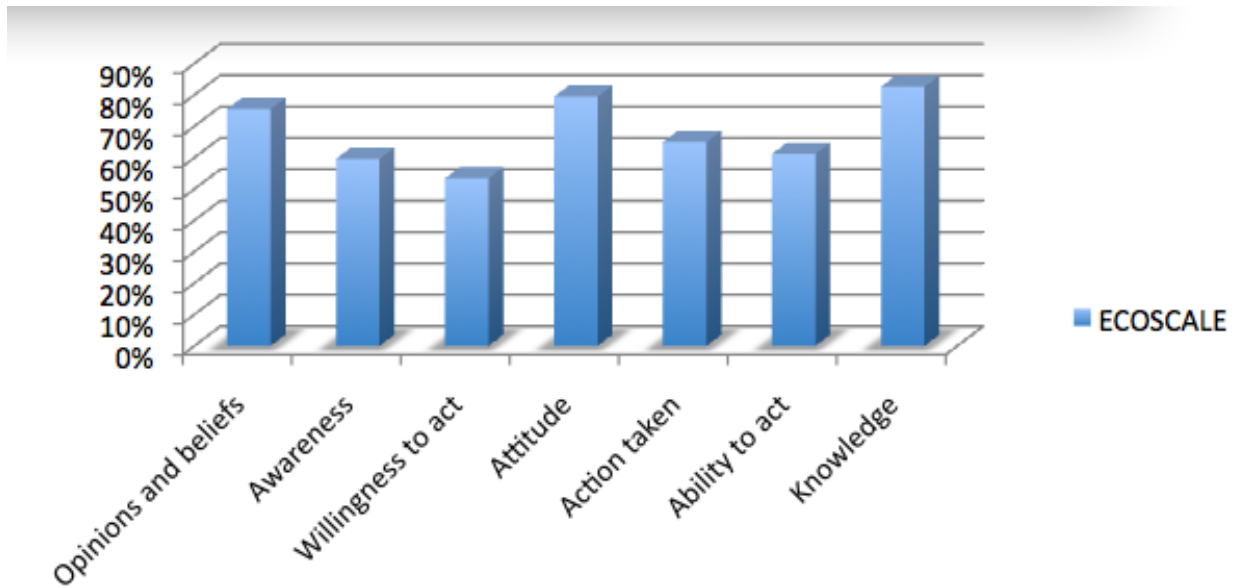
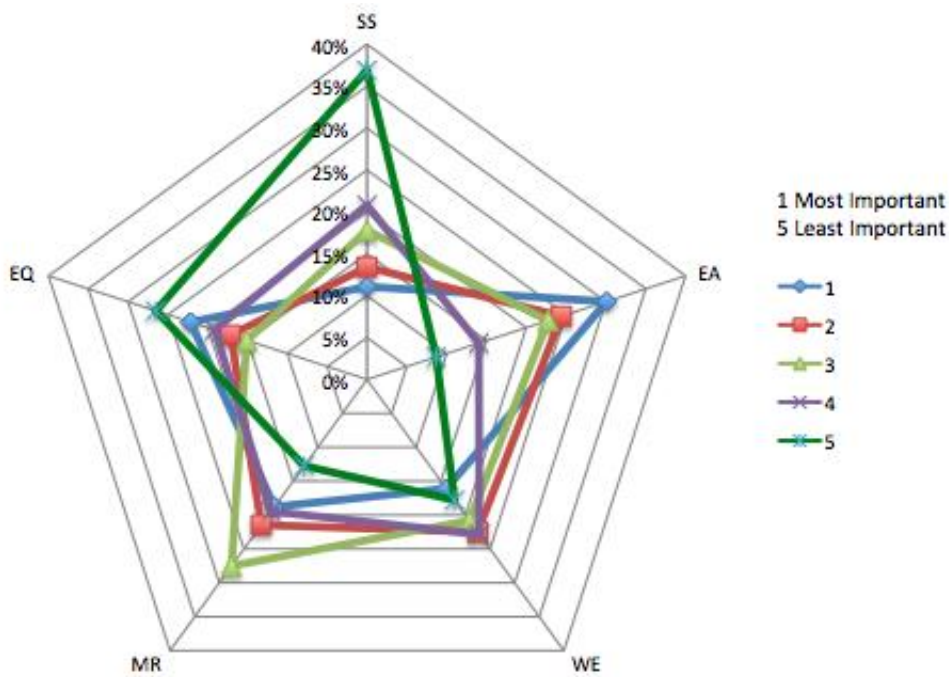**Figure 1.** Respondent results on environmentally responsible behavior.

Figure 2. Distribution of respondents' choices based on rank of each category.

CONCLUSION

Determining the ranking of each LEED certification system categories enables the analysis of occupant value given to sustainable building attributes. In addition, understanding the effects of environmental behavior on the occupant preferences is one of the goals of the proposed study

presented in this paper. Administering extensive surveys has allowed authors to collect valuable data to execute some or all of these analyses. Authors, in this paper, presented one of the initial phases and a preliminary methodology of a multi-phased study. Some of the initial results and the survey parameters are presented while the authors are in the process of completing a more detailed statistical research.

This approach acknowledges that determining the occupant value of sustainable practices will allow a greater implementation of sustainable buildings which is a better way of designing and constructing a sustainable environment. Such information can provide sustainable builders and designers the opportunity to expand their marketing efforts.

REFERENCES

[1] Celik, B. G. (2006). Decision Model to Optimize Indoor Air Quality In Commercial Buildings *in Florida.* Gainesville, FL: University of Florida.

[2] Celik, B. G. and Attaran, S. (2011). Customer-Centric View of Sustainable Buildings: The Case of LEED ®. *Proceedings of the Associated Schools of Construction, USA, edited by Tulio Salbaran.*

[3] Magretta, Joan (2003). *What Management Is: How It Works and Why It's Everyone's Business*. London: Profile Books.

[4] Stone, George, James H. Barnes, and Cameron Montgomery (1995). ECOSCALE: A Scale for the Measurement of Environmentally Responsible Consumers. *Psychology and Marketing,* 12: 595-612.

[5] US EPA. (n.d.). *US Environmental Protection Agency*. Retrieved June 5, 2010, from Tool for the Reduction and Assessment of Chemical and Other Environmental Impacts (TRACI): http://www.epa.gov/nrmrl/std/sab/traci/

[6] USGBC. (n.d.) *United States Green Building Council*. Retrieved February 14, 2010, from USGBC: U.S. Green Building Council: http://www.usgbc.org

AIRPORTS AND SUSTAINABILITY

Abstract : S.P. Roxane Ouedraogo

Airports and Sustainable Development? Here are two terms that seem opposed when one knows the impact of airport operations on society and environment. In particular, noise and emissions from aviation make it difficult for affixing the label "green" to airports. However numerous complaints from surrounding communities, prompted airports, « often forced by governments, to apply different types of noise management measures that range from noise abatement procedures to limits on the total noise allowed» (Morrell & Lu 2006). But environmentally much remains to be done. Indeed, « compared to the introduction of noise management measures, there are fewer airports applying engine emissions mitigation measures (Morrell & Lu 2006) ». Airports are facing constraints thus forcing them to be socially responsible, but environmentally, some obstacles are still to overcome. Especially since « the environmental impacts of aircraft engine emissions include both aircraft landing and take-off and 30-minute cruise». (Morrell & Lu 2006) These environmental problems are born to the cornerstones of airport operations; it is easy to grasp the magnitude of the difficulties which, airports must overcome. How the airports that claim to be socially and environmentally responsible, can overcome such challenges?

The main objective here is first of all, to learn about the tools used and actions implemented by these airports. Then to identify the strengths and weaknesses of those tools and actions. Test their applicability to all airports, before suggesting ways of improvement. A second idea is to find out the different operations at airports that produce garbage and how the garbage is eliminated by airports authorities. For instance, one should know the average daily weight of the garbage produced by the food consumption aboard the hundreds of flights per day and how these garbage are destroyed/eliminated?

Moreover, even if it is a must for airports to implement certain measures to mitigate the impacts of their activities on the community, this does not mean that they are effective. Therefore this will entail thirdly, considering these measures and testing their effectiveness by conducting surveys at communities and councils of cities that host these airports. Recommendations based on the findings of these surveys will then be made to airports in the hope that they go beyond the constraints of community and government if not yet done. The ultimate goal of these approaches is to answer the following question: How can airports contribute to sustainable development and human wellbeing?

Sustainable Alternative Sources of Energy and Their Application: Education of Tomorrow's Managers

Said Dini, Professor

Mechanical Engineering Department
Western New England College
Springfield, MA 01119

Richard B. Mindek, Jr., Associate Professor and Chair

Mechanical Engineering Department
Western New England College
Springfield, MA 01119

ABSTRACT

Reduction in utilization of energy along the activity cost chain of companies, including supply, manufacturing, distribution and sales, plays a vital role in their cost efficiency and resource conservation. Alternative sources of energy, such as wind and solar, have been the center of attention for several decades. With petroleum prices escalating to their highest levels across the world, wind turbines and solar panel arrays have again become symbols of a global growing interest in alternative sources of energy. This study describes a unique program that has been developed at the Mechanical Engineering Department of our university. This program provides students with a hands-on understanding of utilization of indoor and outdoor alternative sources of energy in a laboratory environment. The laboratory facility includes a fully operational geothermal system, which is used to heat and cool the engineering labs. It also includes an array of six PV solar panels and a full scale horizontal axis wind turbine. The scale of these devices allows for useful power and heat to be provided to the engineering building. This facility is also fully instrumented for the collection of key performance data and allows for moderate scale demonstration of efficiency of alternative energy sources to students. Through this lab, students not only understand the operations and capabilities of energy generated by wind and solar devices, but it also allows them to analyze the cost and the extent of efficiency and sustainability of such sources of energy. This form of hands-on education can have wide ranging functional incentives for businesses large and small in terms of energy consumption, in addition to impacting the design and manufacture of different products such as home appliances and automobiles. This paper describes the development, operations and capability of the energy laboratory to educate and train future managers/owners of companies in addition to how we can enhance our students' understanding of sustainable alternative sources of energy and their application to more cost efficient operations and production in an organization.