JOURNAL OF MANAGEMENT AND ENGINEERING INTEGRATION

Scope: The Journal of Management and Engineering Integration (JMEI) is a double-blind refereed journal dedicated to exploring the nexus of management and engineering issues of the day. JMEI publishes two issues per year, one in Summer and another in Winter. The Journal's scope is to provide a forum where engineering and management professionals can share and exchange their ideas for the collaboration and integration of Management and Engineering research and publications. The journal will aim on targeting publications and research that emphasizes the integrative nature of business, management, computers and engineering within a global context.

TABLE OF CONTENTS

The Journal of Management and Engineering Integration
Vol. 3, No. 1, Summer, 2010

Importance of Timeliness and Accuracy on Customer Attitude in a Pharmacy Chain Context

Jyoti Kulkarni, B. Andrew Cudmore and Anthony T. Fischetti
Florida Institute of Technology
jdinkar@yahoo.com, acudmore@fit.edu, Anthony_Fischetti0226@yahoo.com

Abstract

A growing number of customers have expressed dissatisfaction with pharmacy customer service. The present study examines consumer attitudes of the customer service at a pharmacy. The results indicate that consumer's desire accurate and timely information regarding the progress and potential delay of their prescription. The lack of accurate and updated information was found to be a major reason for customer dissatisfaction. Managerial implications, such as strategies to minimize customer dissatisfaction are discussed.

1. Introduction

Healthcare is sometimes referred to as a recession-proof industry, as human health is independent of economic condition (*Zpryme News,* 2008). Individuals may reduce expenditures on items, such as food, but healthcare – specifically medications are a requirement for many making them perfectly inelastic. Thus, in spite of the increase in co-pays and changes in insurance plans consumers must continue to buy prescriptions. This is evidenced by growth in the retail drug industry from 2007 to 2008 resulting in total retail prescription drug sales well above $250 billion for the industry. Pharmacy transactions account for more than 65% of net drugstore sales, thereby making it essential for store profit. In fact, eighty three percent of the market share is controlled by retailers having 200 plus pharmacies (List, 2008). Such giants as Walgreens, CVS and Rite Aid pharmacies have experienced multiplicative gains, due to successful combination of convenience and affordability. In fact, Walgreens is the largest drugstore chain in the United States with sales reaching $53.76 billion and approximately $35 billion generated via pharmaceutical sales (*Walgreen Co.*, 2007). Thus, our focus will be on pharmacy chains and their prescription sales.

Two important issues pertaining to consumer attitudes about pharmacy prescription sales are discussed in this paper in the following manner. First, we provide an overview of the impact of delayed prescription-filling on customer attitudes. Second, we use queuing theory and the theory of attribution to explain the positive effect of timeliness and accuracy of delayed information on customer attitudes. From this theoretical framework a methods section is devised, explaining the research design, sampling procedures, and manipulation checks used. The study concludes with a discussion of the results, limitations of the current findings, and implications for future research.

1.1 Customer relations in the pharmacy

To achieve customer loyalty and trust a business needs to establish, build and maintain positive consumer relationships (Keiningham, Aksoy, Cooil, & Andreassen, 2008). While the primary function of a pharmacy is to dispense prescription medication, customer relations are important in establishing repeat sales (Buckley, 2000). In particular, it is important to deliver the prescription medication within the quoted time frame, and inform the customers of potential problems in a timely and accurate manner. It is through building strong customer relations that a pharmacy will continue to thrive.

Previous findings suggest that low pharmaceutical prices alone do not retain loyal customers. Rather, customers who are loyal to a particular pharmacy seek quality service, and long term, positive relationships with pharmacy employees. In fact, it was found that more than 24% of pharmacy customers engage in conversation with pharmacists. This finding demonstrates the importance of engaging in positive interactions with consumers. One way in which a positive kinship can be achieved is by offering expected services in an appropriate amount of time with accurate information, especially where a customer may wait (Desjardins, 2006).

1.2 Queuing theory

Substantial research exists on the negative effects queuing can have on customer satisfaction (Hui & Tse, 1996; Katz, Larson, Blaire, & Larson, 1991; Taylor, 1994). However, no study has focused on the impact of timeliness and accuracy on customer attitude in the context of pharmacy chains. Customer satisfaction is no longer based solely on the 'best buy,' rather it is influenced by a myriad of factors, including timeliness and accuracy of information.

Queuing information relates to a consumer's position in a line of people waiting their turn (e.g. queue) (Hui & Tse, 1996). Theory of queue psychology was developed in 1985 by David Maister, who investigated the psychology of waiting lines. Findings from Maister's (1985) study resulted in the proposal of eight principles organizations could use to influence customer satisfaction with waiting, including:

(1) Unoccupied time feels longer than occupied time

(2) Preprocess waits feel longer than in-process waits

(3) Anxiety makes waits seem longer

(4) Uncertain waits are longer than known, finite waits

(5) Unexplained waits are longer than explained waits

(6) Unfair waits are longer than equitable waits

(7) The more valuable the service, the longer people will wait

(8) Solo waits feels longer than group waits (p.115-122).

Tactics of queuing psychology have long been employed by Disney World and Disney Land theme parks to make waiting feel shorter (Leerhsen, 1989). The use of overestimated wait times at the beginning of each ride or attraction aids in improving customer satisfaction by paralleling principle 4 of Maister's theory. Ensuring that people wait in groups and that wait lines are continuously moving results in individual's considered being 'in-process," thereby utilizing principle 2, and 8 of Maister's theory of queuing. Applying the principles of queuing psychology, as evidenced at Disney theme parks allows for the negative effects of waiting to be circumvented, thereby improving overall customer satisfaction.

Use of queuing information can effectively reduce consumer dissatisfaction with waiting (Larson, 1987; Maister, 1985). Conversely, lack of queuing information can result in consumers blaming an establishment when experiencing long waits (Bitner, 1990). The effects of waiting on service evaluation were illustrated by Katz et al. (1991) via investigating customer perceptions of waiting, and ways to improve satisfaction. While it was found that perception of waiting time negatively correlated with customer satisfaction, distractions were positively related to customer satisfaction. These findings suggest it is not only important to determine how long customers will wait, but it is essential to utilize distracters to increase customer satisfaction. As the above discussion indicates queuing information is widely used to maintain and improve customer satisfaction with waiting.

2. Conceptual development

2.1 Attitudes toward pharmacies

To ensure customer retention pharmacies must provide excellent customer service (Keiningham et al., 2008). However, this is often challenging since consumer attitudes tend to deteriorate as a result of issues pharmacies have when filling prescriptions. A

major problem stems from delays in filling prescriptions which may or may not be the fault of the pharmacy. A substantial amount of research has investigated the effects of waiting on perception of service (Hui & Tse, 1996; Katz et al., 1991; Taylor, 1994).

When a customer approaches the pharmacy counter or drive-thru with a prescription to be filled, satisfactory service is expected. Standardizing service as much as possible with clear policies and procedures can only assist in this service. For example, according to the *Walgreens Employee Handbook* (2008) the pharmacy technician who greets the customer is instructed to ask certain questions, including:(1) "Hello, how may I help you?", (2) "Are you picking up or dropping off prescriptions?", (3) "Could you tell me the date of birth of the patient please?", and (4) "And when would you like to pick up your prescription?"

Of importance to the present study is the last question -the amount of time the customer is willing to wait for the prescription(s) to be filled. During the wait period that ensues Walgreens anticipates customers will walk around the aisles, shopping for other merchandise on display (Heller 2003). It is through this combination of both pharmacy and front-end sales that the industry remains economically sound (Shott, 2007). However, prior studies have demonstrated that waiting for service, as in the aforementioned case of waiting for prescriptions, has a negative effect on customers' perceptions of the service provider (Katz et al., 1991; Taylor, 1994). Thus, we are interested in how pharmacies can mitigate such perceptions by better managing the timeliness of information about delays and the accuracy of reason for the delay.

2.2 Timely information about the delay

Customers often choose larger pharmacies due to convenience (Mullis, 2007). In addition, prescription medication, as well as photo finishing, printer cartridge refilling, cosmetics, and other basic needs, also bring customers to large chains. However, if the customer is in pain, is sick, or uncomfortable he/she requires timely relief from these feelings. Therefore, time is one of the most important factors in service at the pharmacy as evidenced by numerous blogs and complaint reports filed about pharmacies on such websites as www.ripoffreport.com.

Customer relationships at the pharmacy not only impact quality of service but also dictate the levels of pain and annoyance experienced by the customer (Desjardins, 2006). Customer expectations are set high to begin with due to the option of drive thru service and express pay methods, but these methods are successful only if the prescription is ready for pick-up on time. If the customer uses the drive thru window to pick up the prescription after a promised period of time and the prescription is not ready, s(he) may be required to wait in the parking lot until it is ready. This inconvenience often leads to unfavorable attitudes toward the pharmacy.

The rise in the number of prescriptions filled per day has lead to a perception that a pharmacy is analogous to a McDonalds for drugs, as millions are served per day (Ross & Schwartz, 2007). Service in the pharmacy is of utmost importance and cannot afford to be compromised, as healthcare is always in the limelight of litigation. However, according to statistical studies there is a regional shortage of pharmacists across the country. In fact most drug chains rely on untrained pharmacy technicians (Brady & McCoy, 2008). Technicians assist the pharmacist in filling prescriptions, and entering the information printed on the prescription: Name of the patient, address, phone number, and health care provider's name. A prescription may not have all of this information, or it may not be clear. To avoid delays it is often necessary for the pharmacist, who is may be busy to intervene.

To better serve customers who desire brief stays, pharmacies must have all the information required to fill the prescription assembled and readily accessible. In turn, this process helps the pharmacy learn of potential problems in filling the prescriptions and convey them to the customers without delay. Yet, the combination of a scarcity of pharmacists and required compliance with laws, such as HIPPA increases the total time required to fill the

prescriptions (Brady & McCoy, 2008). The more time it takes to provide consumers with information regarding the prescription's delay the more negative the consumer attitude. Conversely, if the information is given in a timely manner the customers often feel confident the pharmacy is providing the right service, resulting in a positive effect on their assessment of the situation.

Various other industries provide wait time guarantees. For example, pizza companies often set a time frame, such as 30 minutes for their deliveries, setting expectations for the customer (Kashyap, 2001; Kumar, Kalwani, & Dada, 1997). Similarly, if the integrity of a pharmacy is compromised due to an inability to deliver a prescription in a promised period of time, customers' attitudes toward the pharmacy are affected negatively (Panizzo, 2008). This leads to the following hypothesis:

H1: The more timely the information about the delay in filling prescriptions, the more positive the customer attitude towards the pharmacy.

2.3 Accurate information about the reasons for the delay

Pharmacies often estimate wait times for the filling of prescriptions. Waiting for service can be considered an opportunity cost since the individual is not able to engage in a more productive activity during that time. Waiting beyond an individual's wait threshold often results in frustration. Chambers and Panagiotis (2006) found that Walgreens pharmacy customers expect their prescription to be filled before the estimated wait time they were given. Therefore, it is reasonable to assume that customer expectations of completing the sale increase upon hearing their name over the store intercom. However, these customers are often paged over the intercom to be informed of further delay, creating a satisfaction gap that could lower the customer's attitude toward the pharmacy (Zemke, 1989).

One common reason for delay is the denial of prescription coverage by a consumer's medical insurance (Medicare.gov, 2007). While there are many reasons for denial of coverage the most often cited example is that the consumer's physician failed to authorize a refill. Although the pharmacy may not be at fault for the delay in filling one's prescription, the theory of attributions suggests that consumers will still attribute causes to situational factors (Heider, 1958). Ultimately, this results in the blame being placed on the pharmacy. However, by informing the customer of the reasons for the delay in an accurate and timely manner, the pharmacy may mitigate negative associations with the pharmacy itself.

Another major reason for delays is that Medicare prescription drug plans are experiencing problems filling prescriptions. For example, pharmacies are unable to confirm that a customer qualifies for extra help since the databases are not always updated (*Medicare.gov*, 2007). In this case the pharmacists' require proof including the Medicaid card and an award letter from Social Security, or the yellow auto-enrollment letter from Medicare. If the customer does not have the required documentation the pharmacy is unable to fill prescription, or may charge more than the expected amount. While this is not the pharmacy's fault the information about the delay should be made available to the customer as soon as possible in an attempt to meet the customer's service expectations. To avoid unrealistic customer expectations pharmacy technicians can voice a caveat about insurance changes to customers immediately upon greeting them.

A final reason for delays is that the prescribed drug may simply be out-of-stock. Although the pharmacy may have placed the replenishment order in time there may be shortages for that particular drug in the market. Immediately informing the customer about such problem helps retain positive attitudes toward the pharmacy. However, the high daily volume of customers in each pharmacy makes identifying problems in each order an arduous task. Therefore, providing the customer with a realistic timeframe on filling their prescription is one way to keep customers from getting upset. Informing the customer, in advance that the pharmacy is overwhelmed and explaining a realistic wait time, albeit longer than expected, helps neutralize customers' attitudes. This type of communication results in the customer

appreciating the honesty of the pharmacy, building a positive image (Young & Albaum, 2003). This leads us to the second hypothesis:

H2: The more accurate the reason for delay in filling a prescription the more positive the customer attitudes toward the pharmacy.

2.4 Timely and accurate information about the delay

While revenues from sales at pharmacies continue to increase customer satisfaction ratings due to poor pharmacy services continue to decrease. Improving customer relationships in the pharmacy, aids in reducing the potential loss of customers to competitors. Timeliness of the information regarding the delay and accuracy of the reason for the delay are predicted to be important influences on customer attitudes. Timely service is expected to be the most important factor in impressing the customer and therefore will have a greater influence on the attitude of the customer (Brill, 1992). And, according to attribution theory the pharmacy will therefore be put at fault. If inaccuracy in the expected service results in further delays and costs, the customer's attitude is negatively affected (Olivia, Oliver & McMillan, 1992). This results in our third hypothesis:

H3: The timing of the delay in filling the prescription leads to more positive customer attitudes toward the pharmacy, only if the reason for delay in filling the prescription is accurate.

3. Methodology

3.1 Design

A 2x2 (early and late timeliness of delay duration information x accurate and inaccurate reason for delay) ANOVA, between-subjects, balanced design was utilized for this research. A survey, consisting of 48 items was used. The survey included a 7-point Likert type scale included semantic differentiation scales with open-ended items. These items offered choices from number one through seven, with one being the most negative and seven being the most positive point (e.g., positive= 7; favorable = 7; strongly agree = 7). Higher numbers indicated a stronger positive response.

Table 1. Demographics

Sub-group	Description	Pop.	%
Gender	Male	48	39.7
	Female	73	60.3
Ethnicity	African American	16	13.2
	American Indian	2	1.7
	Asian	31	25.6
	Caucasian	58	47.9
	Hispanic	7	5.8
	Pacific Islander	1	0.8
	Others	6	5.0
Income range	0 – $ 24,999	21	17.4
	$ 25,000 - $49,999	17	14.0
	$50,000 - $ 74,999	48	39.7
	$ 75,000- $ 99,999	25	20.7
	More than 100,000	10	8.3

Table 2. Sample descriptive statistics

N	121
Range of age	20-85
Average age	41.03
Sampled area	SE, USA
$M_{\text{prescriptions /year}}$	12.44
$M_{\text{drive thru}}$	7.58
M_{wait}	4.86

The sample demographics presented in Table 2 parallel those of the normative population of pharmacy customers. These pharmacy customers can be sub-grouped into two groups- Young adults up to the age of 49 and older adults aged 50 and above. Expectations of each group differ based on the age of the individual (Gold, 2004). Older adults spend more than $100 per month on the prescriptions and therefore represent an important sector of the business. Generally, however, this group prefers to order their prescriptions via the US mail. When older individuals do frequent the store, satisfaction depends mainly on the available parking areas, friendliness of the store employees, and readable

vials. Whereas the younger adults, likely to be busy and preoccupied, are more likely to be satisfied with the timeliness of the service.

In the chain drug store sample, cited above, the average age of the shopper is 57 years, but is drawn from a sub-group with the highest household income level (Gold, 2004). Older customers look for added value that community pharmacies bring to the nation's health- network. Consumers polled by WilsonRx reported visiting their chain drug store pharmacy more frequently than their primary care physician -An average of 3.1 times per month, compared to 3.1 visits per year, respectively (*Pharmacy satisfaction digest survey*, 2007). This indicates that the importance of timeliness and accuracy does not significantly affect their attitudes. However, approximately 60 percent of the participants in this survey were less than 42 years of age and more than 60 percent of the total population fell into the income range of $50,000 – $99,999.

Whether one values timeliness or accuracy is somewhat dependent upon an individual's age. Retired customers have more leisure time than working adults; hence they care about accuracy more so than timeliness. Whereas, working adults, who typically spend time during the workday waiting for their prescriptions, are more concerned about timeliness. Our sample showed a higher proportion of females, consistent with recent targeting strategies by such pharmacies as CVS (Shott, 2007). Mothers, who typically serve as chauffeurs for their families are on the run and have timelines to meet, therefore timeliness is of importance to them. Other factors in addition to age impact whether an individual values accuracy or timeliness. Young, working individuals, considered those aged less than 57 year, typically earning $50-$100K annually take their jobs seriously, and as a result strongly value timeliness. Therefore, this was a representative sample appropriate for examining the importance of the time and accuracy of the service.

3.2 Dependent variable

The attitude variable was measured utilizing three, seven point semantic differential items: "Based on the scenario described above, my attitude about the pharmacy is _____: (Negative/Positive; Unfavorable/Favorable; Bad/Good)."

3.3 Manipulated variables

The survey scenarios were designed to depict the frustration mentioned in the blogs by the pharmacy customers. The survey represented a customer belonging to the younger adult group in the light of pain and hurry to get back home. The attitude of this group is important, as their loyalty will determine the future success of chain drug stores. By 2050, the Census Bureau projects the 65 and older population will grow from 36.3 million in 2004 to 86.7 million, or 21 percent of the U.S. total. This means that from the sustainability point of view, it will be necessary for the retail drug stores to please customers which are now younger and will grow older along with increasing need for medication. For example, an individual aged 35 years today will be 75 in 2050. If the individuals is pleased with the customer service s(he) will still be a loyal customer of the same drug store. Therefore, retail drug stores need to assess the lifetime value of the presently young customer, making efforts to satisfy them for purposed of future loyalty.

In the depicted scene the subject is asked to imagine being at a pharmacy for the purpose of getting a prescription filled. The scenario begins with a technician accepting the prescription as usual; promising to fill the prescription in 30 minutes. The scenario describes the condition of the subject being just out of surgery and in need of pain medication. The subject therefore chooses to wait for the prescription while it is filled. To make the story realistic and to make the time factor important, there is the spouse waiting at home to get to work; described as caring for the child. The timing, in combination with information about the delay and the reason for the delay, was manipulated in four different combinations (scenarios). Finally, the survey narrates other factors that may affect one's attitude, including:

- Attitude of the pharmacist
- Politeness of the technician

- An apology from the pharmacist about the delay
- Being able return later for the medication
- Being able to ask the pharmacist questions
- Availability of chairs near the pharmacy
- Having the ability to use a cell phone to talk to one's family
- Availability of a massage chair, along with other seating near the pharmacy counter
- Being able to navigate freely, around the air conditioned store and view various merchandise on sale
- Availability of restrooms for use with proper identification

The particular factors were chosen, as they were mentioned in blogs detailing scenarios considered unsatisfactory by many. The presence of a massage chair was chosen for all scenarios to act as a "positive constant" in order to provide a more conservative test of our hypotheses. The time constraint was established by the need to get home within the next hour or so. The timing delays were set in such a way that subjects were free to leave the pharmacy after 60 minutes. These considerations also made the experiment a more conservative test of our hypotheses.

The information about the delay provided 5 minutes after the prescription was accepted and 15 minutes after it was accepted, represents the early/late conditions of timeliness. Note that there was no actual passing of the amount of time indicated in the scenarios for the respondents; rather we depended on their perception of time as they imagined themselves in such a situation. It was assumed that most people have some reference point for what is too long for waiting in certain situations. This manipulation was intended to tap into these time reference points and endeavor to capture at what point is the waiting time too long. A similar utilization of this form of stimuli for perception of time can be seen in Baranishyn, Cudmore and Fletcher (2010).

Providing the accurate reason for the delay and inaccurate reason for the delay was represented by ability/inability to cover for the cost of the medication by the insurance coverage carried by the subject. The 5 minutes timing was set as early to describe a scenario where the case was handled right away and the 15 minutes were considered late since 50 percent of the time promised (30 minutes) had elapsed before the pharmacy paid any attention to the case. The accuracy of the reason was designed to take the blame away from the pharmacy (providing a more conservative test of hypotheses) and put it on the insurance changes, while inaccuracy of the reason reflects the inability of the pharmacy in detecting the problem. Additionally a picture of a person waiting at the pharmacy counter with nobody in the pharmacy in the view was shown to amplify the effect of the scenario on the subjects.

Table 3. Pairing and manipulation of independent variables

	Timeliness E	Timeliness L
A	-Delay inform- 5 minutes -Insurance does not cover medication	-Delay inform- 15 minutes after drop-off -Insurance does not cover medication
I	-Delay-inform 5 minutes after drop-off -Insurance does not cover medication - incorrect Insurance not updated – correct	-Delay inform- 15 minutes after drop-off -Insurance does not cover medication - incorrect Insurance not updated – correct

*E=early; L=late; I=inaccurate; A= Accurate

Manipulation checks were included toward the end of the survey to verify manipulation of the independent variables. Timeliness of the delay information: "In the scenario described above, I believe the information about the delay was…(Early/Late)". Accuracy of the reason for the delay: "In the scenario described above, I believe the reason for the delay was…(Inaccurate/Accurate)".

3.4. Procedure

The survey was administered in the residential area of a southeastern state of USA. The total number of subjects polled was one hundred and twenty one. The administrator was careful to incorporate a diverse age and gender populous when distributing questionnaires to the interested volunteers. The survey was administered in a quasi-random pattern choosing every third person who came across in the general public places like waiting lounges in clinics, parents waiting at schools to pick up their kids and people in the residential sub-divisions in the area. Online users from a regional email also participated via email. These participants were chosen from the list of subscribers by picking every third subscriber. The survey was administered only to the willing subjects in a reasonably simple manner and to be able to accurately measure the attitude towards the pharmacy. The subjects were asked to read and sign a consent form for the paper-survey users while the internet users were asked to click "I agree" radio button to indicate their consent to taking the survey. If the respondent clicked "I disagree" the last page with just a "Thank you" note was displayed. Similarly the age of over 18 was confirmed before displaying the survey. In order to maintain anonymity, the consent forms were removed from the remainder of the survey for the paper-survey users. The first page of the survey included the picture of a person waiting at the pharmacy counter with no help, followed by the scenario and the survey questions. The sequence of the questionnaire was- measuring the attitude, intention, trust, other items influencing the attitude, manipulation checks, demographics and finally some open ended questions.

4. Results

Cronbach's alpha (.991) for the attitude measure was good. Manipulation checks for timing of the delay information ($M_{early} = 2.95$, $SD_{early} = 1.741$; $M_{late} = 5.51$ $SD_{late} = 1.330$) ($F_{(1,119)} = 83.318$, $p<.0001$) and accuracy of the reason for the delay were ($M_{inaccurate} = 2.20$, $SD_{inaccurate} = 1.459$; $M_{accurate} = 5.62$, $SD_{accurate} = 1.290$) ($F_{(1,119)} = 186.429$, $p<0.0001$) were successful (see Table 4).

The primary purpose of this study was to examine the effect of timeliness of the delay information and the accuracy of reason for the delay on the attitudes towards the pharmacy.

Table 4. Descriptive statistics and between subjects results: manipulation checks

Timing	M	SD	N	F	Adj. R^2
Early	2.95	1.741	58		
Late	5.51	1.33	63	83.3	.407
Accuracy					
Inaccurate	2.20	1.459	61		
Accurate	5.62	1.29	60	186.4	.607

The overall model was found to be significant ($F_{(3,117)} = 14.054$, $p < 0.001$) with an adjusted $R^2 = 0.246$. Table 5 and 6A provide information about the analysis of variance, including the descriptive statistics and the between subjects effects. Table 6B provides information about the interaction.

Table 5. Descriptive statistics: attitude toward the pharmacy

Timing	Accuracy	M	SD	N
Late	Inaccurate	2.66	1.382	29
	Accurate	3.66	1.726	34
	Total	3.20	1.644	63
Early	Inaccurate	2.88	.850	32
	Accurate	4.92	1.628	26
	Total	3.79	1.616	58
Total	Inaccurate	2.77	1.130	61
	Accurate	4.21	1.786	60
	Total	3.48	1.651	121

Table 6A. Attitudes toward the pharmacy - between subject effects

Source	Type III SS*	df	MS*	F*	P
Overall	86.7	3	28.9	14.1	.000
T	16.53	1	16.53	8.0	.005
A	69.62	1	69.62	33.9	.000
T*A	8.2	1	8.2	3.99	.048
Error	240.5	117	2.055		
Total	1794.2	121			
Correct	327.1	120			

Total					

*Rounded
**R^2= .265 (Adjusted R^2= .246)

H1: The more timely the information about the delay in filling prescriptions, the more positive the customer attitude towards the pharmacy was supported ($M_{early} = 3.79$, $SD_{early} = 1.616$; $M_{late} = 3.20$, $SD_{late} = 1.644$; $F_{(1,119)} = 8.043$, p= 0.005) (See table 5 and 6A)

H2: The more accurate the reason for delay in filling a prescription the more positive the customer attitudes toward the pharmacy was supported ($M_{inaccurate} = 2.77$, $SD_{inaccurate} = 1.130$; $M_{accurate} = 4.21$, $SD_{accurate} = 1.786$; $F_{(1,119)} = 33.874$, p = 0.0001) (See table 5 and 6A).

H3: The timing of the delay in filling the prescription leads to more positive customer attitudes toward the pharmacy, only if the reason for delay in filling the prescription is accurate was supported ($F_{(1,117)} = 3.988$, p = 0.048. (see tables 6A and 6B)

Table 6B. Interaction means tested

Scenario	Scenario	T-test	P
E/I	E/A	-6.14	P<.05
E/I	L/I	0.76	P>.05
E/I	L/A	-2.31	P<.05
E/A	L/A	2.87	P<.05
E/A	L/I	5.57	P<.05
L/I	L/A	-2.51	P<.05

*E=early; L=late; I=inaccurate; A= Accurate

5. Discussion

All of our hypotheses were supported with an adjusted R-squared (25%) that suggests that timing and accuracy of the delay information are not only significant, but are major influences on customer attitudes toward the pharmacy. It is important to note that in table 5 the only mean that is above 4 is for the combination of early and accurate. This means that any other combination of the independent variables left the customer having a negative attitude toward the pharmacy. Yet, all combinations led to significantly better attitudes than the late and inaccurate condition (see table 6B). Thus, pharmacy management should endeavor to make improvements on at least one (e.g., accuracy), or preferably both timing and accuracy of delay information in order to better manage customer attitudes and inevitably their satisfaction with the pharmacy.

More specifically, timely (early) delay information puts the customer in a position to anticipate the delay and set expectations accordingly. Conversely if the delay information is given late, the customer is expecting to be out of the pharmacy by the end of 30 minutes. But in reality the customer has to wait for twice the amount of time to finally receive the prescription. Also, since the information was provided 15 minutes after accepting the prescription, the customer perceives a loss of control of the situation. In the scenarios described above the locus of control is internal since the customer is free to return for the prescription if he/she does not wish to wait as described by Rotter (1954). But due to the external factors such as pain and time constraints, the locus of control shifts towards an external focus. Once externally focused, the customer will tend to attribute all the events to external factors such as the delay to be the fault of the pharmacy. This perception will affect the attitude towards the pharmacy negatively.

Focusing on the accuracy of the delay information, if the customer is told the accurate reason for the delay he/she will place trust in the service process (VanRaaij & Pruyn, 1998). However, the customer will attribute an invalid reason for the delay to be the fault of the pharmacy, which in turn may affects their trust, which arguably will negatively affect the customer intention of repeat services (Grandison & Sloman, 2001). In our research trust was found to be positively correlated with the customer attitude toward the pharmacy (r=.652; p<.0001), indicating the damage that can be done to the relationship with the customer if timing and accuracy of delay information are not improved.

Additionally the study also revealed other variables that may affect the attitude towards the pharmacy. The following table (see Table 7) shows the importance given by respondents to

each of the factors. For example, the attitude of the pharmacist was the highest ranked factor by the customers which was followed by the politeness of the technicians. Both were assessed as similar in importance to the respondents ($M_{attpharmacist}$ = 6.19, $SD_{attpharmacist}$ = 1.035 and $M_{politeness}$ = 6.36, $SD_{politeness}$ = 0.876). When asked about the importance of the ability to ask questions to the pharmacist and an apology from the pharmacist the respondents rated both as second most important factors. This emphasizes the need of interaction between the pharmacist and the customers (Desjardins, 2006). Less than 50 percent of the respondents thought that being able to use the restrooms and being able to return for the medication was important. Similarly customers rated chairs near the pharmacy to be of medium importance. This may be due to a certain level of anxiousness when waiting which may keep a person from sitting down while waiting (Pan, 2007). Being able to look at the various merchandise on display and the massage chair were rated as least important. This lower rating may due to the fact that there is usually only one massage chair near the pharmacy counter and the customer is in pain, so looking at other merchandise loses its importance ($M_{merchandise}$ = 3.80; $SD_{merchandise}$ = 2.132).

Table 7. Importance assessment and rankings

Other factors	M	SD	% Rank Top 5
Attitude pharmacists	6.19	1.035	86.0
Politeness	6.36	0.876	81.0
An apology	5.84	1.522	76.0
Can ask questions	5.85	1.376	76.0
Can use restrooms	4.97	1.807	41.3
Chairs	4.52	1.728	37.2
Can use cell phone	4.63	1.598	28.1
Massage chair	4.34	1.631	22.3
Can shop	3.80	2.132	22.3

When respondents were asked about when in general they would notice a delay in service, 50 percent replied they would notice in 10-15 minutes. This is reasonably long considering the wait time they had imagined in the scenario was supposed to be 30 minutes. This suggests that only after about 45 minutes would these customers begin to experience negative attitudes toward the pharmacy. In any case, this range of acceptable delay is useful information for pharmacies when experiencing delays in filling prescriptions.

Further, the respondents indicated that they purchased prescriptions 10-12 times per year or approximately 1 prescription per month with the majority choosing to wait (M_{wait} = 7.58) compared to those who would use the drive thru ($M_{drive\ thru}$ = 4.86) indicating once again, the relevance of the context chosen for this experiment. In this study, respondents expressed that they would feel "stuck" in these scenarios were they were forced to wait (M_{stuck} =5.21; SD_{stuck} =1.643). In addition, respondents indicated that their attitudes toward the pharmacy were positively correlated with pharmacists that took the time to explain the reason for the delay. This indicates the important role of the pharmacist in delay situations.

In summation, pharmacies can improve customer service by endeavoring to improve the accuracy of their delay estimations for filling prescriptions and by reducing the overall service time. When this accuracy increases, then the earlier the delays are announced, the better for maintaining positive attitudes toward the pharmacy. Often this accuracy depends on fast access to insurance information. The reasons for delays in accessing this insurance information should be explored further by pharmacies. If the fault lies with the insurance companies, then this needs to be negotiated, as it is hurting the brand equity of the pharmacies. Given the volume of prescriptions for large pharmacies, insurance companies should be willing to listen. If the fault lies with the pharmacies, then perhaps either their computer systems need to be faster, or more user friendly, or more personnel is needed.

Our research has indicated that the personnel are the fulcrum of the situation, where

a calm, polite, positive, and apologetic technician or pharmacist can make all of the difference in compensating for delays in service. Unfortunately, this is harder to maintain than one might think as despite the skill set of overseas labor, a high degree of job security and high pay potential, there is a scarcity of pharmacists. To overcome this bottleneck in the business process, pharmacies should fund scholarships' in educational institutions, help these universities to attract students to this discipline during the admissions process, and advertise more heavily in the job fairs to show students that there is a lucrative market for these skills. As this is a long term problem that requires a long term solution, pharmacies need to start now to increase the pool of pharmacists and to focus on ways to improve customer satisfaction as their volume of transactions continues to escalate to new heights.

5.1 Limitations

A few respondents questioned the administrator regarding the relationship between the pharmacy and the physician's office. There are times in which there is a mistake in a prescription and the pharmacy must deal with the issue, although they may not be at fault. In such cases the delay is still viewed as the incompetence of the pharmacy.

Arguably, sampling from other locations in the country would increase the representativeness of our sample and in turn its external validity. However, we feel justified in having an adequate cross section of the population to substantiate the results as generalizable. In addition, we recognize that this paper studied the attitudes of the customers in a controlled scenario where only two variables were manipulated. In turn, we limit some realism for the sake of control and internal validity of the research design.

Respondents were asked to imagine a scenario. The presence or absence of real, physical pain would be more realistic, resulting in a more significant effect on the attitude of the customers. This would mean we would be more likely to find significant results. However, the fact that significant results were found in our more conservative test (without pain) shows that timing and accuracy of delay information are indeed important variables worthy of examination. Lastly, it is important to note that this experimental design methodology and use of scenarios is widely accepted in consumer behavior studies (Amyx & Bristow, 2001; Choi & Mattila, 2009; Estelami & Bergstein, 2006; Fernandes & Pizzutti, 2007; Grewal, Roggeveen, & Tsiros, 2008; Liu, Huang, & Minghua, 2007; Martin, Ponder, & Lueg, 2009; Mittal, Huppertz, & Khare, 2008; Smith & Cooper-Martin, 1997; Till & Shimp, 1998).

5.2 Future research

Older adults prefer to use the mail order system (Gold, 2004). Their attitudes can be studied to further aid the industry to enhance their popularity and increase market share. Perhaps the use of drive thru service needs to be advertised more to increase its popularity with examination of any problems for the customers using this service could become another topic of study.

A large number of pharmacy customers are diabetic patients and are regular customers of the pharmacy. Once loyal to a particular pharmacy such customers seldom change their where they fill their prescriptions. But, if their pharmacies fail to provide good customer service along with availability of the test strips and other equipment on time they will lose such long term customers and loose many more due to word of mouth (Buckley, 2000). A stratified sampling of this segment of the population may be worthy of future study.

In our study we found indications that the attitude of the pharmacist and the politeness of the technicians were the most important factors rated by participants. As such, these factors should be considered in future research as employee training could easily improve such aspects of the customer experience.

6. References

Amyx, D., & Bristow, D. (2001). An empirical investigation of customer satisfaction with health care services. *Marketing Intelligence & Planning, 19*(6/7), 515-526.

Baranishyn, M., Cudmore, B., & Fletcher, T. (2010). Customer service in the face of flight delays. *Journal of Vacation Marketing*. *16*(3), 201-215.

Bitner. M.J. (1990). Evaluating service encounters: The effects of physical surroundings and employee responses. *Journal of Marketing, 54*, 69-82.

Brady, E., & McCoy, K. (2008, February 14). Drugstore chains rely on pharmacy techs. *USA Today*. Retrieved from http://www.usatoday/community.com

Brill, J.E. (1992). You want it when? *ABA Journal*, *78,* 98.

Buckley, B. (2000). Loyal diabetes customer reflects untapped sales potential. *Drug Store News, 22*(11), 27-31.

Chambers, C., & Kouvelis, P. (2006). Modeling and managing the percentage of satisfied customers in hidden and revealed waiting line systems. *Production & Operations Management*, *15*(1), 103-116.

Choi, S., & Mattila, A. (2009). Perceived fairness of price differences across channels: The moderating role of price frame and norm perceptions. *Journal of Marketing Theory and Practice, 17*(1), 37-48.

Desjardins, D. (2006). Pharmacy customer satisfaction centers on services. *Drug Store News*, *28*(12), 18.

Estelami, H., & Bergstein, H. (2006). The impact of market price volatility on consumer satisfaction with lowest-price refunds. *The Journal of Services Marketing, 20*(3), 169-181.

Fernandes, D., & Pizzutti Dos Santos, C. (2007). Consumer complaining behavior in developing countries: The case of Brazil. *Journal of Consumer Satisfaction Dissatisfaction and Complaining Behavior, 20*, 86-110.

Gold, G. (2004). Age a factor in pharmacy customer satisfaction. *Chain Drug Review, 26*(14), 34.

Grandison, T., & Sloman, M. (2001). *A survey of trust in internet applications [IEEE Communications Surveys and Tutorials]*. Retrieved from http://www.comsoc.org/pubs/surveys/ Imperial College

Grewal, D., Roggeveen, A., & Tsiros, M. (2008). The effect of compensation on repurchase intentions in service recovery. *Journal of Retailing, 84*(4), 424–434.

Heider, F. (1958). The psychology of interpersonal relations. New York, NY: John Wiley & Sons.

Heller, L. (2003). Experimental store has a new merchandising twist. *Drug Store, 25*(4), 52.

Hinchee, D., & Chase, T. (2007). *Delayed: Opportunities for better service*. Unpublished manuscript, Department of Business, Florida Institute of Technology, Melbourne, FL.

Hirsch, I.J., Bilger, R.C., & Deatherage, B.H. (1956). The effects of auditory and visual background on apparent duration. *American Journal of Psychology, 69*(4), 561-574.

Hornik, J. (1984). Subjective versus objective time measures: A note on the perception of time in consumer behavior. *Journal of Consumer Research*, *11*(1), 615-18.

Hui, M.K., & Tse, A.C. (1996). What to tell consumers in waits of different lengths: An integrative model of Service Evaluation. *Journal of Marketing*, *60*(2), 81-90.

Katz, K.L., Larson, B.M., & Larson, R.C. (1991). Prescription for the waiting-in-line blues: Entertain, enlighten and engage. *Sloan Management Review, 32*(2), 44-53.

Keiningham, T. L, Aksoy, L., Cooil, B., & Andreassen, T.W. (2008). Linking customer loyalty to growth. *Sloan Management Review, 49*(4), 51-57.

Larson, R.C. (1987). Perspectives on queues: Social justice and the psychology of queuing. *Operations Research, 35*, 895-905.

Leerhsen, C. (1989, April). How Disney does it. *Newsweek*, *3*, 52.

List, B. (2008). 2008 drug store and HBC chains industry overview. *Chain Store Guide,* 1-6.

Liu, M., Huang, Y., & Minghua, J. (2007). Relations among attractiveness of endorsers, match-up, and purchase intention in sport marketing in China. *The Journal of Consumer Marketing, 24*(6), 358-367.

Maister, D.H. (1985). The psychology of waiting lines. In J.A. Czepiel, M.R. Solomon, & C.F. Surprenant (Eds.), *The Service Encounter: Managing Employee/Customer Interaction in Service Businesses* (p.113-123). Lexington, MA: Lexington Books.

Martin, W., Ponder, N. & Lueg, J. (2009). Price fairness perceptions and customer loyalty in a retail context. *Journal of Business Research, 62*(6), 588-593.

Mittal, V., Huppertz, J., & Khare, A. (2008). Customer complaining: The role of tie strength and information control. *Journal of Retailing*, *84*(2), 195–204.

Mullis K. (2007). A SWOT analysis of Walgreens in the competitive pharmacy marketplace, 1-15.

Olivia, T.A., Oliver, R.L., & McMillan, I.C. (1992). A catastrophe model for developing service satisfaction strategies. *Journal of Marketing*, *56*(3), 83-95.

Pan, M. (2007). *The impacts of the number of people front or rear, and perceived waiting time on the mood states*. Unpublished manuscript, Department of Business Management, Tatung University, Taiwan.

Panizzolo. R. (2008). A methodology to measure the value of services provided to customers in manufacturing firms. *Measuring Business Excellence, 12*(3), 3-15.

Pharmacy Satisfaction Digest Survey. (2007). Retrieved from http://www.pharmacysatisfaction.com/pharmacysatisfactionWeb/resources/pdf/2007/2007-pharmacy-customer-and-medicare.pdf

Shott, C. (2007). *CVS targets women in new ad campaign*. Retrieved from http://www.projo.com/business/content/BZ_CVSADS_10_09_07_DJ7DQPM.25d9b64.html

Smith, C., & Cooper-Martin, E. (1997). Ethics and target marketing: The role of product harm and consumer vulnerability. *Journal of Marketing, 61*(3), 1- 20.

Taylor, S. (1994). Waiting for service: The relationship between delays and evaluation of service. *Journal of Marketing, 58*(2), 56-67.

Till, B., & Shimp, T. (1998). Endorsers in advertising: The case of negative celebrity information. *Journal of Advertising*, *7*(1), 67-83.

VanRaaij, W.F., & Pruyn, H. (1998). Customer control and evaluation of service validity and reliability. *Psychology and Marketing, 15*(8), 811-832.

Walgreens Co. (2007). *Annual Report*. Retrieved from http://news.walgreens.com/article_display.cfm?article_id=1048

Walgreens Co. (2008). *Employee Training Guide*. Retrieved from www.walgreens.com/employee

Young, L., & Albaum, G. (2003). Measurement of trust in salesperson-customer relationships in direct selling. *Journal of Personal Selling & Sales Management*, *23*(3), 253-270.

Zemke, R. (1989). Auditing customer service: Look inside as well as out. *Employment Relations Today*, *16*(3), 197-203.

Zpryme News. (2008, April 4). Is the health and wellness industry 'Recession Proof.' Retrieved from http://zpryme.com/blog/?p=68

The Journal of Management and Engineering Integration
Vol. 3, No. 1, Summer, 2010

A Novel Methodology to Enhance Business Alignment using Quantitative QFD and Business Modeling Tools

Diala Tawfig Gammoh, Dima Nazzal, Ahmad Elshennawy, Sandra Furterer
University of Central Florida
dgammoh@knights.ucf.edu

Abstract

The intensive implementation of Quality function deployment (QFD) in the design of projects with six sigma initiatives revealed some of the problems enterprises can encounter while trying to map, or align, business strategies to customer expectations. This paper discusses the use of QFD in strategic planning. A business alignment clock and a new framework are introduced that integrate enterprise business architecture (EBA) with QFD. Relationships inside the QFD matrix are quantitatively defined using various multivariate data analysis techniques. The results of these relationships are exploited to generate competitive priorities and verify the consistency with enterprise strategic plans.

1. Introduction

Business enterprises have recognized quality as an important metric to maintain their strategic objectives in a randomly changing economy when there is no stable or predictable business [1,2]. One important quality design and measuring tool is the QFD, also known as the house of quality (HoQ), which acts as a structured approach that helps companies understand and prioritize customers' requirements, identify their relationships with the technical specifications and then evaluate the company's performance in comparison to its competitors [3, 4].

At this point, business and market analysis can be conducted to assess the actual performance of producing a product or a process based on customer needs and desires known as the voice of the customer (VOC); thus, HoQ has become a vital documentation tool implemented in many enterprises to identify the company's selling points that most of the companies are curious about [4].

Unfortunately, in most cases, HoQ relationships are subjectively defined. This causes an uncertainty in the design of a new product or process to occur and subsequently affecting the alignment between business strategies and customer expectations [5, 6].

EBA and Unified Modeling Language (UML) provide the foundational architecture that links corporate strategy, process initiatives and software development so that individual projects can build capabilities in addition to fulfilling immediate needs. This paper introduces a novel framework to map a holistic business strategy overview that integrates QFD, multivariate data analysis, EBA, and UML.

2. Methodology

Business strategies are the generalized mechanisms used in achieving the objectives of a business enterprise. For instance, a business desires to guarantee that it can meet customer expectations throughout all phases of the business process and stay within the enterprise resources. Since customer requirements can change rapidly, enterprises want to be cognizant

of all changes that might affect their strategic goals and processes.

Ideally, enterprises seek to attain a stable flow of processes that will exceed customer satisfaction yet still meet enterprise strategic goals and objectives. To do so requires constructing a foundation (architecture) that creates a holistic overview of the enterprise and facilitates tracking all inputs and outputs associated with any changes [7, 8].

2.1. Proposed business alignment clock

The business alignment clock in Figure 1 illustrates a novel representation of the changes that occur in business strategy, capabilities and processes as well as in customer requirements. The speeds of the three arms of the clock are the main drivers in mapping the criteria between an enterprise's real-world clock and a proposed quality clock; the hour arm moves slowly, minute arm moves faster, and seconds arm moves the fastest. Figure 1 shows how the business strategies arm maps to the hours' arm, and the customer requirements arm maps to the seconds' arm. To cope with the change in customer expectations, the business capabilities and processes arm maps to the minutes' arm which move faster than the strategies but slower than the customer requirements.

Figure 1A and 1B illustrate two states of the enterprise business alignment during the evolvement of its business clock. To ensure quality in enterprises; business capabilities, processes and customer requirements must be aligned to the business strategy. This alignment should be reviewed whenever the business enterprise receives new data on customer requirements.

Figure 1 (a), represents the ideal case where 100 % alignment exists between the three arms of the business clock. In this case, customer requirements and business processes align, or "map" to the business strategy, indicating that the business processes are capable of meeting the customer requirements as well as matching the enterprise's strategic goals. The direction of the arms indicates the role of the QFD in achieving the 100 % alignment.

Most often, the ideal alignment is not the case, especially when a change in customer requirements creates a conflict in business strategic goals and processes as shown in Figure 1B. Step A represents the alignment between the customer requirements and business strategies and step B represents the alignment between the business strategies, capabilities and processes. The proposed mechanism for the two steps is described in Section 2.2.

2.2. Clock mechanism

This research proposes a business model that integrates four core elements, EBA, UML, QFD, and multivariate data analysis to achieve the needed business alignment. Multivariate data analysis along with QFD is utilized to check the degree of alignment in response to a change in customer requirements which in turn provides

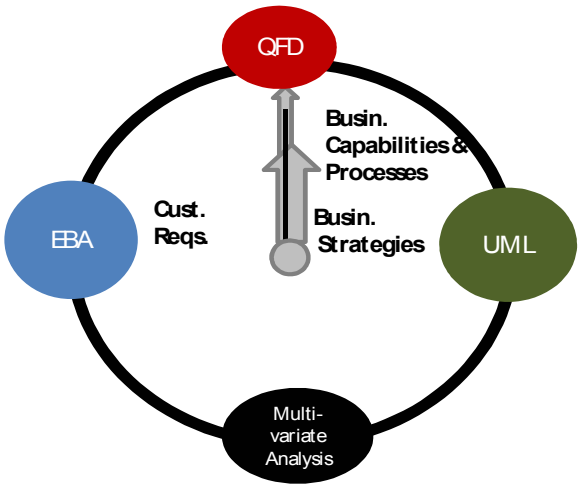


Figure 1 a). Proposed business alignment clock with 100 % alignment.

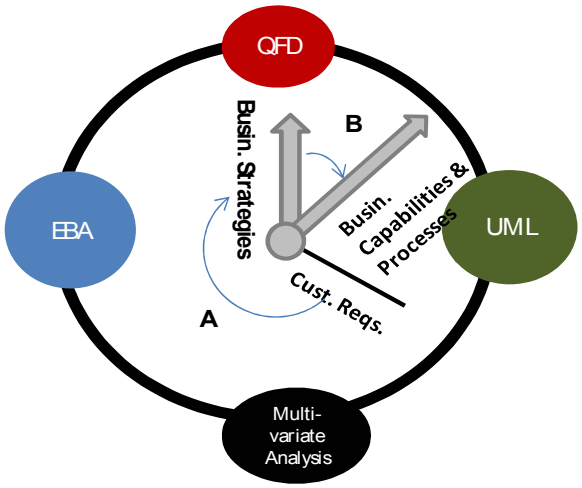


Figure 1 b). Proposed business alignment clock with <100% alignment.

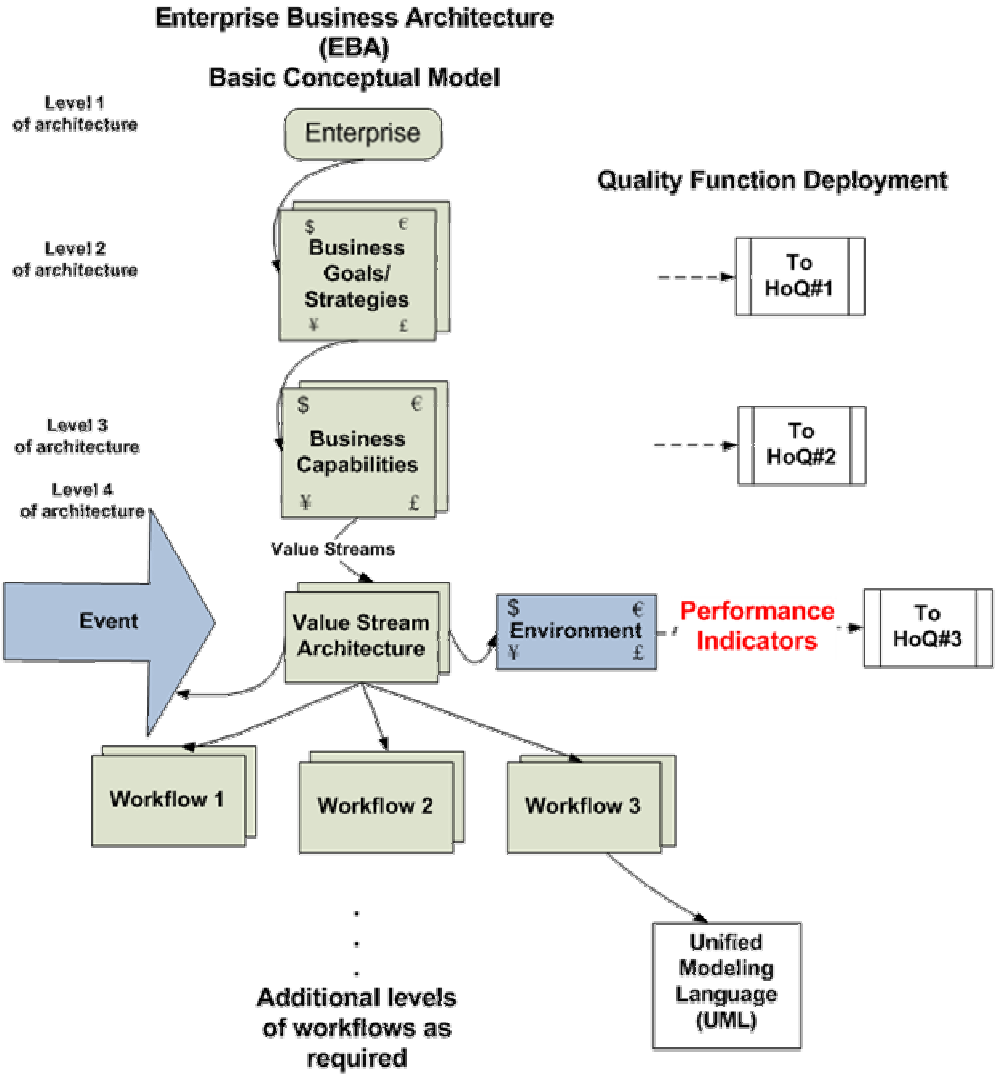


Figure 2. Re-illustrated figure of the basic conceptual structure of the EBA [8]

an accurate measure of the current state of the enterprise strategies. EBA and UML are anticipated to support the alignment process by providing QFD with information about the important processes which the enterprise must address in order to optimize the needed alignment.

The rotation of the clock's three arms indicates a change in the customer requirements that must be investigated. Hence, when a change occurs in the customer requirements, the four core tools contribute to align the clock. We use EBA, the basic conceptual model commonly used by enterprises to focus on specific model elements for analysis while understanding their interrelationships. Figure 2 depicts the basic EBA structure which provides a conceptual overview of the major components and the integration schema.

For example, an enterprise may decide to focus on a specific strategic objective in response to a change in customer requirements. Since all strategic objectives link to the enterprise capabilities with supporting metrics and measures; our hypothetical enterprise can analyze how the value streams affect a specific objective and identify the improvements needed to meet the new strategic expectations.

3. Model integration

Our model implements three HoQs to align the arms of the proposed business clock. The integration of the four elements of the model during the alignment is implemented inside the three houses in two phases: the analysis phase, and the correction phase. Figure 3 shows a diagram that describes the model and the interaction between its elements.

3.1. Analysis phase

The analysis phase (HoQ1 and HoQ2) is responsible for checking the change in the strategic goals' priorities against the change in customer requirements. HoQ1 assigns each strategic goal a weight according to its importance to the customer. The difference between the weights is then investigated to check if this change should be reflected in the current business capabilities in HoQ2, and thus whether a corrective action is needed in HoQ3.

3.1.1. House of Quality 1: Aligning customer requirements with business strategic goals

Customer requirements are inputs to HoQ1 and are gathered and prioritized through surveys, customer complaints, interviews, focus groups, etc.; they represent the WHATs in HoQ1. The initiation phase of HoQ1 is called the *base model* where the house is fed with customer requirements for the first time. After feeding the house with the new customer requirements, we refer to the house as the *dynamic model*. A comparison of the strategic goals' weights between the base and dynamic model has to be performed at this point to decide whether to move to HoQ2.

In addition to the customer requirements, the strategic goals are inputs to HoQ1, and they represent the HOWs in HoQ1.

The relationships inside the body of HoQ are quantitatively defined using the appropriate multivariate data analysis and consequently produce a prioritized list of business strategic goals. The column weights are the summations of the column values; each value is the product of the importance of the customer requirement and its strength with the strategic goal.

After finding the results of HoQ1, the strategic goals' weights have to be checked against their weights in the base model. If the new weights (after feeding the house with the new requirements) have changed from the first time the house is fed with customer requirements (base model); then we proceed to

HoQ2 to study the effect of this change on the business capabilities. However; if the change in

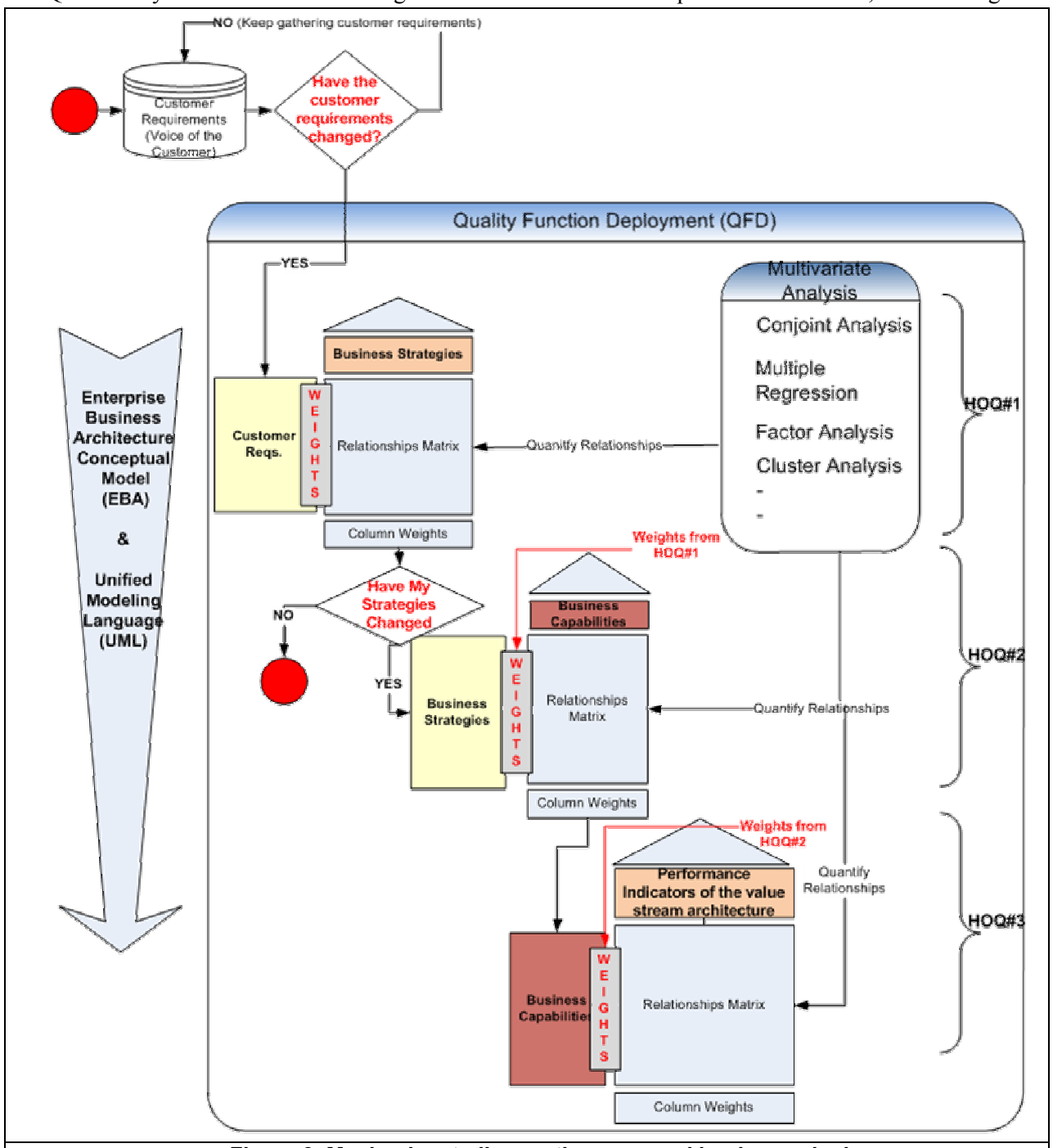


Figure 3. Mechanism to line up the proposed business clock

the customer requirements has not affected the weights of the strategic goals, the enterprise is maintaining the alignment needed and should check the strategic goals' weights against new customer requirements periodically.

3.1.2. House of Quality 2: Aligning business strategies with business capabilities

The prioritized business strategies generated in HoQ1 are inputs to HoQ2 (the WHATs), and their relationships with the business capabilities (the HOWs) are quantitatively defined using the appropriate multivariate data analysis technique. HoQ2 results in a prioritized list of critical business capabilities.

Multivariate data analysis refers to all of the statistical techniques that simultaneously analyze multiple measurements on individuals or objects under investigation [9].

Researchers may select the appropriate multivariate technique according to the application in which the proposed framework is used, type of data and relationships.

3.2. Correction phase

This phase represents the corrective actions that the enterprise should adopt to account for the strategic changes identified. HoQ3 examines the current processes (workflows), events or environment using their performance indicators and their relationships to the prioritized business capabilities.

The current relationships between the processes and the capabilities are examined and compared to the expected relationships. For example finding a large gap indicates that more attention has to be given to a particular process or event which may result in process improvement, infrastructure expansion, or software development in one or more of the business capabilities.

The Unified Modeling Language (UML) helps modify or add any IT-related processes since it provides the technical team (programmers) with a clear set of diagrams (class, use case, sequence diagrams…etc) that help them accommodate the new change.

3.2.1. House of Quality 3: Aligning business capabilities to business processes

The prioritized list of critical business capabilities are inputs from HoQ2 (the WHATs) in addition to the performance indicators of the business processes (workflows), events or environment from the business architecture (the HOWs).

The relationships' matrix inside the body of HoQ3 represents the differences between the current and the expected strength of each business capability and each process or any other architecture component. The greater the difference the more effort is needed by the enterprise to modify the architectural components associated with this difference.

In addition, UML, an extension to EBA, can be utilized to enhance the alignment in IT-enabled business processes. UML translates the business model into an IT model. The feedback loop from the IT architectures to the business architectures produces additional process improvement ideas. For example, an enterprises may want to run a simulation of the new processes to test and predict the results of the new improvements. Since the enterprise has the inputs and outputs modeled along with the events, most of the data required by a simulation product or tool will already be located in the EBA. The EBA serves as the single repository of enterprise information required by most strategic initiatives.

4. Summary and conclusion

This paper defined a novel business alignment clock and a framework that integrates EBA and the HoQ to reach an optimal level of alignment between business strategy and customer expectations. Statistical tools such as

multivariate analysis can be used to increase the robustness of the HoQ relationship matrix and to avoid any ambiguity or biases in the quantitatively defined relations.

The proposed framework increases the efficiency of quality assurance in business enterprises since the integration between QFD and business architecture leads to a more precise design of enterprises with high levels of customer focus. Using UML increases the efficiency of translating this design into an IT framework. Investment in the design of quality will pay off to the business enterprise in the short and long term.

5. References

[1] Crowe, T. J., & Cheng, C. C. (1996). Using quality function deployment in manufacturing strategic planning. *International Journal of Operations and Production Management, 16*, 35-48.

[2] Dikmen, I., Talat Birgonul, M., & Kiziltas, S. (2005). Strategic use of quality function deployment (QFD) in the construction industry. *Building and Environment, 40*(2), 245-255.

[3] Camevalli, J. A., & Miguel, P. C. (2008). Review, analysis and classification of the literature on QFD - types of research, difficulties and benefits. *International Journal of Production Economics, 114*(2), 737. Retrieved from http://proquest.umi.com/pqdweb?did=1515553171&Fmt=7&clientId=20176&RQT=309&VName=PQD

[4] Teck Khim, N., & Cher Ming, T. (2000). Integrating device modeling in QFD implementation for power electronics application. Paper presented at the *Electrical and Computer Engineering, 2000 Canadian Conference on,* Halifax, NS, Canada. *, 2* 1187. doi:10.1109/CCECE.2000.849651

[5] Chan, L. K., & Wu, M. L. (2005). A systematic approach to quality function deployment with a full illustrative example. *Omega, 33*(2), 119-139.

[6] Martins, A., & Aspinwall, E. M. (2001). Quality function deployment: An empirical study in the UK. *Total Quality Management & Business Excellence, 12*(5), 575-588.

[7] Ross, J. W., Weill, P., & Robertson, D. (2006). *Enterprise architecture as strategy: Creating a foundation for business execution* Harvard Business School Pr.

[8] Whittle and Conrad B. Myrick. (2005). *Enterprise business architecture*. US: crc press llc.

[9] Hair , J. F., Black, B., Babin, B., Anderson, R. E., & Tatham, R. L. (2006). Multivariate data analysis (6th Edition ed.) Prentice Hall.

The Journal of Management and Engineering Integration
Vol. 3, No. 1, Summer, 2010

Bank of America's Acquisition of Merrill Lynch: A Multi-Theoretical Analysis

Randall Harris and Jarrett Kotrozo
California State University, Stanislaus
raharris@csustan.edu, jkotrozo@csustan.edu

Abstract

Bank of America merged with Merrill Lynch following the global financial crisis of 2008. This unprecedented action, with the cooperation of the U.S. Federal Government, was undertaken in part to prevent the collapse of the US financial system. This merger raises a number of theoretical issues from the point of view of the management literature. Numerous perspectives, including Corporate Governance, Agency Theory, Diversification and Stakeholder Theory, are examined. Our results suggest that the extant literature does not adequately explain this merger. However, Stakeholder Theory may partially explain this merger if stakeholders are construed in a much broader context than normally suggested by the theory.

1. Introduction

Events surrounding the details of Bank of America's merger with Merrill Lynch, consummated in January of 2009, pose substantial challenges to existing management theory. Specifically, the Federal government's role in the merger, through Treasury Secretary Henry Paulson and Federal Reserve Chairman Ben Bernanke, among others, suggests that an additional layer may need to be considered with respect to monitoring in the Corporate Governance literature (Tosi, Katz and Gomez-Mejia, 1997).

Agency Theory (Jensen and Meckling, 1976; Fama, 1980; Eisenhardt, 1988) is generally silent with respect to the role of other principle stakeholders such as the U.S. government; it instead focuses on the misaligned interest of owners and managers. Thus, Agency Theory also does not seem to adequately explain this merger.

The diversification literature suggests that firm diversification destroys firm value (Lang and Stulz, 1994; Berger and Ofek, 1995). Regardless of this broadly accepted research result, however, the merger was consummated. Shareholders, in particular, appear to have been negatively impacted.

Stakeholder Theory may provide a partial explanation (Donaldson and Preston, 1995), but only by moving far away from Friedman's (1962) value-maximization version of Stakeholder Theory. Somewhat consistent with Stakeholder Theory, though, the shareholders of Bank of America appear to have been forced to bear the brunt of the risk and cost associated with the government's effort to curtail the global financial crisis of 2008.

The purpose of this paper is to examine the theoretical issues raised by this merger and to briefly explore some of its complex implications. Fruitful suggestions for future research are also presented and discussed.

The paper begins with a brief discussion of the financial crisis of 2008 after which we discuss a broad outline of the merger itself. With this foundation, the focus then shifts to a theoretical analysis of the merger through the lens of corporate governance, agency theory, and the diversification literature and stakeholder theory. Our results broadly suggest that the merger is a-theoretical; that is, extant theory

appears to be inadequate to explain this merger. Stakeholder Theory may provide a partial explanation for the merger, but only if stakeholders are construed in the broadest sense possible. We also briefly address the implications of this merger, with a discussion about the implications for both scholars and executives.

2. Roots of the crisis

The trigger that sparked the global financial crisis of 2008 is widely regarded to be the collapse of the US market for subprime mortgages in February 2007 (Greenspan, 2010). Although subprime mortgages were a relatively small part of the overall US mortgage market, its collapse started the ball rolling that eventually culminated in the global financial crisis of 2008.

Events quickly accelerated, and eventually engulfed the entire banking system. By the close of trading on Friday, September 12, 2008 it was clear that the investment bank Lehman Brothers would not survive the weekend without immediate assistance, the insurance firm AIG would be on life support after taking catastrophic losses, and Merrill Lynch executives would be certain that they were running out of time (Sorkin, 2009).

3. Fateful weekend in September

Negotiations over the fate of Lehman Brothers, Merrill Lynch and AIG began early on the morning of Saturday, September 13th. By 6am, Greg Fleming of Merrill Lynch was on the phone with John Thain, the company's CEO, regarding the need for Bank of America (BAC) to acquire their Merrill Lynch (MER). Thain, originally hesitant, agreed to keep an open mind throughout the day. This call was timely because, later in the day, Merrill Lynch would experience extreme liquidity pressures. Margins were under severe duress, with JP Morgan alone upping their margin call to MER by $10 billion.

By 2pm on Saturday, executives from BAC and MER had begun negotiations. MER wanted at least $30/share for their firm, which was a hefty premium over the company's trading price of $17.05 on that Friday. Teams from both firms, along with outside counsel, prepared to conduct due diligence on the merger within 24 hours, an unprecedented timetable. By 5pm, Secretary Paulson had concluded that MER was in imminent danger of collapse, and urged MER CEO Thain to conclude a merger, despite Thain's increasing unease (Sorkin, 2009).

At 3am on Sunday morning, bankers and lawyers from BAC and MER were still working at the Law Offices of Wachtell, Lipton, Rosen & Katz, attempting to hammer out a deal.

The meetings on Sunday morning started early as well. By 8am, all of the CEOs of the major New York investment banks were in a meeting at the New York Federal Reserve Building in lower Manhattan. At this meeting, led by Treasury Secretary Paulson, New York Federal Reserve President Tim Geithner and Securities and Exchange Commissioner Chris Cox, it was announced that the US Government had an offer on the table for Lehman Brothers. Barclays Bank, they said, would acquire Lehman, and what was needed from the CEOs present was financing of $33 billion for a "bad bank." This "bad bank" would guarantee the assets that Barclays had refused to acquire as part of their deal for Lehman Brothers. As a result of this announcement, furious negotiations ensued between the New York CEOs regarding how to structure the deal to save Lehman.

By 11am Sunday morning, Treasury Secretary Paulson was reasonably confident that he had a deal to save Lehman well under way. One wrinkle that had emerged, however, were UK regulators. Some in the UK had expressed

reservations regarding what was occurring in New York. Further, and to a great extent, they felt out of the loop. To smooth the waters, Secretary Paulson called Alistair Darling, Britain's Chancellor of the Exchequer who informed Secretary Paulson that he wouldn't support a deal between Barclays and Lehman Brothers. This led Secretary Paulson and New York Fed President Geithner to prepare for Lehman's bankruptcy, including the unwinding of its trading positions.

By mid-afternoon on Sunday, Bank of America and Merrill Lynch had reached agreement on a merger. Terms for the deal included a $29/share price for MER shares, bonuses for Merrill employees (the source for great controversy later), and a strong material adverse change (MAC) clause (also the source for later controversy). *The closing date on the merger was set for January 1, 2009*. Bank of America then scheduled a board meeting for 5pm, and Merrill's board scheduled to meet at 6. Before the vote, Bank of America board members were surprised that MER was acquired, and not Lehman, suggesting that the board was unaware of the negotiations with Merrill until the meeting to approve the merger (Sandler, 2010).

That evening, in addition to rumors about Lehman, the news quickly making the rounds was that AIG was "in a bind." Michael Bloomberg quipped that evening in New York, "The world is about to end tomorrow" (Sorkin, 2009, p. 369).

4. Global financial crisis

The headline from the Wall Street Journal said it all on Monday, September 15th: *CRISIS ON WALL STREET***.** The damage to the financial system began to accelerate at this point. Lehman Brothers had filed for bankruptcy protection at 1:45am that morning. Bank of America's stock had closed down 21% after the announced merger with Merrill Lynch. The Dow would end the day down 504.48 points. At 4pm that afternoon, AIG would receive a $14 billion loan from the Federal Reserve, who would take an 80% equity stake in the company, effectively nationalizing it. Top AIG officials were told that they no longer had jobs.

As the crisis spread, Treasury Secretary Paulson declared that the situation was an "economic 9/11" (Sorkin, 2009, p. 417). He would tell then President Bush on Wednesday, "If we don't act boldly, we could be in a depression greater than the Great Depression" (Sorkin, 2009, p. 440). After consultation with members of Congress, Paulson would announce the Troubled Asset Relief Program (TARP), on Friday, September 19th. It would take two votes in the US House, and the single greatest one-day drop in the Dow (777.68 points, or over $1 trillion in stock market value), to get the agreement approved. The $700 billion TARP, originally designed to purchase toxic assets, was then used (among other things) to buy equity stakes in major US banks, and succeeded in ultimately offering some stability to the global. Global stock markets, however, would continue to descend until March of 2009.

5. Losses mount at Merrill Lynch

The agreement between Bank of America and Merrill Lynch did nothing to stop losses at Merrill Lynch from continuing to mount, in the third quarter of 2008; Merrill reported a $5.1 billion loss. By November, Merrill's internal reporting units were projecting a $6.1 billion loss for the month of October alone. There was heated discussion within the Merrill Lynch ranks about the rapidly increasing losses. Further, consultation between Merrill executives and their legal staffs, both internal and external, show discussions regarding whether or not the mounting losses should be disclosed.

The extent of the losses, the timing of the losses, who knew about the losses, their method of calculation and the assumptions underpinning these estimates are under investigation and litigation as of March 2010. What is clear from the litigation is that by mid-December, the losses at Merrill were significant enough to merit discussion at Bank of America regarding whether or not to consummate the merger on New Year's Day, 2009.

Much of the discussion revolved around whether or not to invoke the MAC clause. Legal opinion around whether the losses at MER were significant enough to invoke the clause vacillated between yes and no. There was also discussion regarding whether or not to attempt to renegotiate the merger agreement between the two companies on terms more favorable to Bank of America and its shareholders. Unfortunately, as Bank of America executive Greg Curl observed on December 15th, Merrill would probably not survive if they had to renegotiate. In other words, either invoking the MAC clause or attempting to re-negotiate the merger deal could have potentially destroyed Merrill Lynch. Further, the failure of Merrill Lynch had the potential to destabilize the global financial system (once again).

6. Endgame

Finally, after again consulting with outside counsel, Ken Lewis called Secretary Paulson on the afternoon of December 17th. Lewis explained that the Bank of America board had significant concerns about whether to go ahead with the deal to buy Merrill Lynch. Lewis conveyed to Paulson that he had learned that the losses at Merrill in the 4th quarter of 2008 could be as much as $18 billion, pretax. Lewis further stated that the Bank of America board was considering invoking the MAC clause to get out of the deal with Merrill. After some discussion, Paulson arranged a 6pm meeting in Washington, DC at the Federal Reserve.

At approximately 5pm, Secretary Paulson, Fed Chairman Ben Bernanke, and several Treasury and Fed staffers met at the Fed. Federal Reserve officials at the meeting indicated that they were aware that Bank of America would post a loss in the fourth quarter and that the bank had a weak capital ratio, indicating a good deal of stress on their balance sheet. Paulson and Bernanke agreed, after some discussion, to take a tough stand against Bank of America exercising the MAC clause. Secretary Paulson also expressed concern about market reaction to an $18 billion pretax loss for Merrill Lynch.

The Bank of America team, consisting of Ken Lewis (CEO), Joe Price (CFO), and Brian Moynihan (General Counsel) arrived promptly at the Federal Reserve at 6pm. At the meeting, Lewis described the mounting losses at Merrill and raised the possibility of exercising the MAC clause of the merger agreement. Chairman Bernanke pushed back strongly once the discussion broached the MAC clause, saying that such a move could lead to a run on the bank. When Lewis asked Bernanke if he was referring to Merrill, Bernanke replied, "No, both Merrill and Bank of America..." (Paulson, 2010, p. 426).

Someone at the meeting raised the possibility of the US Government giving Bank of America a support package similar to one received by Citigroup at an earlier stage in the crisis. Chairman Bernanke said that such support was given because of systemic risk, not to facilitate a merger. At this point both Secretary Paulson and Chairman Bernanke indicated that they were committed to preventing the failure of any systemically important institution. The meeting ended with agreement for each side to talk further and to take no action without further consultation. Secretary Paulson, as a result of the

meeting, was concerned that losses at Merrill Lynch now threatened the viability of both Merrill and Bank of America.

By the next morning, Friday, December 19th, Bank of America staff had reported to the Federal Reserve that losses at Merrill had exceeded $22 billion, pretax. Paulson, increasingly distressed, called Lewis that afternoon to ask how losses at Merrill could increase by $4 billion in less than 48 hours. Lewis replied, "I am trying to understand that myself..." (Paulson, 2010, p.428).

An hour later, Paulson received a conference call from Lewis, the Bank of America team, and dozens of Federal officials from New York, Washington, DC, and Richmond, Virginia. On the call, Lewis repeated that the Bank of America board was still considering invoking the MAC clause. Officials from the Federal Reserve questioned the enforceability of the clause, noting that the clause was unusually restrictive. Paulson then stated that invoking the MAC clause would pose a risk to Bank of America and the entire banking system. When Lewis suggested using the MAC clause to renegotiate the deal with Merrill Lynch, Paulson stated that renegotiating would cause the same problems: uncertainty and fear in the market. The conference call ended with agreement to discuss the issue the following week.

Secretary Paulson flew to Colorado for some much needed rest on Saturday, December 20. The following morning, though, Ken Lewis called Paulson. Paulson would recall that Lewis sounded shaken on the call. Lewis reiterated to Paulson that the Bank of America board was concerned about Merrill's losses and was still considering invoking the MAC clause. Paulson told Lewis that the Treasury Department and Federal Reserve were committed to saving systemically important institutions and that they would work on a support package, if needed. "You know how strongly we feel about this," Paulson said (Paulson, 2010, p.429).

Later in that day, Paulson called Lewis back. The two rehashed their previous discussion. Further, Paulson reemphasized that the government would not let any systemically important institution fail, that exercising the MAC clause would show a colossal lack of judgment by Bank of America, that such an action would jeopardize the bank, Merrill Lynch and the entire banking system; and that under such circumstances, the Federal Reserve, as Bank of America's regulator, could take extreme measures, including the removal of management and the board of Bank of America.

"I understand," Ken (Lewis) said. "Let's de-escalate" (Paulson, 2010, p. 430).

7. Aftermath

Merrill Lynch was acquired by Bank of America on January 1, 2009. Two weeks later, on January 15th, the US Treasury gave $20 billion in TARP money to Bank of America. In return for this cash, the Treasury was given preferred stock in Bank of America with an 8% dividend. As part of the agreement, BAC would absorb the first $10 billion in losses on a $118 billion pool of toxic loans, much of it coming from Merrill Lynch. Losses accrued beyond the first $10 billion in losses would be split 90/10 by the government and then BAC.

One day later, BAC would release their Q4 earnings. BAC would report a $1.79 billion loss for their company (pre-merger) and a $22 billion pretax loss for Merrill Lynch. After this report was released, BAC shares would fall 14% to $7.18 per share.

John Thain lost his job over the Merrill losses. In a meeting that would reportedly last

less than 15 minutes, Thain was fired by CEO Ken Lewis as BAC's President of Wealth Management and Global Corporate Investment Banking on January 22, 2009.

Lewis, however, would also lose his job over the deal. Ousted as Chairman of the Board in April, he would announce his resignation as CEO and member of the Board on September 30, 2009. Most of the other executives closely associated with the deal, including Greg Curl and Joe Price, would also be fired or pushed out as well.

One executive to weather the storm was Brian Moynihan. After an extensive and very public global search, Moynihan was named CEO of Bank of America on December 16, 2009.

8. Theoretical analysis

The events surrounding Bank of America's acquisition of Merrill Lynch has ties to numerous research streams. Arguably the most important are the implications this case has with respect to Corporate Governance. However, Agency Theory, Stakeholder Theory and Diversification Theory are also relevant, and related, areas of discussion.

8.1. Corporate Governance and Agency Theory

Corporate governance attempts to align the interest of the dispersed owners of an organization with the divergent interests of other stakeholders, most commonly the managers of the organization (Monks & Minow, 2004). Within the Management literature, there are two primary foundations on which corporate governance mechanisms rest: incentive alignment and monitoring (Beatty and Zajac, 1994; Zajac and Westphal, 1994, Tosi et al., 1997).

According to Tosi et al., monitoring is, "[O]bservation of an agent's effort or outcome that is accomplished through supervision, accounting controls and other devices" (1997: 588) that ensures behavior is consistent with the interest of firms' shareholders. Monitoring has a number of mechanisms which include the influence of large shareholders (Dalton, Daily, Ellstrand and Johnson, 1998; Rediker and Seth, 1995; Zahra and Pearce, 1989; He and Wang, 2009), management provisions (Gompers, Ishii and Metrick, 2003), the Board of Directors (Jensen and Meckling, 1976; Demsetz and Lehn, 1985) and outside regulatory agencies such as the Federal Reserve, the Department of the Treasury, the Securities and Exchange Commission (SEC) and the Federal Deposit Insurance Corporation (FDIC).

During the last days of 2008, Bank of America's CEO, Ken Lewis, was increasingly concerned about the impending MER merger. Growing losses at Merrill Lynch had members of BAC's Board of Directors concerned that consummation of this deal could be potentially devastating to the long-term welfare and perhaps the very survival of Bank of America. As members of the Board of Directors (a prominent monitor) discussed whether these losses were relevant under the MAC clause of the agreement, Lewis is characterized as having a difficult time with his Board (Paulson, 2010, p. 429) which has a fiduciary duty to act in the best interest of BAC shareholders (Colley, 2005).

Upon learning of these reservations from both Lewis and BAC's Board, Secretary Paulson, on December 21, 2008 is reported to have told Lewis that a decision of Bank of America to back out of their agreement regarding MER would jeopardize the entire financial system and that, under such circumstances, the Federal Reserve as Bank of America's regulator (yet another monitor), could

take extreme measures including the removal of management, including Lewis himself, as well as the members of the Bank of America's Board (Paulson, 2010, p. 430).

To our knowledge, the comments by Secretary Paulson are unprecedented. Although regulatory agencies are not always expected to act in the best interest of shareholders, particularly if illicit or inappropriate behavior of a company's management is involved, Paulson's virtual insistence that this merger proceed as previously agreed, when the Board of Directors are performing their legally-bound fiduciary duty to their owners, is in direct contradiction to their role as a monitor on behalf of shareholders.

Both the literature and the law are clear about the role of a company's Board, which does not include embarking on a potentially harmful agreement in order to balance the interests of all of the stakeholders of the organization.

Agency Theory, acting in one's own best interest instead of in the best interest of the shareholders of the firm (Berle and Means, 1932; Jensen and Meckling, 1976), is precisely what corporate governance mechanisms are intended to mitigate.

The Agency Theory and corporate governance literature is (understandably) silent with respect to government agencies, and those within these agencies, acting in the best interest of the "whole," or even acting in their own individual interest within a predominantly market economy such as the United States. However, with Secretary Paulson suggesting that Bank of America take on a severely crippled company such as Merrill Lynch because it is in the best interest of the US financial system, there is reason to suggest that a new agency problem has been created at a government level of analysis.

The Federal government's actions raise a number of relevant questions including whether it is appropriate for a representative of the Federal government to use his or her influence at the expense of the shareholders of an organization, and if companies will take on additional risk in an effort to become "too big to fail" in order to receive similar treatment from the Federal Government. It is also possible that Secretary Paulson may have been acting in his own best interest from a political or a career standpoint. These, and other, questions arise from the details of this case with no clear direction on how to address, let alone mitigate, the potential for similar issues in the future.

8.2. Stakeholder Theory

Under Stakeholder Theory, the purpose of a firm is to create value for all of its stakeholders (Donaldson and Preston, 1995). From this standpoint, the actions of Secretary Paulson, and ultimately CEO Lewis and the other members of BAC's Board of Directors should be to the benefit of all interested parties. Although the actions could impair or destroy value for one sub-group of stakeholders, such as BAC's shareholders, the aggregated benefits to all of the stakeholders is of primary importance.

While Stakeholder Theory has merit, and is particularly relevant to this case based on the actions of Secretary Paulson and other members of the top management team and the Boards of Directors of both companies, there are a number of unresolved issues both with respect to this case and within the literature. The first has to do with the relative importance of each of the stakeholders of the organization. The literature is unclear regarding the "most important" stakeholder and just how much more important that stakeholder is relative to others.

Furthermore, there is ambiguity regarding when the relative importance of stakeholders change.

For example, one can strongly argue that shareholders are the most important stakeholder of the organization. However, just how much more important they are than the employees, the customers or even the environment is difficult to quantify. From the case, it also appears that the relative importance of both Bank of America's shareholders as well as its management was subordinated in favor of the overall health of the US economy.

From Bank of America's perspective, the acquisition of Merrill Lynch increased the risk to BAC shareholders without a corresponding increase to expected returns. Other stakeholders of both organizations may be satisfied by the merger, consistent with Stakeholder Theory. However, this was not enough to satisfy BAC shareholders who unwound their positions *en masse*, as is evident by BAC's falling share price during this period. Should investors find that Stakeholder Theory is the primary purpose of the firm, they will begin to require additional returns on their investments to compensate for the increased risk associated with the variety and variance of desired outcomes. Should they fail to be compensated for this increased risk, investment will migrate to companies, industries or countries where this risk is minimal which could have an equally devastating impact on the US economy if investors are wary that the events during this economic crisis will recur.

Although corporate governance, Agency Theory and Stakeholder theory are distinct literature streams, they are highly inter-related as Wearing (2005) notes. Corporate governance is intended to help mitigate the agency problems prevalent in market-based economies. Stakeholder Theory also helps to overcome agency issues by reducing the emphasis on a firm's profit maximizing role (Friedman, 1962); instead emphasizing the importance of parties both directly and indirectly affected by an organization. In this way, Stakeholder Theory is able to compliment corporate governance.

8.3. Diversification Theory

Indirectly related to the previous discussion, the details of BAC/MER also have implications regarding the diversification literature. The merging of these two organizations constitutes a related diversification event of two financial institutions.

Amit and Livnat (1988) nicely articulate two broad categories as to why a firm would engage in such an event. First, there may be synergistic benefits of such a merger. Second, companies may capitalize on financial benefits. Although unresolved in the literature, related diversification strategies, such as BAC/MER can be expected to lead to better performance (Rumelt, 1982).

However, it is unclear from the information on this merger which, if any, of these potential benefits may be derived for Bank of America. Related to financial benefits, efficient internal capital markets is unlikely as the agreement stated that Bank of America would absorb the first $10 billion in losses from Merrill Lynch, with any losses greater than $10 billion absorbed by both Bank of America and the US Government in a 10/90 split (Paulson, 2010). If anything, this type of arrangement is an inefficient use of internal capital markets, diverting resources from successful areas of BAC to support failing areas of MER.

Based on previous discussions, risk is not likely to be reduced through this merger, and may actually increase. Furthermore, risk reduction is a weak argument in support of

diversification as investors are much better at reducing risk through portfolio diversification than companies are at reducing risk through diversifying cash flows.

It is difficult to determine if the merged entity is able to capitalize on the human capital from both companies. However, both BAC and MER appeared to have slack resources as both companies reported losses during this period. Furthermore, it is unlikely that economies of scale and scope were present as the bad assets, which are suggested to be at the root of the crisis, were present at both companies.

Also of considerable importance is that, prior to the crisis and the government's involvement with both companies, it is questionable whether a merger between these two organization would have been approved without considerable government scrutiny from the Department of Justice because of market power considerations. However, with the government's blessing (or insistence) during the financial crisis, it is yet to be determined what impact this merger of two large financial institutions will have on competition over the long-term assuming the new entity survives.

While the management literature offers some support for value creation associated with related diversification (Rumelt, 1982; Palepu, 1985), the finance literature offers findings to suggest that any kind of diversification is value-destroying (Lang and Stulz, 1994; Berger and Ofek, 1995).

On two separate occasions, considerable shareholder value was destroyed for BAC shareholders as a result of the merger. The first was immediately following the announcement of the merger when BAC shares fell 21% destroying more than $70 billion in value. The second was immediately following the announcement of $22 billion in losses attributed to MER in January, 2009 when more than $11 billion of value was destroyed.

Since the start of the financial crisis, the market capitalization of BAC has fallen as low as $32 billion (March 6, 2009) roughly 90% lower as the value of the company in mid 2008. (BAC's market value has since rebounded to more than $100 billion as of this writing.)

While it is not possible to attribute all of BAC's market decline to its merger with MER, it has clearly had a negative impact on the value of the firm and has harmed its shareholders with almost $200 billion of value destroyed since mid 2008.

Under the umbrella of diversification theory, the results of Rumelt and Palepu notwithstanding, those with a fiduciary responsibility to its shareholders (specifically, BAC's Board of Directors) were clearly derelict in this particular merger. Stakeholder theory would suggest that, while $200 billion is a substantial sum, it is small relative to the potential collapse of the US financial system. Agency theory would suggest that the government (a potential new principle) was acting in its own best interest vis-à-vis the agents (both the executives of BAC, its Board of Directors and ultimately the firm's shareholders). However, this result is outside the normal scope of Agency Theory, which normally only considers the direct owners (shareholders) and managers of the firm. Agency theory would suggest that the government (a potential new principle) was acting in its own best interest vis-à-vis the agent (both the executives of BAC, its Board of Directors and ultimately the firm's shareholders) which raises the question: Who monitors the monitors?

9. Implications

The events surrounding the Bank of America/Merrill Lynch merger in 2008 and 2009 are thus a-theoretical with respect to the extant literature. From a Corporate Governance perspective, managers and directors must now consider recourse from the owners of an organization when their company is confronted with pressure from outside entities, such as regulatory agencies, to take action which may not be in their best interest. Managers must consider what governance mechanisms related to large shareholder or management provisions may be used to address this potential, and how should this be accomplished. The events of this case appear to contradict the traditional arguments related to corporate governance.

Knowing that the government may step in for the "overall good of the economy," both monitors and researchers need to consider an expanded scope of agency problems related to moral hazard, as managers may take on excessive risk in an effort to become "too big to fail" to receive government assistance. Furthermore, additional research could consider whether the Federal government overstepped its authority in a predominantly market-economy like the United States both with respect to its involvement in this merger as well as the decision to allow Lehman Brothers to fail, after regulators in the United Kingdom refused to bless the merger of Lehman with Barclays (Paulson, 2010). Understanding why one financial institution was chosen to survive through a merger while another was allowed to fail has strong implications for managers and academics alike in anticipation of and preparation for the inevitable "next big crisis" (Greenspan, 2010).

From a stakeholder perspective, research must focus on whether some stakeholders are more important than others. If so, it must also be determined just how much more important they are, and how their desired outcomes are subordinated when these desired outcomes are at odds. Investors must also begin to consider how to evaluate the increased risks associated with Stakeholder Theory.

Boards of Directors and executives too must explicitly consider situations in which the 'greater good' may mean adverse consequences for themselves or their company's shareholders. The spate of firings of executives and shareholder lawsuits since the consummation of the merger suggest that these are important and, as yet, unresolved issues deserving of consideration. These issues may be correlated with literatures that examine the effects of corporate governance on market development and firm value (La Porta, et. al., 1999).

Finally, with respect to diversification theory, little work has been done to examine the effects stakeholders have on the value of the firm when mergers are influenced external stakeholders such as government entities. Event studies which examine the effect of the various announcements during the 2008-2009 time periods may offer insight into this (and other) questions. The long-term impact of such a merger on the financial services landscape has yet to be determined, but will also benefit from further examination and research. Further unresolved is the question of whether Bank of America benefitted from the demise of Lehman Brothers. From a consumer perspective, it is also unclear whether consumers benefited from the merged entity or whether competition in retail banking will be stifled. Additional research is needed to draw definitive conclusions regarding these questions.

10. Conclusion

The events surrounding the Bank of America/Merrill Lynch merger represent unprecedented events in US history. High-ranking officials such as Federal Reserve Chairman Ben Bernanke and U.S. Treasury Secretary Paulson recognized the potential that the U.S., and possibly global, financial system was close to collapse in 2008. This was an almost unthinkable scenario. Markets also understood the risk as the Dow Jones Industrial Average fell more than 500 points on two separate occasions during the crisis.

While the crisis was averted, in large part to the tireless work of Secretary Paulson, Chairman Bernanke and others, the implications from these events during the height of the financial crisis will be felt and studied for many years to come.

One outcome from the events of late 2008 was the merger of Bank of America with Merrill Lynch. This action was taken, by all accounts, to mitigate the possibly catastrophic consequences from the collapse of one of the largest investment banks in the world. What is unknown, however, is whether this merger, and others like it, addressed an immediate concern at the price of increased systemic risk in the future.

From an academic perspective, the merger is deeply troubling. Bank of America's merger with Merrill Lynch does not conform to the tenets of standard management theories. Only an extremely broad and unusual interpretation of Stakeholder Theory appears capable of interpreting the merger. Additional research is necessary to determine whether this event was an anomaly, or whether extant managerial theory will need revision in light of current events.

An overarching question to be asked is: now that the genie is out of the bottle, will these events represent the first step onto a slippery slope with far-reaching, long-term consequences? Or do they represent a curious anomaly which will constitute not much more than an anecdote in our yet-to-be-written history books? This paper draws upon a number of related, but distinct, literatures to provide a framework to begin thinking about, and ultimately, answering this important question.

11. References

Amit, R. & Livnat J. (1988). "Diversification Strategies, Business Cycles and Economic Performance." *Strategic Management Journal*, 9:2, 99-110.

Beatty, R. P. & Zajac, E. J. (1994). "Managerial Incentives, Monitoring, and Risk Bearing: A Study of Executive Compensation, Ownership, and Board Structure in Initial Public Offerings." *Administrative Science Quarterly*, 39, 313-335.

Berger, P. G. & Ofek, E. (1995). "Diversification Effects on Firm Value." *Journal of Financial Economics*, 37, 251-273.

Berle, A. A. and Means, G. C. (1932). *The Modern Corporation and Private Property.* New York, NY: Macmillan.

Colley, J., et. al. (2005). *What is Corporate Governance?* New York, NY: McGraw Hill/Irwin.

Dalton, D. R., Daily, C. M., Ellstrand, A. E. & Johnson, J.L. (1998). "Meta-Analytic Reviews of Board Composition, Leadership Structure and Financial Performance." *Strategic Management Journal*, 19, 269-290.

Demsetz, H. & Lehn, K. (1985). "The Structure of Corporate Ownership: Causes and Consequences." *Journal of Political Economy*, 93, 1155-1177.

Donaldson, T. & Preston, L. E. (1995). "The Stakeholder Theory of the Corporation: Concepts, Evidence and Implications." *Academy of Management Review*, 20, 65-91.

Eisenhardt, K. M. (1988). "Agency- and Institutional-Theory Explanations: The Case of Retail Sales Compensation." *Academy of Management Journal*, 31, 486-511.

Fama, E. F. (1980). "Agency Problems and the Theory of the Firm." J*ournal of Political Economy*, 88, 288-307.

Friedman, M. (1962). Capitalism and Freedom. The University of Chicago Press, Chicago, IL.
Gompers, P., Ishii, J. & Metrick, A. (2003). "Corporate Governance and Equity Prices." *Quarterly Journal of Economics*, February, 107-155.

Greenspan, A. (2010). *The Crisis*. Paper Presented at the Brookings Institution. Second Draft Dated March 9, 2010. Downloaded at www.wsj.com.

He, J. & Wang, H. C. (2009). "Innovative Knowledge Assets and Economic Performance: The Asymmetric Roles of Incentives and Monitoring." *Academy of Management Journal*, 52, 919-938.

Jensen, M. C. & Meckling, W. H. (1976). "Theory of the Firm. Managerial Behavior, Agency Costs and Ownership Structure." *Journal of Financial Economics*, 3, 305-360.

La Porta, R., Lopez-de-Silanes, F., Shleifer, A... (1999). "Corporate Ownership Around the World." *Journal of Finance*, 54, 471-517.

Lang, L. H. P. and Stultz, R. M. (1994), "Tobin's *q*, Corporate Diversification and Firm Performance." *Journal of Political Economy*, 102:6, 1248-1280.

Monks & Minow (2004). *Corporate Governance, 3rd Edition*. Cambridge, MA: Blackwell Publishers.

Palepu, K. (1985). "Diversification Strategy, Profit Performance and the Entropy Measure." *Strategic Management Journal*, 6:3, 239-255.

Paulson, H. (2010). *On the Brink: Inside the Race to Stop the Collapse of the Global Financial System*. New York: Business Plus.
Rediker, K.J. & Seth, A. (1995). "Boards of Directors and Substitution Effects of Alternative Governance Mechanisms." *Strategic Management Journal*, 14, 107-122.

Rumelt, R. P. (1982). "Diversification Strategy and Profitability." *Strategic Management Journal*, 3:4, 359-369.

Sandler, L. (Feb. 4, 2010). Bank of America E-Mails show Lehman was buy target. Retrieved at bloomberg.com.

Sorkin, A. R. (2010). *Too big to Fail: The Inside Story of the How Wall Street and Washington Fought to Save the Financial System from Crisis - and Themselves*. New York: Viking.

Tosi, H. L., Katz, J. P. & Gomez-Mejia, L. R. (1997). "Disaggregating the Agency Contract: The Effects of Monitoring, Incentive Alignment and Term in Office on Agent Decision Making." *Academy of Management Journal*, 40, 584-602.
Wearing, R. (2005). *Cases in Corporate Governance*. London, UK: Sage Publications.

Zahra, S. A. & Pearce, J. A. I. (1989). "Boards of Directors and Corporate Financial Performance: A Review and Integrative Model." *Journal of Management*, 15, 291-334.

Zajac, E.J. & Westphal, J. D. (1994). "The Costs and Benefits of Managerial Incentives and Monitoring in Large U.S. Corporations: When is More not Better?" *Strategic Management Journal*, 15 (Winter Special Issue). 121-142.

The Journal of Management and Engineering Integration
Vol. 3, No. 1, Summer, 2010

Distributed Intelligence Framework for P2P File Sharing

James Jacobs
University of Southern Mississippi
usm.james.jacobs@gmail.com

Abstract

We propose a framework that provides an intelligence mechanism to address problems relevant to P2P networking. The framework is designed with the intelligence integrated and exchanged from peer-to-peer. The peers communicate, design, and administer the network for the good of the user, the network, and the infrastructure.

1. Introduction

Peer-to-peer (P2P) networks are a collection of computer systems organized in a manner which allows the sharing of the peers collective resources. The resources being shared by the network may include CPU cycles, disk space, or content (files). The sharing of resources provides an outstanding mechanism for the distribution of media. The primary focus of this work is to enhance this ability.

P2P networks use a variety of protocols to create and maintain the interconnections and list the shared resources. One of the key components of the P2P operation is the distributed hash table (DHT) which stores the (key, value) pair mapping (a stored reference to the location of the resource). As such, the DHT provides the basis for listing and querying needed to route a peer looking to download a resource and the peer sharing it. Several protocols and applications relating to P2P networks utilize DHT, including CAN (Content Addressable Network) [1], Chord [2], Pastry [3], Tapestry [4], Kademlia [5], Bittorent [6] and Gnutella [7].

Several open problems exist in P2P file sharing which still need to be addressed. Efficiency to find and transfer files is always a big concern for P2P network as the primary focus is the ability for peers to share this information. Considerable work has been focused on both the ability to search the network and transmit the information between peers much to the delight of the end user and the content provider. However, the leaps in efficiency have caused concerns for internet service providers (ISPs) as the sharing of files consumes considerable bandwidth and creates additional expense for the ISP.

File distribution and availability are two of the keys to efficiency of the P2P network. Where the files are located and how they are distributed in the P2P overlay network drastically affects the efficiency to locate the information. The efficiency of the actual transfer of the resource is affected by where the files are located in respect to the peer retrieving the information.

Malicious files and peers is a constant concern. In addition to causing disruption for the user and the P2P network, the malicious behavior causes extra traffic between peers which can be eliminated with the appropriate mechanism.

The solutions proposed in this paper shift the focus of a P2P network from a resource centric approach to a collective approach adding awareness to the user, the network and the subject when grouping the resources. Using the added awareness and information gathered an

intelligent DHT is proposed to organize the P2P network and increase the efficiency of searching and transferring information in a P2P.

The remainder of this work is organized in the following order: Section 2 covers related work. Section 3 defines the awareness of the file subject, file location and the user's interest when utilizing the P2P network. Section 4 defines the decision support mechanism utilized to gather and share the information. Section 5 explains how the information gathered is used to intelligently arrange the DHT based on peers interest. Section 6 covers file availability and enhance distribution techniques that are also made possible by this proposal. Section 7 concludes the paper.

2. Related work

Segmenting the P2P network has been shown to increase the efficiency of searching and reducing updates via multiple techniques. Topic segmentation to enhance search and reduce network traffic is proposed by applying SETS [8]. SETS utilizes a topic-segmented overlay network to using short distance links to connect related items and long distance links to connect the short segments to the complete the network. OntoZilla [9] combines ontology with P2P in order to group the peers within the network based on semantics. GridVine [10] is an overlay on top of P-Grid P2P, which applies semantic inheritance and semantic gossip to build the relationships between peers and files.

Peer location awareness, as it relates to the files within the P2P, has been shown to increase the efficiency of P2P file sharing by giving precedence to the files within the local domain over files outside. The use of proxy-trackers at the ISP to redirect the requests to a local copy and the modification of the BitTorrent trackers to add prefix rules, domain names or more sophisticated methods to cluster the information are the proposed in [11].

Unlike the previous works described, our approach is to use a combination of awareness of the Subject, Location, and Users requirements entered into a DSS to address several problems in P2P networking. Instead of using an overlay to create the relations, the name of the file being added contains the relations used to link the semantics to the DHT.

3. Awareness of P2P

An aware P2P is a network that understands, adapts, and manipulates the defining structure based the provided information. The current state of P2P awareness consists mainly of a data centric view of the network by focusing on the addition and removal of the resource from network. In this section, we define additional awareness to the subject, the user and the network which provides the building blocks of the distributed intelligence framework.

3.1. Subject aware P2P

The current approach for most P2P networks is to assign keys to peers based on resource title without regard to subject or other relevant information. This approach works well for the user as most people are searching based on a title of interest, however, this approach has negative effects on the P2P network. One of these effects is displayed when hashing dissimilar resources to locations local to each other in the hash. The results returned by the search are not entirely what the user desired causing further manual searching or worse another query adding load to the network.

The negative effect of the filename only approach is also observed in the search for the resource in question by requiring a search in a space originally containing all N number of keys in the P2P network. This global search paradigm is expensive when the user is only interested in one of the N items in the P2P network.

In order to address these problems subject awareness is added to the P2P network to reduce

the N number of P2P items to a search with a number closer to the one item of interest. The grouping of resources by subject is one of the fundamental organization techniques and allows for a more efficient search and retrieval of the resource in question. A subject aware P2P network consists of information that allows for the organization of the data based on the topical description of the material as well as the title. This works similar to the way a library separates books into bookshelves based on the subject the book contains. The subject can also be divided into sub-subjects to increase the granularity of the results to the desired effect where one or approximately one item is returned for a query. The benefits of adding subject awareness will also decrease the *N* number items in the P2P network to only *n* number of items within the subject of interest. This reduction reduces the search and has the potential to reduce hash updates depending on implementation.

One possible format for this implementation is to append the subject to the beginning of the file name. For example, OS-Linux-FILENAME.EXT would allow the key to contain a main subject of OS subdivided by sub-subject Linux routing to FILENAME.EXT.

The lack of integration of subject awareness of the P2P network has lead to the added complexity of adding overlays to the structure and increasing the complexity of the overlay solution. It is noted that some subject related information is being exploited by some P2P applications using the overlay available. For example a search may be limited to extension contained in the filename (filename.mp3) if the available. However, this is a function of the application, where as this proposal is to modify the structural properties of the P2P network to better serve the requests. The application solution simply filters the responses with no benefit to the search or the updates. This exploit also fails when the files are compressed or the extension is not available.

3.2. Expanded location aware P2P

A network aware P2P is one which takes into account the underlying network topography. A network aware P2P has shown to have great benefits for all entities involved in the P2P network. The efficiency of the resource sharing is increased for the users and the content distributers adhering to the primary goal of the P2P network. In addition, the use of locality and caching has advantages for the ISP and networks on which the P2P is operating. Due to the inherit benefit, we propose adding the locality to the structure of the P2P itself. One possible format for this implementation is to append the location to the beginning of the file name similar to the subject. For example, ExternalIP-InternalIP-FILENAME.EXT would allow the key to contain the external IP for the host network or proxy and the internal IP of machine within the local network.

Addition subdivision may also be needed depending on the size of the network involved to obtain adequate results. As with the subject, the location can also be divided into sub-locations to increase the granularity of the results to the desired effects.

This approach can easily be combined with the subject aware example to form OS-Linux-ExternalIP-InternalIP-FILENAME.EXT. Notice the subject is the parent category of the location to apply the appropriate hierarchy.

Adding the local awareness to the subject provides additional benefits to the search by decreasing the number of items even further and providing the hierarchy to display the most efficient transfer.

The gathered information combined format is very important as it creates the structure to increase the efficiency of the P2P. It is possible to add user defined subjects and locations, but this should be limited to super-peers or discouraged all together. The reason for the limitation pertains to the efficiency of the network revolving around the subject and

location granularity. If the subject and location information is too fine the search resolves to only one file when multiple exist.

3.3. User aware P2P

Typical users are not interested in every file in the P2P network; they are only interested in the files that are the subject of interest. Conversely the files shared by the user will also be limited in subject based on their interest and the files transferred to them in the sharing process. Due to the relationship between the user and the information transferred and shared the P2P network should be aware of this relationship.

As in the previous two Sections we could encode the relationship in the filename, which in turn will become the hash key to provide a mechanism for these relationships. The main differences between the previous two awareness models and user awareness are the dynamic nature and the diversity of the interests of the user. The user's interest will also evolve over time and some will expire.

The model of this evolution from a network perspective begins with the user having a blank slate of interests. As the user searches, transfers and shares files the slate can be filed with the subjects of interests taken from Sec. 3.1. In addition, like minded people can be found from the network IP relation to the other users in the network. This information can be used to build subsections of the P2P network that are amiable to the user's interest by joining the user to the like minded peers. The original network with no information gather would have the same efficiency of the current methodologies.

4. Gathering and sharing information

Section 3 outlines the information and awareness to the P2P. In order to make the awareness possible additional information must be gathered and shared with the appropriate peers using a decision support system. The first action a peer takes is joining the P2P network. As part of this action, information is gathered via a survey of the users of the users interest in the subject. For example, the user may be interested in operating systems (OS). The information gathered will add the peer in a location within the OS subject area. The structure is maintained when the user transfers the OS files without requiring the re-hash. If the user is interested in multiple areas this same mechanism can be used by adding the peer to multiple locations within the network.

In the second step the user searches for the resource desired and selects the file to transfer. This step is no different than how the user interacts with the network except that the information will be used to define the peer's interest. The survey, the search and the transfer actions are added to the users profile for the P2P. The survey, search, and transfer are assigned different weights in the profile. Generally speaking the survey has the highest weight as it most accurately defines the user's interest. The search has weight as the terms chosen represent the user's desire in the subject area. For example, Ubuntu versus Redhat represent two different interests in the Linux OS. The transfer has weight, but at this point, only for the description of the resource until the user actually reviews the retrieved information.

Offline of the P2P the user accesses the resource that was retrieved and at this point will be aware of the quality and accuracy of the downloaded material at this point. In order to report this information back to the P2P community a post-download survey is utilized to rate the resource. The user is only to rate the correspondence between the property of the resource as described by the poster and the resource as viewed by the recipient peer.

No-reviews returned for a download are viewed as positive by the user as people generally have little interest in responding positively. If the no-reviews were weighted as

neutral or negative, the end result would be too much negative weight.

The equation defining the subject interest is:

$$S_1 = w_{e1} + w_{q1} + w_{p1},$$

$$S_2 = w_{e2} + w_{q2} + w_{p2},$$

$$S_n = w_{en} + w_{qn} + w_{pn}.$$

S_n represents the subjects of interest by the user. w_e represents the weight of the entry survey. w_q represents the weight of the queries for the searches within the subject. w_p represents the weight of the post download survey. The sum on the subjects S_{1_n}, and weights w_{e1_n}, w_{q1_n} and w_{e1_n} equal 100%. This maintains that an increase in a particular subject will result in a decrease in other subjects. This summation gives the subjects the ability to expire as the interest decreases below a threshold.

Once the information is gathered it needs to be shared in order to be of use. A simple measure is to use and information file similar to the approach used in newsgroup file sharing. The info file contains all the pertinent information to the corresponding resource. An easy way to identify this file is by using the .nfo extension to the resource name with the appended information. The internal contents should format in an XML format to provided easy extensibility and parsing. The .nfo file should be available for view along with the associated resource to allow the peers searching for the resource to analyze the pertinent details about the resource before downloading.

All information added by the return surveys should be appended to info file. Modification should not be allowed as this provides a mechanism for the unscrupulous to corrupt the process. The .nfo file can be parsed and ratings and comments combined to provide a quick look for those interested in the contents.

Malicious and corrupt files and files not of adequate content or quality can be removed from the network via this peer review mechanism. If a set number of peers denote the file as bad, the filename is appended with “garbage-” along with any copies of the file distributed in the network. The number of complaints from peers is up to the demand of the community. Although it should be set high enough to discourage false reports and reports limited to only those which actually retrieved the file. Too many complaints on files being shared by a particular peer would result in the removal of the peer.

5. Organization using intelligent DHT

The information gathered from the peers and the network allow for the DHT to be segmented based on the information gathered. The segmentation and hierarchy created is displayed in Fig. 1. The key containing the appended subject and location information organizes the resources within the DHT in a manner where the overall DHT can be segmented and recombined easily. This ability is used to further enhance the efficiency of the overall DHT by only linking the peers interested in the segmented sections and creating a smaller DHT then the original. This linking is accomplished by using the weight function from Sec. 4 and the reduction of the overall P2P network to a smaller DHT designed around the interest of the peer joining and participating in the network.

Figure 2 displays this concept with three subjects and two peers interested in subsets of the overall DHT. Peer 1 is only interested in subjects 1 and 3, so combining those subjects and excluding Subject 2 the DHT is reduced based on the user’s interest to the solid line. The same holds for the user only interest in Subject 1 and 2 represented by the inner most dashed line.

The effect of the reduction of the DHTs is felt throughout the P2P network. The search is reduced to only the areas of interest, even for a “global search”. The updates within the P2P network are reduced to only the subset of the overall DHT based on the peer’s interest. This segmentation allows for the subjects to be re-

hashed and reorganized without having an effect on the entire DHT. The basic form of the scheme requires the hash function maintain locality of the keys similar to the alpha representation of the original topic. This can be accomplished by hashing the subject, location, filename, and extensions separately and recombining the results. It is also possible to leave these values in their raw form or compressed using a simple compression scheme to shrink the length while maintaining the relative location.

Expanding the searches that return negative results is accomplished by expanding the groups of peers that have similar interests. Figure 3 displays four groups with collective interest in four different subjects S1-S4. The connections are the common interest between the groups based on subject. The connection builds an ontology of user interest based on the segmentation of the DHT based on the user's decisions. The ontology built allows for a more efficient expansion of the search by looking in the most likely place while maintaining the exclusion of unwanted material.

For example, a search by a user is first executed in the subject chosen in the survey; it is then expanded to the users other subjects of interest. The final step is the search of the subject connected directly to the user's interest. Multiple connections will be chosen based on weight of interest given in Sec. 4.

Expanded searches can also be used to indicate the strength of the subject classification defined. A large number of expanded searches followed by positive download would indicate the subject classification needs to be expanded to include the other subject as the subject class may be too small. Very few expanded searches may indicate the subject is too large and may need to be sub-divided to promote efficiency.

It is noted that files not assigned a subject are placed in a default segment of the DHT. This insures that obscure files still have a place within the DHT while maintaining the segmentation by subject. If all files are not segmented the resulting network would be the complete P2P network as currently defined.

6. BitTorrent and file availability.

BitTorrent and enhanced transfers can be added by the addition of a file extension to identify copies of the original through the network. The collection of copies can serve multiple purposes for the end user. Typically, the user would choose the best file to retrieve based on bandwidth of the peer or the location in the network in order to achieve the most efficient transfer. In addition, the file copies can be used to provide basis for BitTorrent without the need for a tracker. The search provides the ability to centralize the file much in the same way as the torrent file would in standard BitTorent.

The lack of copies and the awareness of the user's interest can be used to identify files of high priority for distribution to users of that subject area. This knowledge should increase the transfer of the file and increase the availability. An automatic distribution can also be implemented to transfer the files to the peers interested in the subject using location awareness to distribute the resource for the benefit of the P2P network. An "I am feeling lucky" mechanism, where the user allows files to be force fed to the peer based on the profile created in the P2P for past history, is possible to allow downloading without a search.

Failure to find a file of interest can also be addressed using the proposed framework. When a file is not located, a request can be initiated to the peers within the subject area. The request can be rated by interest and responded to via a post of the requested resource to the P2P community.

7. Conclusion

The proposed framework collects knowledge from the three primary components of the P2P network: the user, the files and the interconnect network. The knowledge is gathered from the peers during the process of joining the P2P, transferring files and rating the downloaded files using a decision support system consisting of data collection mechanism and multiple surveys. The knowledge gained represents the awareness of the P2P network in regards to the file subject and location as well as the users of the network. This awareness allows the files to be grouped by subject and the location within the DHT providing a structure of the intelligent P2P network. The knowledge of the files is shared with the rest of the peers using an .nfo file, which allows for easy surveying of the material and the historical information. The info is not modifiable to prevent malicious activity. The communication of information allows the peers to administer the network on a distributed basis and rate the files and the peers in the network. The rating system provides the ability for the peers to make decisions on the files to transfer and the ability to remove malicious files and peers from the network.

User awareness is provided by the weight of the knowledge gathered and using the results to define the user's interests. The intelligence mechanism uses this knowledge to create sub-networks within the P2P network by combining only subject segments that adhere to the user's interests. The segmentation of the three components allows for more efficient search, update mechanisms, and file transfers by reducing the area for each peer activities.

Additional benefits of file availability and fast transfers may also be implemented by using awareness gathered to provide BitTorent ability without a tracker and using the user's interest to pre-deploy the files.

The use of a specific P2P protocol is not mentioned as this framework is meant to apply to all protocols using DHT. Although it is noted that certain implementations will be easier than others. For Example, P2P protocols with a peer hierarchy, such as Gnutella, are structured in a way that makes the peer segmentation easier to implement. Other P2P protocol may need to be modified or layered to produce such a hierarchy.

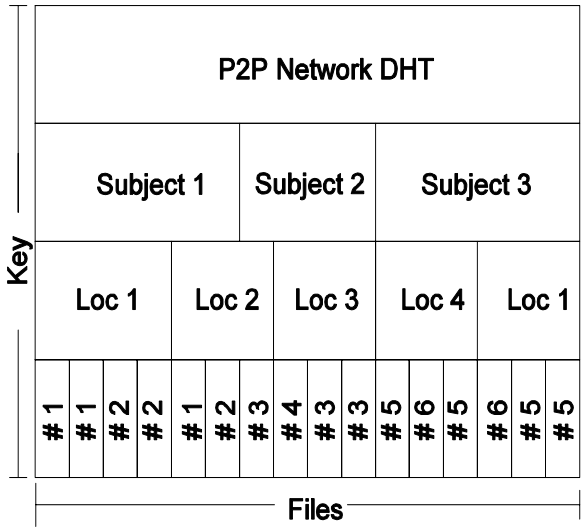


Figure 1. Segmented DHT based on subject and location of files.

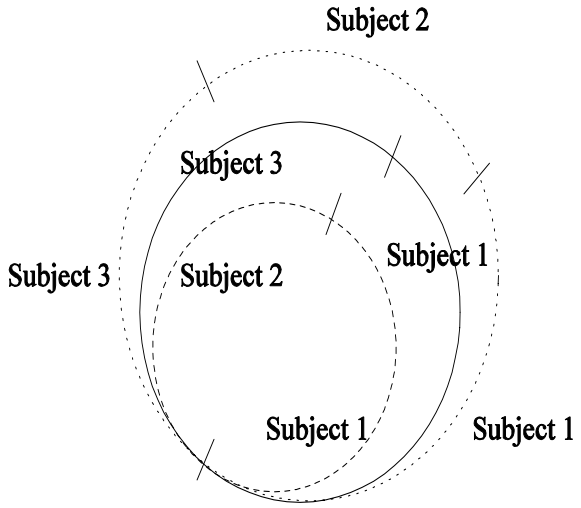


Figure 2. A view of segmented DHT. The dotted line represents the complete DHT. The solid line represents the view of peers interested in Subject 1 and 3. The dash line represents the view of peers interested in Subject 1 and 2.

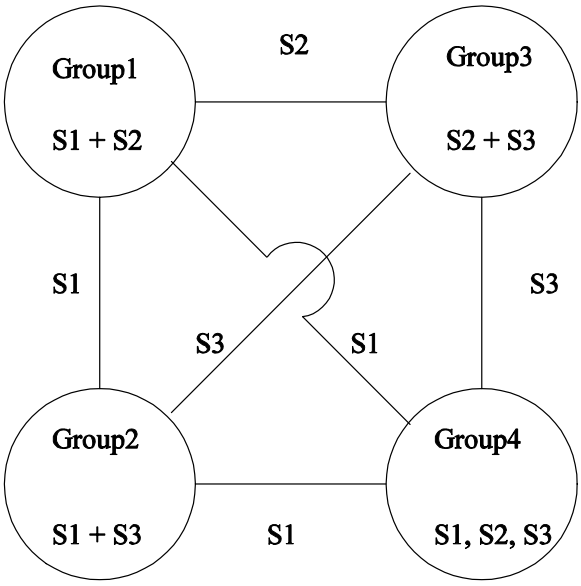


Figure 3. Four group's interest in subjects S1-S3. The interconnect shows the relations between the group based on the subject used to create the ontology of user's interest. The relations are utilized to expand the search area on failure.

8. References

[1] S. Ratnasamy, P. Francis, M. Handley, R. Karp, and S. Shenker, "A scalable content-addressable network", SIGCOMM, ACM , San Diego, CA, Aug. 2001, pp. 161-172.

[2] I. Stoica, R. Morris, D. Karger, M. F. Kaashoek, and H. Balakrishnan, "Chord: A scalable peer-to-peer lookup service for Internet applications", SIGCOMM, ACM, San Diego, CA, Aug. 2001, pp. 149-160.

[3] A. Rowstron and P. Druschel, "Pastry: Scalable, distributed object location and routing for large-scale peer-to-peer systems," IFIP/ACM International Conference on Distributed Systems Platforms (Middleware), ACM, Heidelberg, Germany, Nov. 2001.

[4] B. Y. Zhao, L. Huang, J. Stribling, S. C. Rhea, A. D. Joseph, and J. D. Kubiatowicz, "Tapestry: A resilient global-scale overlay for service deployment," IEEE Journal on Selected Areas in Communications, IEEE, Jan. 2004, vol. 22, no. 1, pp. 41–53.

[5] P. Maymounkov and D. Mazieres, "Kademlia: A peer-to-peer information system based on the xor metric", Proceedings of International Workshop on Peer-to-Peer Systems, Cambridge, MA, Mar. 2002.

[6] M. Izal, G. Urvoy-Keller, E. W. Biersack, P. Felber, A. Al Hamra, and L. Garces-Erice, "Dissecting bittorrent: Five months in a torrent's lifetime", Passive and Active Measurements (PAM), 2004.

[7] M. Ripeanu. "Peer-to-peer architecture case study: Gnutella network. Technical report", University of Chicago, 2001.

[8] M. Bawa, G. S. Manku, and P. Raghavan, "SETS: search enhanced by topic segmentation", SIGIR, ACM, Toronto, Canada, 2003, pp. 306-313.

[9] Yuh-Jzer Joung, Feng-Yuan Chuang, "OntoZilla: An ontology-based, semi-structured, and evolutionary peer-to-peer network for information systems and services", ACM International Conference Proceeding Series, ACM, Jan. 2009, Vol. 25, Iss. 1, Pages 53-63.

[10] K. Aberer1, P. Cudre-Mauroux1, M. Hauswirth1, and T. van Pelt2, "GridVine: Building Internet-Scale Semantic Overlay Networks", ISWC, Hiroshima, Japan, Nov. 2004.

[11] T. Karagiannis, P. Rodriguez, and K. Papagiannaki, "Should Internet service providers fear peer-assisted content distribution?", Proceedings of Internet Measurement Conference, Berkeley, CA, Oct. 2005.

The Journal of Management and Engineering Integration
Vol. 3, No. 1, Summer, 2010

A Simulation Study for a Modified Sudden Death Test

Erik E. Kostandyan
Western Michigan University
erik.kostandyan@gmail.com

Abstract

Total Accumulated Time is defined as total time needed to produce a selected number of failures. So it is a length of time during which the test equipment was used and utilized. The *total accumulated time* and a *cost of test equipment per unit testing time of operation* will result in the cost of the utilization of testing facility. In this paper a simulation study of total accumulated time for a Modified Sudden Death Test strategy is developed, under the assumption that the failure distribution of tested components is Weibull with a shape parameter greater than or equal to unity. The theoretical distribution of Total Accumulated Time is derived for a special case where a shape parameter in Weibull distribution equals to unity. The optimum testing strategy from Total Accumulated Time perspective is found under Modified Sudden Death Test strategy.

Index Terms— Total Accumulated Time, Type II Reliability Test, Failure Truncated Test, Modified Sudden Death Test, Weibull distribution.

Notations

T_{ac}	Total Accumulated Time
T_{ac_j}	Total Accumulated Time in an arbitrary j^{th} test
C	Cost of test equipment per unit testing time of operation
n	Number of available testing facilities
r	Number of results collected per group
k	Number of groups
R	Required number of results from the test
N	Total components used during the test
$t_{r:n}$	The rth failure-time out of n components
$t_{r:n,k}$	The rth failure-time out of n components in the kth group
β	Weibull distribution shape parameter
θ	Weibull distribution scale parameter
χ_a^2	Chi-square distribution with an a degrees of freedom

1. Introduction

Testing mechanical components for their intended purpose under predetermined working conditions is a common practice used by industries to prevent failures and determine reliability of products. Failure tests are performed in specialized testing laboratories, where the capacity of the test stations is limited. This is a major restriction for reliability tests and a reason splitting components into the groups in a way that test stations would utilized to their maximum capacity.

Reliability tests are categorized as Time Truncated or Failure Truncated, known in the literature as Type I and Type II tests, respectively. In Failure Truncated tests, the mechanical components are tested until the desired number of results is obtained. Arbitrary testing strategy for a Failure Truncated test might be presented by Modified Sudden Death Test. A *Modified Sudden Death Test (MSDT)* is a failure (Type II) test, where the components are divided into groups (k), and each group is

tested without replacement until a predetermined number of failures occur (r). Once there is the predetermined number of failures in a group, the test is terminated for that group and the next group can be tested. The predetermined number of failures in each group is the same. In the MSDT the number of failures in each group may be less than or equal to the number of components in the group. If r is the number of failures in the group and n is the number of components in the group, then $r \leq n$ holds.

Endeavors aimed to collect the desired number of results during a test *(R)* costs money. So, the expenditures for a reliability test might be categorized as costs associated with number of components used during a test *(N),* technician cost that supervised the test and a cost associated with test equipment utilization. The test equipment utilization cost is a product of *Total accumulated time* (T_{ac}) and *Cost of test equipment per unit testing time of operation (C).* This means that as higher T_{ac} so expensive a reliability test.

This paper investigates the behavior of T_{ac} in Modified Sudden Death Test due to changes in a shape parameter from a Weibull distribution and a number of possible grouping combinations in a Modified Sudden Death Test. Also, the distribution of T_{ac} is theoretically derived for a unit shape parameter from a Weibull distribution.

2. Background Work

Pascual and Meeker [9] compared sample sizes and the corresponding cumulative testing time estimators for a technique, which they termed Modified Sudden Death Test (MSDT). In this work the quantiles of T_{ac} were approximated by Cornish - Fisher expansion and it was mentioned that the distribution of T_{ac} did not have a simple form. Freudenthal and Gumbel [3] have discussed fatigue phenomenon and the failure of mechanical components. They showed that a failure of mechanical components has a Weibull [11] distribution. Currently, this work is a basis for many research manuscripts that investigate a failure of mechanical components and assume a Weibull failure distribution for them. Based on Vlcek, Hendricks and Zaretsky [10] simulation study, it was assumed that 50 mm deep-groove ball bearing life has a two-parameter Weibull distribution, with a shape parameter equal to 1.11 and scale parameter of 47,227 hrs. The authors stated that results from past studies had shown that the most bearings with rolling elements have a slope parameter between one and two. Also, Vlcek, Hendricks and Zaretsky [10] documented that the number of available test facilities would be between two and eight. It worth to mention that Maximum Likelihood Estimation (MLE) is a widely used technique for estimating the Weibull distribution parameters. This technique utilizes some collected data from the test to estimate the unknown parameters. It is well known that MLE is asymptotically unbiased and asymptotically efficient. Cohen [1] discussed and showed shape and scale parameter MLEs from complete and Type II censored data for the Weibull distribution. Also, he mentioned that estimates from the medium sample sizes (n= 30 to 32) deviate by 10% from the true parameters. However, in the literature there are many works aimed for reduction both *Biased* and *Mean Square Error* (Hirose [4] and Cousineau [2]).

3. Formulation of Total Accumulated Time in MSDT

If n components are tested simultaneously until the rth failure, then the accumulated time for the test is a summation of the all “r” failures plus “n-r” survived components testing times.

Let $t_{r:n}$ be the rth failure time for an arbitrary j group of n components, then the accumulated time for an arbitrary j group is computed by (1).

$$T_{ac\_j} = \sum_{i=1}^{r} t_{i:n} + (n-r) * t_{r:n} \qquad (1)$$

If k groups of n components are tested serially until the rth failure in each group, then the total accumulated time for the test is the sum of the accumulated times in each group.

Let $t_{r:n,k}$ be the rth failure time out of n components in the kth group, then the total accumulated time for the test is computed by (2).

$$T_{ac} = \sum_{j=1}^{k} T_{ac\_j} = \sum_{j=1}^{k} \left(\sum_{i=1}^{r} t_{i:n,j} + (n-r) * t_{r:n,j} \right) \qquad (2)$$

4. Simulation Study

In this section a Monte Carlo Simulation study is performed for a Type II reliability test to study T_{ac} behavior. Some assumptions were maid and justified by the summary in background work section. It is assumed that a component under the investigation is a 50 mm deep-groove ball bearing. It has a Weibull failure distribution with scale parameter of 47,227 hrs ($\theta = 47{,}227$) and shape parameter that might vary from one to two ($1 \le \beta \le 2$). The number of available testing facilities is six ($n = 6$) that is utilized by 100%, and the desired number of results from the test is thirty ($R = 30$). By summarizing these assumptions for the sake of Monte Carlo simulation the following might be stated:

- Number of testing facilities available is six ($n = 6$)
- Required number of results from the test is thirty ($R = 30$)
- Failure distribution is Weibull
- Scale parameter is 47,227 ($\theta = 47{,}227$)
- Shape parameter varies from 1 to 2 with increments of 0.2 ($\beta = \{1 : 0.2 : 2\}$)
- Number of replications for each MSDT plan is 10,000

Based on the six available testing facilities and the desired thirty results from test, the following five MSDT strategies are possible to construct (*see Table 1*).

Table 1: All the possible configurations of MSDT strategies

MSDT strategies	*Groups (k)*	*Results / group (r)*	*Component / group (n)*	*Total results (R)*	*Total components (N)*
1	5	6	6	30	30
2	6	5	6	30	36
3	10	3	6	30	60
4	15	2	6	30	90
5	30	1	6	30	180

In order to distinguish the strategies from each other, the *MSDT(k,r,n)* terminology will be used. For example, the strategy number three will be termed *MSDT(k=10,r=3,n=6).* This means that ten groups of six components are tested until the third failure in each group. Six components in a first group was tested until the third failure, after that the second group of six components was tested until the third failure and this procedure was continued until the tenths group of six to be tested till third failure. Test is in serial mode between groups, but in parallel mode within a group. A computer algorithm is developed, where each *MSDT(k,r,n)* strategy is replicated 10,000 times. Based on these 10,000 results, a histogram was constructed for each *MSDT(k,r,n)* strategy. The results of simulation study are depicted in Figure 1.

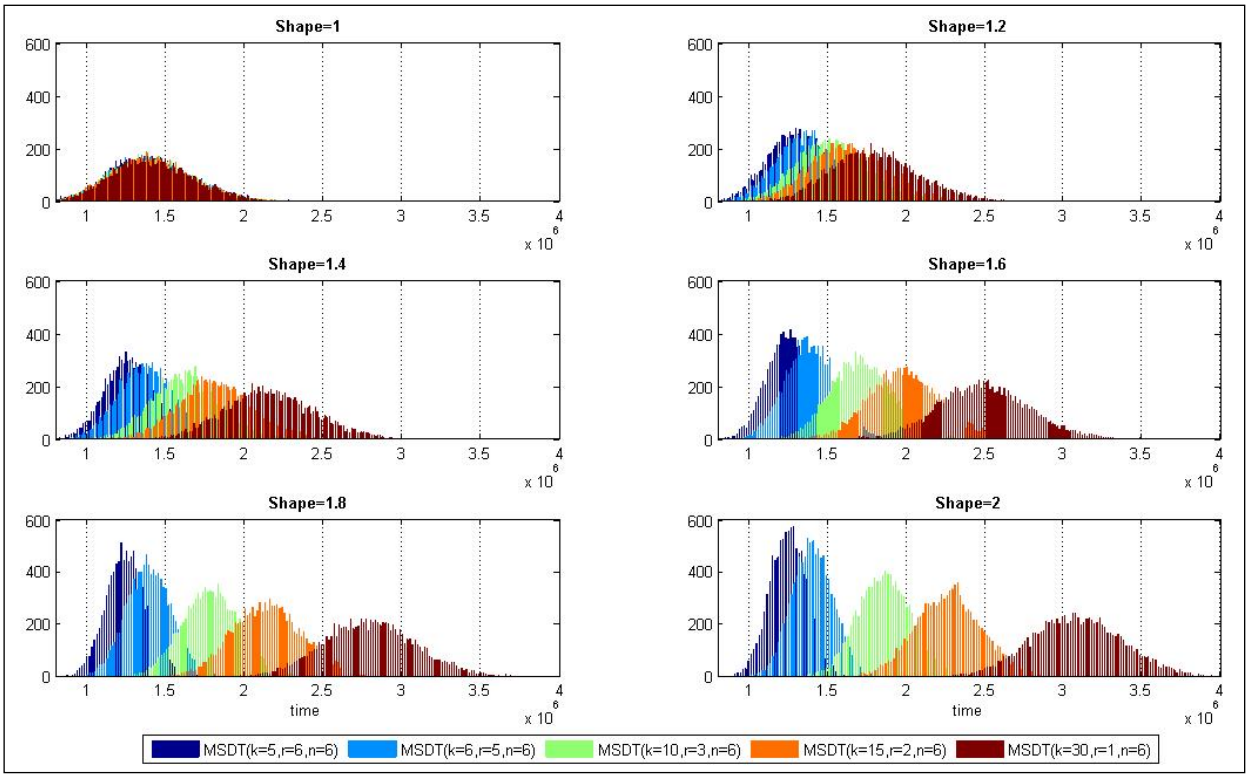


Figure 1: Histograms of the Total Accumulated Time in all possible MSDT configurations

5. Results and Discussion

For the fixed shape parameter ($\beta > 1$) that greater than unity (*see Figure 1*), an increase in number of groups *(k)* such that a total collected results from test *(R)* is the same have an consequence to increase location and spread parameters of T_{ac}. Also it is noticed, that an increase of shape parameter affects on location and spread parameters of T_{ac} to increase. It means that from T_{ac} perspective, MSDT with the possible smallest number of groups is superior to the all possible MSDT configurations when shape parameter greater than unity. Moreover, by increasing number of groups in a such way that to keep a total collected results from a test unchanged, increases a total number of required components *(N)* for a test (*see Table 1*).

For a particular case, when shape parameter equals to unity, an increase in number of groups does not affect on location and spread parameters of T_{ac}. This phenomenon is discussed and theoretical distribution of T_{ac} is derived when shape parameter is one ($\beta = 1$).

6. Distribution of Total Accumulated Time when shape parameter equals to unity

Claim: $T_{ac} \sim (\theta/2)\chi^2_{2R}$ or $2*T_{ac}/\theta$ has a chi-square distribution with $2R$ degrees of freedom, when shape parameter equals to unity.

Proof: If n components are tested simultaneously until the rth failure and failure distribution is a Weibull distribution with $\beta = 1$ and some θ, then $2*T_{ac\_j}/\theta$ has a chi-square distribution with $2r$ degrees of freedom (Lawless [6], Meeker and Escobar [7], Lamberson and Kapur [5]). T_{ac_j} is a total accumulated time in an arbitrary j^{th} test, which is stopped at the rth failure out of n tested components. If one repeats this test k times, then a total accumulated time for the all k repetitions would be a sum of total accumulated times in each test, see (2).

It is noticed that for $j \neq g$ any T_{ac_j} and T_{ac_g} are independent random variables. Due to the fact that every $2*T_{ac\_j}/\theta$ has a chi-square distribution with $2r$ degrees of freedom, it follows that $2*T_{ac}/\theta$ distributed as a chi-square random variable with $2rk$ or $2R$ degrees of freedom. So, it follows that $T_{ac} \sim (\theta/2)\chi^2_{2R}$ and T_{ac} in any MSDT strategy is the same random variable.

7. Conclusion

It is under our interest to run the cheapest test. Shorter *total accumulated time* would imply less costly test. Based on this simulation study results, desired testing strategy from the total accumulated time perspective is a strategy where the number of groups is as small as possible, under the condition that shape parameter is more than or equal to unity. This also will result in useing fewer components during the test and will save money from total components perspective. The theoretical distribution for T_{ac} is derived for a special case where shape parameter equals to unity. If a shape parameter equals to unity then T_{ac} depends on both the scale parameter (θ) and the number of desired results from the test (R), yet it does not depend on the number of groups. This will help a practitioner to calculate *total accumulated time* with a desired confidence and make an estimate of expenses in test equipment utilization.

8. Acknowledgment

The work is partially funded by the U.S. Army TARDEC through the Center for Advanced Vehicle Design and Simulation (CAViDS) at Western Michigan University.

9. References

[1] A. C. Cohen, "Maximum likelihood estimation in the Weibull distribution based on complete and on censored samples", Technometrics, 7(4), 1995, pp. 579-588.

[2] D. Cousineau, "Nearly unbiased estimators for the three-parameter Weibull distribution with greater efficiency than the iterative likelihood method", The British Journal of Mathematical Statistical Psychology, 62(1), 2009, pp. 167.

[3] A. M. Freudenthal, and E. J. Gumbel, "On the statistical interpretation of fatigue test", Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences, 216(1126), 1953, pp. 309.

[4] H. Hirose, "Bias correction for the maximum likelihood estimates in the two-parameter Weibull distribution", IEEE Transactions on Dielectrics and Electrical Insulation, 6(1), 1999, pp. 66-68.

[5] Lamberson L. and K. Kapur, Reliability in engineering design, Wiley, New York, 1977.

[6] Lawless J., Statistical models and methods for lifetime data, Wiley New, York, 1982.

[7] Meeker W. and L. Escobar, Statistical methods for reliability data, Wiley, New York, 1998.

[8] Nelson W., Applied life data analysis, John Wiley Sons Inc., Hoboken, 1982.

[9] F. Pascual and W. Meeker, "The modified sudden death test: Planning life tests with a limited number of test positions", Journal of Testing and Evaluation, 26(5), 1998, pp. 434.

[10] B. Vlcek, "Monte carlo simulation of sudden death bearing testing", Tribology Transactions, 47(2), 2004, pp. 188.

[11] W. Weibull, "A statistical distribution function of wide applicability", Journal of Applied Mechanics, 18(3), 1951, pp. 293.

Biography

Erik E. Kostandyan completed his Ph.D. program in Industrial Engineering at Western Michigan University. Currently he is in a PhD program in the Department of Statistics at Western Michigan University. He has a Master's Degree in Industrial Engineering and Systems Management from American University of Armenia. His areas of research are reliability test configurations, statistical models in reliability, improved reliability test designs, systems simulation and modeling. Since 2006 he has been studying different set-ups' for reliability tests. The title of his dissertation was *"Optimum Failure Truncated Testing Strategies."* Results from this research have been accepted in conferences and peer review journals.

Improving the Accuracy of a Mobile Robot Using Factorial Analysis

Tamer A. Mohamed
The British University in Egypt
tamer.mohamed@bue.edu.eg

Ayman S. Abbas
The British University in Egypt
aabbas@bue.edu.eg

Abstract

Mobile robots are increasing in popularity in the industrial applications due to the increased precision versus usability. The most common of these types are path following and wall tracking at a pre-defined distance. This paper studies the influential factors that can affect the accuracy of the Mobile robot in tracking a wall. Doing so, these factors can be controlled to improve the robot accuracy.

Three controllable factors are used with two response factors. The controllable factors are robot speed; platform inclination; and the friction between the robot wheels and platform material. The robot speed was analyzed at three levels (300, 400, and 500 cm/min), the platform inclination is kept at 0, 1.6, and 3.2 degrees upward. Increasing the platform inclination angle over 3.2 degrees makes it hard for the robot to withstand its weight and climb on the platform. Finally the platform material ranges from high resistance material (vinyl), medium resistance material (abrasive sheets) and low resistance material (wood). The material of the robot wheels is not changed throughout the whole experiment.

The response factors that determine the accuracy of the robot motion are the mean deviation from the target value and the motion standard deviation. Estimation of the response factors as a function of the controllable factors can be used in future work to enhance the accuracy of the robot motion by feeding the error function into the robot motion control programme to improve its performance.

1. Robot Accuracy

The aim of this paper is to improve the accuracy of mobile robots operating in an industrial setting which are affected by influential factors. Several factors have been identified which have affected the accuracy of industrial robots. These are; the tolerance of robot components, elasticity of components and joints, encoder resolutions, and control system limitations resulting from the inaccuracies between actual kinematic relationships and the definition of those relationships in the controller. These are the four root causes of the inaccuracies in the mobile robot [9].

The study of robot accuracy in general and industrial manipulators in particular are extensive. Junchuan et al. [4] have presented an extensive review of position accuracy of industrial robots. Their study also advocates the assessment of the overall accuracy of the robot system rather than the accuracy of individual components. An approach to improve the overall robot accuracy is presented based on a revised Denavit-Hartenberg kinematics model which takes into account the small variations in robot joints. In addition, a verification of the overall accuracy

of a surgical robot used in Computed Tomography (CT) scan has been presented by Morgan [8]. The method presented emphasizes that measurement of the positional error must be more precise than that of the application error. Another approach used to detect robot position errors during operation is to utilize radar motor detection techniques [10].

Several methods for improving industrial robot accuracies have been suggested in the literature. For example, one particular approach is to mimic the position error with analytical functions. This method approximates the expected error positions with a class of polynomial functions of joint angles [1]. A method for prediction of geometric errors by utilizing particle swarm optimization methods has been presented. Another approach which has been suggested by Martin-Gorostiza [6] is to use a coverage mapping method to aid positioning of the robot in space thus reducing the associated error. A different approach taken by Lightcap et al. [5] is to use a two level non-linear optimization algorithm which utilizes measurements obtained from a coordinate measuring machine. This is in converse to the early work by Borm et al. [2] which does not require the use of a position measuring machine. The approach rather relies on using a nominal inverse kinematics algorithm.

A more novel approach is to circumnavigate the accuracy issue of robots by utilising a fine positioning module which operates at the nano scale. This module is built up using piezo axes with very minute resolutions. The approach relies on convergence of the robot normal axes with the axes of the fine position module [11].

The review of the above literature that involves the study of the inaccuracies in industrial robots has identified an approach which is based on a holistic manner. This approach tackles the problem of robot accuracy by analysis of the root causes of inaccuracies in a controlled study in which factorial analysis techniques are employed. The reason for this approach is to identify the factors which affect most the overall accuracy of an industrial robot. The case of wall tracking is taken as a case study for evaluating this approach.

Three controllable factors representing real factory setting facets have been identified in this paper. These are the mobile robot speed, the inclination angle of the operating plane and the level of friction between the robot wheels and the ground. The objective of this paper is to:

- Examine the positional accuracy of a mobile robot while tracking a vertical wall
- Develop a control architecture to measure the accuracy of the mobile robot
- Design and implement an experimental setup to measure the accuracy
- Determine the influential factors that affect the mobile robot accuracy
- Reduce the robot inaccuracies using non linear optimization

2. Design of Control Architecture

In order to be able to assess the performance of the robot accuracy while tracking a vertical wall, suitable control architecture has been devised. The architecture starts by taking the actual measured distance from the infrared sensor to a tracked wall and compares it to a preselected reference distance. The robot is either moving away from the wall, or is moving closer to the wall. Accordingly, the direction of the robot in the lateral direction of motion is adjusted to move closer or move away from the tracked wall.

Simultaneously, the actual measured distance as compared to the reference value is plotted on an oscilloscope in real time. The developed control architecture is illustrated in Figure (1).

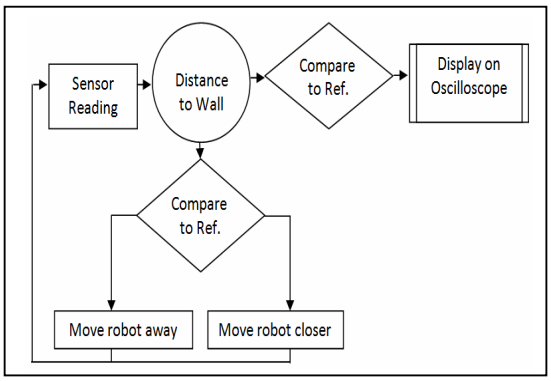


Figure 1. The Control Architecture

This control architecture is implemented in the Matlab Simulink environment. Voltage signal from the tracking infrared sensor is read from the robot in real-time. The measured sensor voltages are calibrated into distances measured in centimetres using the following interpolated equation based on calibration of the infrared sensor:

$$y = 10^{-5}x^4 - 0.0007x^3 + 0.019x^2 - 0.257x + 2.1076$$

Subsequently, the detected distance is compared to a pre-defined constant value which is the required distance to be maintained against the tracked wall. If this distance increases, the robot motors are directed towards maintaining their original approaching direction in order to bring the robot closer to the wall. If the detected distance decreases less than the pre-set value, the robot motors are directed towards reversing the lateral direction of the robot in order to return to the required constant distance from the wall. The predefined constant value is set at 8.4 centimetres (within the optimum operating range of the sensor) measured from the tip of the infrared sensor.

The robot lateral direction obtained from the previous step and constant speed together with the constant longitudinal direction and speed are the input to each of the three omniwheel drives that are responsible for realizing the required speeds and directions. Control signals from the omniwheel controller drives are the input to each of the three motors in order to accomplish the actual robot movements. As a safety feature of the control architecture, a signal from the collision bumper is used to activate all motor brakes when required.

Whereas the above control architecture implementation are responsible for controlling the wall following behaviour of the robot, an additional section is required to present a usable and clearly identifiable measure of the robot accuracy. Therefore, the actual detected distance obtained from the infrared sensor is also compared to a predefined reference value and plotted on a graphical oscilloscope dynamically in real-time. The predefined constant value is set at 8.91 centimetres which is measured from the tip of the robot. These plotted data are used in the factorial analysis elucidated latter in this paper.

The control architecture design (Figure 1) is then built in the Matlab Simulink environment to be used for the study to assess the robot accuracy. After developing this design in Matlab Simulink environment, a snap shot of the window was taken and illustrated in Figure (2).

3. Experimental Work

The experimental setup used to measure the location accuracy of the robot consists of a mobile robot moving on a plane while tracking a vertical wall seeking to maintain a suitable constant distance from it.

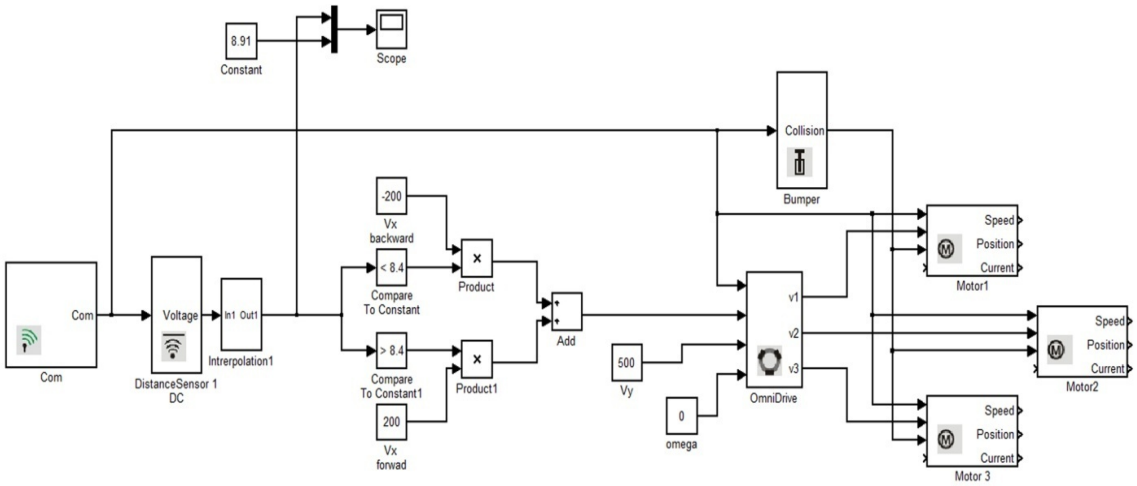


Figure 2. The Matlab Simulink Implementation

The selected robot is a commercially available wheeled cylindrical mobile robot (Robotino from Festo GmbH). It runs on three omniwheel drives which allow it to manoeuvre accurately and rapidly in response to commanded control signals. The robot is fitted with 9 infrared sensors which are used to detect the surrounding environment. One of these infrared sensors was used to track the vertical wall at a desired distance of 8.91 cm.

The Robotino robot is also fitted with a contact bumper sensor which detects collisions with surrounds such as obstacles and walls. Communication between the robot and the controlling programme is achieved through built-in WiFi support which allows remote control of the robot via a PC. The Robotino also comes supplied with an Application Programme Interface (API) which interfaces between the robot hardware and the Matlab Simulink programming environment.

The robot tracks the wall on a movable plane which can be inclined at any desired angle to measure the effect of the inclination angle on the accuracy of the wall following. Similarly, the surface of the plane on which the robot moves can be fitted with any material which can be used to evaluate the performance of the robot in any simulated condition of friction. The layout of the experimental setup is illustrated in Figure (3).

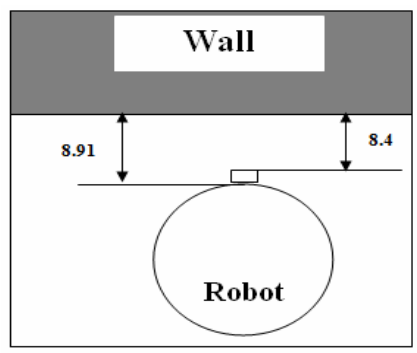


Figure 3. Experimental Setup

4. Results

The above experiment uses the general factorial design with three controllable factors and two response factors (3^3 factorial-experiment). The controllable factors (robot speed, platform inclination, and the friction between the robot wheels and the platform material) with their respective levels are shown in Table (1).

The Robot Speed is tested at three levels (300, 400, and 500 cm/min), the Platform Inclination is tested at three levels (0°, 1.6°, and 3.2°) and finally the Friction between the Robot Wheels and the Platform Material was tested at three levels (Wood, 0.1; Abrasive Sheets, 0.14; and Vinyl, 0.17). Each combination of the controllable factors is replicated 3 times to assess the value for experimental error.

The first response factor is the mean deviation from the target value (cm). Its data is obtained by taking the output of the oscilloscope and finding its mean and then subtracting this mean from the target value (8.91 in our case) to get the deviation. It may be worth saying that the deviation was taken as absolute value.

The same was done for the second response factors, the motion standard deviation (cm) but with calculating standard deviation instead of mean. These two data sets are analyzed to find the influential factors that affect the robot accuracy and find the optimum combination of these factors that will minimize both of the two response factors.

The design matrix and the corresponding values of the mean deviation from the target value and standard deviation of the variation from the target value are shown in Table (2).

The reported data in column 5 and 6 represent the mean deviation from the target value and motion standard deviation (response factors) at each combination of the controllable factor. The value represents the average of the three replications at this specific combination. The statistical analysis of the observed data is performed using the Design Expert Software [3]. Analysis of Variance is performed for the response factors. Experiments are prepared and performed in a randomized order. Analysis of the first response factor (mean deviation from the target value) is shown in Section 5, while Section 6 depicts the analysis of the second response factor (motion standard deviation). Section 7 explains the multi-objective optimization model. The paper ends with the conclusion.

5. Mean Deviation from the Target Value

Analysis of Variance (ANOVA) is used to study the effect of the controllable factor (speed, inclination, and friction) on the response factor (mean deviation from the target value) [7].

The results of the ANOVA are shown in Table (3). As can be seen, the model used is significant at a significance level of 0.0001 (p-value <0.0001) and the lack of fit is not significant (p-value = 0.2743). This indicates that no significant factors are excluded from the model and the model contains all the significant factors or their interactions, i.e. the model capture all the influential factors and their interactions that will affect the response factor.

The predicted model contains the speed, inclination, friction, the product of the inclination and friction, and the square of the friction. To be able to trust the below ANOVA results, model graphics should be visually tested for any violation to the ANOVA assumptions as shown in Figure (4) through Figure (6).

Table 1. Test Factors and Their Levels

Controllable Factors	Unit	Low Level	Medium Level	High Level
Robot Speed	Cm/min	300	400	500
Platform Inclination	Deg.	0°	1.6°	3.2°
Friction		Low Resistance (Wood, 0.1)	Medium Resistance (Abrasive Sheets, 0.14)	High Resistance (Vinyl, 0.17)

Table 2. Design Matrix and Observed Responses

Serial	Speed	Inclination	Friction	Response Variables	
				Mean Deviation (cm)	St.dev. (cm)
1	300	0	0.1	0.437667	0.5444
2	400	0	0.1	0.584333	0.755867
3	500	0	0.1	0.773	1.022167
4	300	1.6	0.1	0.340667	0.6225
5	400	1.6	0.1	0.492	0.991667
6	500	1.6	0.1	0.473	0.944333
7	300	3.2	0.1	0.213333	0.616
8	400	3.2	0.1	0.342333	0.784867
9	500	3.2	0.1	0.480333	0.929667
10	300	0	0.14	0.381667	0.647567
11	400	0	0.14	0.643667	0.866133
12	500	0	0.14	0.873333	0.904767
13	300	1.6	0.14	0.455667	0.578933
14	400	1.6	0.14	0.563667	0.744733
15	500	1.6	0.14	0.764	0.963
16	300	3.2	0.14	0.417667	0.461167
17	400	3.2	0.14	0.537	0.7702
18	500	3.2	0.14	0.712667	0.8865
19	300	0	0.17	0.410333	0.397733
20	400	0	0.17	0.504667	0.681467
21	500	0	0.17	0.468667	0.806067
22	300	1.6	0.17	0.396	0.3911
23	400	1.6	0.17	0.493	0.601533
24	500	1.6	0.17	0.669667	0.835033
25	300	3.2	0.17	0.371667	0.349233
26	400	3.2	0.17	0.464	0.557133
27	500	3.2	0.17	0.680667	0.912933

Table 3. Analysis of Variance for Mean Deviation From the Target Value

Source	SS	df	MS	F-Value	p-value
Model	0.79	5	0.16	39.5	< 0.0001
Speed	0.49	1	0.49	122.8	< 0.0001
Inclination	0.07	1	0.07	18.2	< 0.0001
Friction	0.05	1	0.05	12.8	0.0006
Inclination * Friction	0.11	1	0.11	26.1	< 0.0001
Friction^2	0.11	1	0.11	27.16	< 0.0001
Residual	0.30	75	0.004		
Lack of Fit	0.10	21	0.005	1.22	0.2743
Pure Error	0.20	54	0.004		
Total	1.10	80			

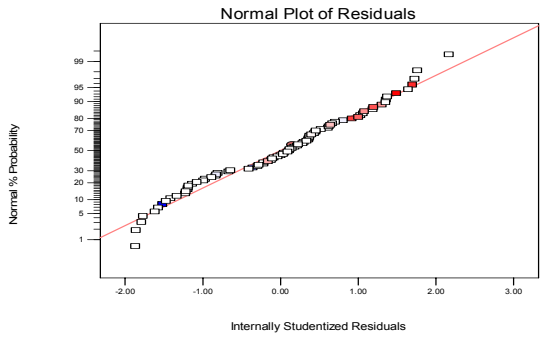


Figure 4. Normal Probability Plot for the Residual

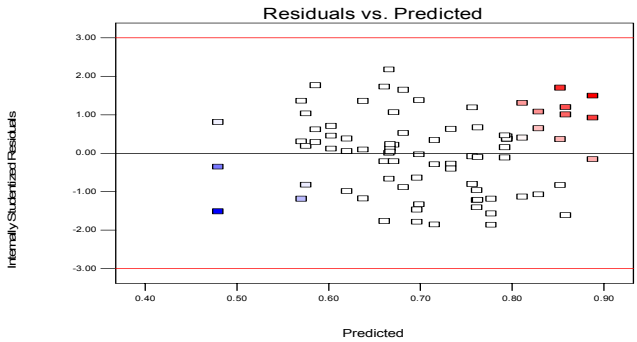


Figure 5. Studentized Residual Versus Predicted Values

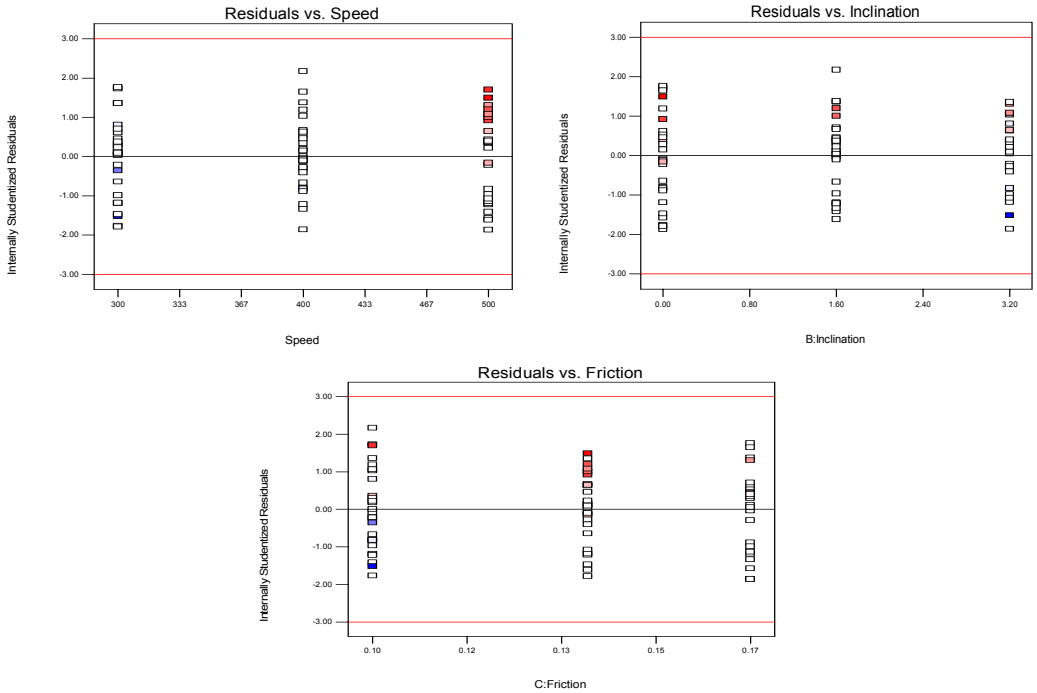


Figure 6. Studentized Residual Versus Each Controllable Factor

As per the normal probability plot, Figure(4), all the plotted points can be represented by a straight line. This means that the residuals follow normal distribution. A statistical test to assess normality of the residuals distribution is performed using a null hypothesis that the correlation coefficient of the data of the normal probability plot is equal to one. This presumes that the normal probability plot data fall on a straight line and thus the residuals follow normal distribution. The results of the test show that the null hypothesis couldn't be rejected at 5% significance level. This confirms the normal probability plot of the residual shown in Figure (4). Thus it can be concluded that the residual follow a normal distribution.

The plot of the residual versus its predicted value, Figure (5), and the plot of the residual versus each factor, Figure (6), show that the errors have constant variance and the data contain no outlier. Thus, the diagnostic examination for the residual revealed no violations of the ANOVA underlying assumptions, which mean that the ANOVA results can be trusted. Therefore, as indicated in Table (3), the fitted model is significant at a significance level equal 0.0001.

The results indicated that the speed, inclination, and the friction with their interactions have a significant effect on the mean deviation from the target value. Some higher order combinations of the controllable factors also have a significant effect on the mean deviation from the target value.

5.1 Mean Deviation from the target value – the fitted regression model

In this section, it is required to fit a mathematical model to the data of the experiment. The independent and dependent factors were fitted to the second order model

equation and examined in terms of goodness of fit.

The regression model that describes the mean deviation from the target value as a function of the robot speed, platform inclination, and the friction between the platform and the robot wheels is given as follows:

Sqrt(mean deviation from the target value) =

$$-0.98 + 9.6E\text{-}004*S - 0.17*I + 21.9*F + 1.1*I*F - 85.98*F^2 \rightarrow (1)$$

Where S, I, and F represents speed, inclination, and friction respectively. As can be seen from Eq. (1), the left hand side of the equation represents the square root of the mean deviation from the target value. This is due to the fact that during analysis of this response, the data required transformation to be able to fit it to a mathematical model.

Utilizing Eq.(1), the mean deviation from the target value can be estimated for any combination of the controllable factors (speed, inclination, and friction). It may be worth to say that the above equation should be used only if the speed is between 300 and 500 cm/min, the inclination between 0° and 3.2^o, and friction coefficient is between 0.1 and 0.17.

The model reports a coefficient of determination R^2 of 0.725 which implies that the model can explain 72.5% of the variation in the mean deviation from the target value. The values of the Adjusted R^2 and the adequate precision are 0.7062 and 23.639 respectively. Both values indicated that the model can be used to navigate the design space and that the model fitness is reliable.

5.2 Mean Deviation from the target value – response surface and contour plot

The mathematical model shown in Equation (1) is used to generate the Response Surface (RS) and the Contour Plot (CP) shown in Figure 7 for the analysis of the effect of the controllable factors (Friction, and Inclination) on the mean deviation from the targetvalue. These plots are performed keeping the speed on the average value of 400. However, changing the speed between 300 and 500 did not affect the shape of the response surface but its location.

As can be seen from Figure (7), an increase in the platform inclination and a decrease in the friction between the platform and the robot wheels led to a definite decrease in the mean deviation from the target value. Therefore, to minimize mean deviation from the target value, the friction should be kept at the minimum level while inclination should be kept at the highest level. Moreover, at the lower friction level (0.1) changing the inclination level from 0^0 to 3.2^0 decreases dramatically the mean deviation from the target value. This can be attributed to the resistance of the weight of the robot to its motion which gives the robot sensors chance to send and receive signal from the wall before continue moving.

5.3 Mean Deviation from the target value – model optimization

In this section, it is required to reach to the values where the controllable factors should be kept to minimize the mean deviation from the target value (the first response factor). The optimization models to minimize the mean deviation from the target value runs as follows:

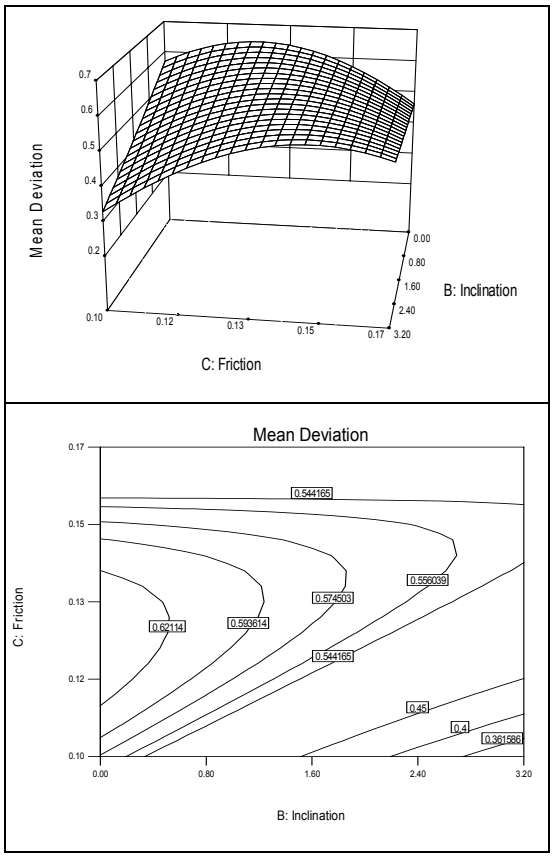


Figure 7. RS and the CP for the Mean Deviation from the Target Value

Minimize: Mean deviation from the target value

Subject to: $300 \leq \text{Speed} \leq 500$

$0.1 \leq \text{Friction} \leq 0.17$

$0^0 \leq \text{Platform Inclination} \leq 3.2^0$

Non-linear optimization was used to find the optimum value for speed, friction, and platform inclination that will minimize deviation from the target value. The optimization model shows that the speed should be kept at a lower level of 300 cm/min; the friction should be kept at the lower level of 0.1, while the platform inclination should be kept at the higher level of 3.2^0. This will reveal a mean deviation from the target value equal to 0.23013. The optimization model results agree with that shown in Figure (4).

6. Motion Standard Deviation

This section studies the effect of the controllable factor (speed, friction, and platform inclination) on the response factor (motion standard deviation) with the aid of the ANOVA. The results of the ANOVA are shown in Table (4).

Table 4. Analysis of Variance For Motion Standard Deviation

Source	Sum of Squares	df	Mean Square	F Value	p-value Prob > F
Model	2.901912	16	0.18137	23.41347	< 0.0001
A-Speed	1.91986	1	1.91986	247.8398	< 0.0001
B-Inclination	0.030586	1	0.030586	3.948369	0.0512
C-Friction	0.275308	1	0.275308	35.54017	< 0.0001
AB	0.004618	1	0.004618	0.596199	0.4429
AC	0.042239	1	0.042239	5.452798	0.0227
BC	0.008936	1	0.008936	1.15359	0.2868
A^2	0.020097	1	0.020097	2.594391	0.1122
B^2	0.014518	1	0.014518	1.87412	0.1758
C^2	0.071316	1	0.071316	9.206427	0.0035

ABC	0.041942	1	0.041942	5.414423	0.0231
A^2B	0.007349	1	0.007349	0.94875	0.3337
A^2C	0.063828	1	0.063828	8.239676	0.0055
B^2C	0.064792	1	0.064792	8.364232	0.0052
A^2B^2	0.004394	1	0.004394	0.567193	0.4541
A^2BC	0.003655	1	0.003655	0.47177	0.4947
A^2B^2C	0.041654	1	0.041654	5.377241	0.0236
Residual	0.495768	64	0.007746		
Lack of Fit	0.105273	10	0.010527	1.455772	0.1819
Pure Error	0.390495	54	0.007231		
Total	3.397681	80			

As can be seen, the model used is significant at significance level of 0.0001. The model contains the speed, platform inclination, and friction with their interactions. The lack of fit was not found to be significant. Testing model graphs for adequacy reveals no severe violations for the ANOVA assumptions. i.e. the residuals are normally distributed with constant variance and the data contains no outlier.

6.1 Motion Standard Deviation – the fitted regression model

The regression model that describes the motion standard deviation (response factor) as a function of the robot speed; the friction between the robot wheels and the platform material and platform inclination (controllable factor) is given as follows:

$$\text{Standard Deviation} = -1.16076 + 1.98E\text{-}003*S - 0.89*I + 11.8*F + 8.4E\text{-}003*S*I + 0.03*S*F + 7.9*I*F + 1.2E\text{-}006*S^2 - 0.15*I^2 - 69.46*F^2 - 0.07*S*I*F - 1.31E\text{-}005*S^2*I - 5E\text{-}005*S^2*F + 0.97*I^2*F + 4E\text{-}007*S^2*I^2 + 1.03E\text{-}004*S^2*I*F - 2.46E\text{-}006*S^2*I^2*F \rightarrow (2)$$

The above model reveals R^2 equals to 0.8541 which means that the fitted model explains 85.41% of the variability in the motion standard deviation (the response variable). The values for the adjusted R^2 and the adequate precision reveals that the model fitness is good and can be trusted for further analysis.

6.2 Motion Standard Deviation– response surface and contour plot

The mathematical model shown in Equation (2) is used to generate the response surface shown in Figure (8) which is aided by the contour plot to assist for the analysis of the effect of the controllable factors on the motion standard deviation.

The figure shows that reducing the robot speed and increasing the platform inclination reduce dramatically the motion standard deviation. This is due to the fact that reduced speed and increased resistance of the platform (elevated platform) gives enough time for the robot to send and receive signals. The response surface and contour plot shown in Figure (8) are drawn at the average friction value of 0.14. Changing the friction from the smaller value of 0.1 to the larger value of 0.17 results in nearly no change in the shape of the plots but the location of the plots is slightly changed.

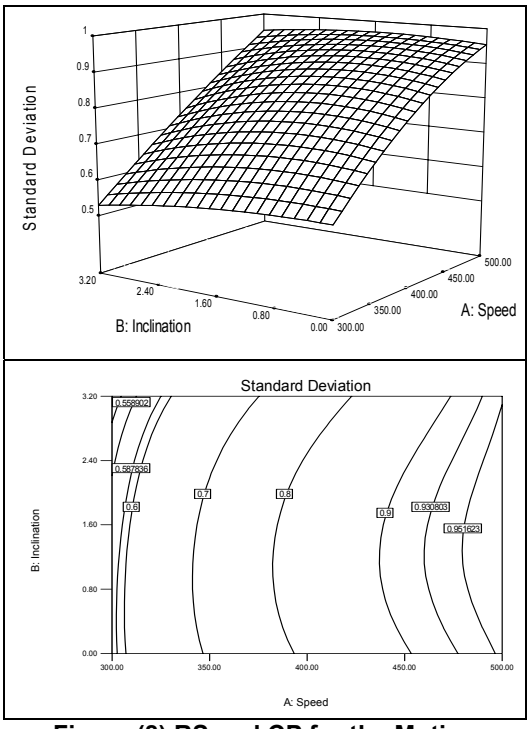


Figure (8) RS and CP for the Motion Standard Deviation

6.3 Motion Standard Deviation – model optimization

Following a similar optimization model for that described in Section 5.3, with a small difference that this model will be a minimization model to minimize the motion fluctuation (standard deviation) the optimization model runs as follows:

Minimize: Motion Standard Deviation

Subject to: $300 \leq \text{Speed} \leq 500$

$0.1 \leq \text{Friction} \leq 0.17$

$0^0 \leq \text{Platform Inclination} \leq 3.2^0$

Non-linear optimization is used to find the optimum value for speed, friction, and inclination that minimize motion standard deviation in order to give more consistent motion. The optimization model shows that the robot speed should be kept at a lower level of 300 cm/min and the platform inclination should be kept at a higher level of 3.2^0 and the friction should be kept at the higher level of 0.17 to reach the minimum motion standard deviation value of 0.325. The optimization results agree with that shown in Figure (8).

7. Multi-objective Optimization Model

Through this section, an investigation is given to the values where the controllable factors (speed, friction, and platform inclination) should be kept to minimize the deviation of both response factors (mean deviation from the target value and motion standard deviation). This multi-objective minimization problem can be expressed numerically as follows:

Minimize: Mean deviation from the target value

Motion standard deviation

Subject to: $300 \leq \text{Speed} \leq 500$

$0.1 \leq \text{Friction} \leq 0.17$

$0^0 \leq \text{Platform Inclination} \leq 3.2^0$

To solve the above multi-objective maximization problem, it is required to find the values where the controllable factors should be kept in order to maximize both of the response factors. In doing so, it is assumed that both of the response factors have equal relative importance. i.e. the two response factors are of the same importance to the decision maker.

Again, the non linear optimization is utilized however; this is a case of multiple objective functions. The solution of the above minimization problem reveals that in order to minimize both of the objective functions, the speed should be kept at 300 cm/min; the platform should be inclined at an angle of 3.2^0; and the Vinyl material should be used. In this case the mean deviation from the target will be equal to 0.3884 and motion standard deviation equal to 0.325.

Conclusion

This paper discusses the influential factors that can affect the accuracy of the mobile robot while tracking a wall. Setting the controllable factors at some specific values will reduce the reponse factors dramitcally and thus improve the mobile robot accuracy.

The study was peformed using 3^3 factorial experiment. Three controllable factors (speed; friction; and inclination) and two response factors (mean deviation from the target and motion standard deviation) are utilized. The paper formulates two mathemtical models that explains the response factors as a function of the controllable factors. These mathemtical models were used to find the values of the controllable factors that will minimize the response factors.

The results of this paper show that in order to minimize both of the response factors, the speed should be kept at 300 cm/min; the platform should be inclined at an angle of 3.2^0 to the ground; and the Vinyl material should be used.

These results of this research can be used in an industrial setting to improve the manoeuvrability of a mobile robot on a factory floor. The user input of the required speed, inclination angle and material of the factory floor will be utilized to compute a correction factor (error). This correction factor is then input to the robot wheel motor controllers which lead to an adjustment of the motion to improve its positional accuracy.

This research has of great benefit to the robot manufacturer as well as the customers or users. The robot manufacturer will accomplish a robot which has a high positional accuracy that will in turn results in improved sales and market share. The users or customers will achieve high positional accuracy from such a low-priced robot.

References

[1] Alici, G. and B. Shirinzadeh, “A systematic technique to estimate positioning errors for robot accuracy improvement using laser interferometry based sensing”, Mechanism and Machine Theory, 40(8), 2005, pp 879-906.

[2] Borm, J.-H. and J.-C. Choi, “Improvement of local position accuracy of robots for offline programming”, Journal of Mechanical Science and Technology, 6(1), 1992, p. 31-38.

[3] Design Expert Software, Statease Incorporation, Version 8.0.1., 2009.

[4] Junchuan, L., Z. Yuru, and L. Zhen, "Improving the Positioning Accuracy of a Neurosurgical Robot System. Mechatronics", IEEE/ASME Transactions, 12(5), 2007, p. 527-533.

[5] Lightcap, C., Hamner, S., Schmitz, T., and Banks, S., "Improved Positioning Accuracy of the PA10-6CE Robot with Geometric and Flexibility Calibration", Robotics, IEEE Transactions, 24(2), 2008, p. 452-456.

[6] Martin-Gorostiza, E., Meca, F. J., Galilea, J. L. L., Martos-Naya, E., Naranjo, F. B., and Esteban, O., "Coverage-Mapping Method Based on a Hardware Model for Mobile-Robot Positioning in Intelligent Spaces", Instrumentation and Measurement, IEEE Transactions, 59(2), 2010, p. 266-282.

[7] Montgomery, D.C., Design and Analysis of Experiments, John Wiley & Sons Inc, 2007.

[8] Morgan, P.S., et al., "The application accuracy of the Pathfinder neurosurgical robot", International Congress Series, 1256, 2003. p. 561-567.

[9] Paziani, F.T., B. Di Giacomo, and R.H. Tsunaki, "Robot measuring form errors", Robotics and Computer-Integrated Manufacturing, 25(1), 2009, p. 168-177.

[10] Smith, G.C. and R.A. Smith, "A non-contact method for detecting on-line industrial robot position errors using a microwave Doppler radar motion detector", International Journal of Advanced Manufacturing Technology, 29(5/6), 2006, p. 605-615.

[11] Sulzer, J. and I. Kovac, "Enhancement of positioning accuracy of industrial robots with a reconfigurable fine-positioning module", Precision Engineering, 34(2), 2010, p. 201-217.

The Design of a Poisson EWMA Control Charts Using Simulation

Tamer A. Mohamed
The British University in Egypt
Tamer.mohamed@bue.edu.eg

Abstract

This paper provides Average Run Lengths (ARLs) tables for various ranges of weighing factor (λ) and the control limit factor (A) to be used for the design of EWMA control charts utilized for Poisson data. It gives the optimum combination of λ and A that satisfies in-control and out of control ARLs conditions using computer simulation.

The computer simulation program is written in MATLAB 7.0 workspace. Once the process average or target (μ_0), and the new expected process average (μ_1) are known, the values for λ and A are determined which satisfies desired conditions of in-control and out of control ARLs.

1. Introduction

Statistical process control techniques such as the control charts have been widely used due to its simplicity and efficiency in detecting any process shift before a large number of products are produced. In manufacturing industries, the most common charts are the $\bar{x}$ and R charts. In these charts, a random sample of size (n) items is repeatedly chosen and the sample mean ($\bar{x}$) and the range (R) of the critical to quality characteristic(s) are sequentially plotted together against pre-calculated control limits [12]. CUSUM, and EWMA have been used extensively to monitor product quality and special causes of variations that might indicate out of control situations, when the shift in critical to quality process parameters is diminutive. This is due to the reason that the Shewhart control charts are relatively inefficient in detecting small changes in the process parameters [1]. The performance of the EWMA is approximately equivalent to that of the CUSUM control charts, and in some ways it is easier to set up and operate [12]. According to Vargas et al [17], the EWMA control charts are more efficient than CUSUM with small changes in the range of $\pm$ 1.0 standard deviation in the process average.

The EWMA control chart was introduced by Roberts [14]. Crowder [6] and Lucas and Saccucci [9] have presented good discussions for the EWMA control charts.

Since the first paper that introduced the EWMA control chart by Roberts [14], a number of researchers have focused on developing and improving different applications for the EWMA control charts to enhance its industrial applications.

Serel and Moskowitz [16] have studied the joint economic design of the EWMA control charts to monitor process mean and variance. They have introduced a cost minimization model to design a joint control scheme for economic or economic statistical design. The weighing factor, sampling interval, sample size, and control chart limits factor were optimized by using a numerical search method. The ARL of the control scheme was computed using a Markov

chain approach. Similar study utilized the cost function of Lorenzen and Vance [8] to determine the optimal values of EWMA control chart parameter which was presented by Serel [15]. The study has assumed that the cost of nonconformities, when the process is in control or out of control, was not constant and depends on a linear/quadratic loss function.

Reynolds and Stoumbos [13] have showed that the combination of EWMA control charts to control process mean and variability is more sensitive to small shifts than the individual and moving range control charts in case of individual data.

Borror et al. [4] have studied the robustness of the EWMA control charts to non normality for collected data in case of skewed and heavy tailed symmetric non normal distributions, namely Gamma distribution and t-distribution respectively. The study has showed that the Shewhart individual control charts were not robust to any change in the normality assumption if false alarms are of major concern. However, the EWMA was proven to maintain its expected power of detecting small shifts in the process mean even if the normality assumption was violated.

Aparisi and Diaz [2] have presented a method for the design of EWMA control charts to be used for detection of shift that is not necessary small in magnitude. Their paper has determined the optimal parameters of EWMA charts for specific expected region of process shift magnitude.

The EWMA control charts for monitoring process mean have received better attention in literature than those control process variance; however, a number of researchers studied the monitoring process variance when using EWMA control charts assuming that the process parameters are known. Among those are Mac Gregor and Haris [10] and Castagliola [5]. In addition, Maravelakis and Gastagliola [11] have proposed a modified EWMA control chart for monitoring process standard deviation when process parameters are estimated instead of being assumed to be known.

2. The EWMA control charts

The exponentially weighted moving average statistic is defined as

$$z_i = \lambda x_i + (1-\lambda)\, z_{i-1} \qquad \rightarrow (1)$$

where $0 < \lambda \leq 1$ is a constant that is mostly called weighing factor, and the starting value required at i = 1 is found by setting $z_0 = \mu_0$ (the process target value).

In the forthcoming discussion, it is assumed that the process target mean μ_0 and process standard deviation σ_0 are known. The control limits for the EWMA control charts are

$$\text{UCL} = \mu_0 + A\sigma \sqrt{\left(\frac{\lambda}{2-\lambda}\right)[1-(1-\lambda)^{2i}]} \qquad \rightarrow (2)$$

$$\text{LCL} = \mu_0 - A\sigma \sqrt{\left(\frac{\lambda}{2-\lambda}\right)[1-(1-\lambda)^{2i}]} \qquad \rightarrow (3)$$

where, $A\sigma$ determines the width of the control limits. For large values of i, the steady-state (asymptotic) EWMA control limits are:

$$\text{UCL} = \mu_0 + A\sigma \sqrt{\left(\frac{\lambda}{2-\lambda}\right)} \qquad \rightarrow (4)$$

$$\text{LCL} = \mu_0 - A\sigma \sqrt{\left(\frac{\lambda}{2-\lambda}\right)} \qquad \rightarrow (5)$$

The EWMA statistics (Eq.1) are then plotted on a control chart with the UCL and LCL and if

any point goes out of the control limits, the process is said to be out of control.

The design parameters of the EWMA control charts are the control limits factor (A) and the value for the weighing factor (λ). It is possible to choose these parameters to give Average Run Length performance for the EWMA control chart, which is the number of samples (items) chosen until a signal (false or true) is found. Many researchers studied the ARL characteristics of the EWMA control charts (Crowder [7]; Lucas and Saccucci, [9]). These studies have provided average run length tables and graphs for different combination of λ and A assuming observations coming from a normal distribution. The optimal design procedure would consist of specifying the desired in-control and out-of control average run lengths and the magnitude of the expected process shift.

Borror et al [3] have developed a methodology for the use of EWMA control charts in case of monitoring the mean of a quality characteristic that follows Poisson distribution. The paper has developed graphs to be used for the selection of the value of the control limit for a given value of in-control average run length.

Most of the above papers have utilized the Markov chain approach to approximate the computations of the ARLs.

In most cases, it might be desired to design the control charts taking both the in-control and out-of-control ARLs into consideration. This paper utilizes a computer simulation program to find the optimal combination of λ and A that will ensure achieving specific conditions (values) for in-control and out-of control ARLs in case of a particular value of expected process shift when the process data follow Poisson distribution.

The following section explains the computer simulation program used to compute the in-control and out-of-control ARL. Section (3) explains the design tables, followed by a numerical illustration (Section 4) to show the use of the program. Finally the paper ends with the concluding remarks.

3. The computer simulation program

When the quality characteristic X, is the number of nonconformities in a unit from a repetitive production process, the data can be represented using the Poisson distribution. i.e., the observed values X_1, X_2,…., are independent and are identically distributed Poisson random variables with mean μ.

The process is said to be in a state of statistical control (stable process) when $\mu = \mu_0$ and out of control and when the mean changes to some other value $\mu = \mu_1$. When the process is out of control, it is required to detect the shift early in the process in order to make adjustments. For stable process, it is required that the rate of false alarms is small (high in control ARL value)

Setting the value for the process standard deviation σ_0 equal to the square of the process mean μ_0 in the Eq. (4) and Eq. (5) (steady state situation), so the UCL and LCL for the EWMA control charts will be as shown below in Eq. (6) and Eq. (7), while the exponentially weighted moving average statistic will remain the same as in Eq. (1).

$$UCL = \mu_0 + A\sqrt{\left(\frac{\lambda\mu_o}{2-\lambda}\right)} \quad \rightarrow (6)$$

$$LCL = \mu_0 - A\sqrt{\left(\frac{\lambda\mu_0}{2-\lambda}\right)} \quad \rightarrow (7)$$

The computer simulation program is written in MATLAB 7.0 workspace. This program is used to calculate the in-control and out-of-control ARL at different values of input parameters and chart parameters (λ and A) as it is shown in the Figure 1 below. The logic for the program can be explained as follows:

1. *Start*
2. *Input process parameters:*
   a. *Process target mean (μ_0)*
   b. *Process new (shift) mean (μ_1)*
   c. *Number of random trials (re)*
3. *Input Range for*
   a. *Weighing factor (λ)*
   b. *Control limit factor (A)*

For Each combination of λ & L do the steps from 4 to 11

4. *Compute UCL an LCL for the EWMA control Chart as follows:*

$$UCL = \mu_0 + A\sqrt{\frac{\lambda\mu_0}{2-\lambda}}$$

$$LCL = \mu_0 - A\sqrt{\frac{\lambda\mu_0}{2-\lambda}}$$

5. *Generate a random numbers (x_i)that follows a Poisson Distribution with mean μ_1*
6. *Compute the exponentially moving value (Z_i) as a function of (x_i) as follows:*

$$Z_i = \lambda X_i + (1-\lambda)Z_{i-1}$$

$$Z_1 = \mu_1$$

7. *Compare (Z_i) to the UCL and LCL*
   if LCL< (Z_i) < UCL, Repeat steps 5, 6, and 7
   Otherwise go to step 8
8. *Compute ARL for this trial (ARL_j) where j, is the number of trials $1 \leq j \leq re$*
9. *Repeat Steps (5-8) for a number of times equal to the specified value re*
10. *Compute A_ARL (Average average run length) for all trials as follows:*

$$A\_ARL = \frac{\sum_{j=1}^{j=re} ARL_j}{re}$$

11. *End*

3.1. Simulation sample size

To come to a decision on the sufficient number of random trials (simulation sample size) it is required to ensure the minimum simulation run time that can produce ARL values as close as possible to true ARL values, the simulation is repeated at different values of random trials (100, 500, 1000, 2000, 5000, 10000, 12000, and 15,000). At each time, the ARL values were compared to the previous one until it was found that increasing the number of random trials did not change the values of the ARL. i.e. the ARL value became approximately constant despite the change of the sample size.

As shown in Figure(1), this stabilization happens at simulation sample size equal to 5000. Increasing the number of random trials from 10,000 and 15,000 did not really change the ARL values but increased the run time dramatically.

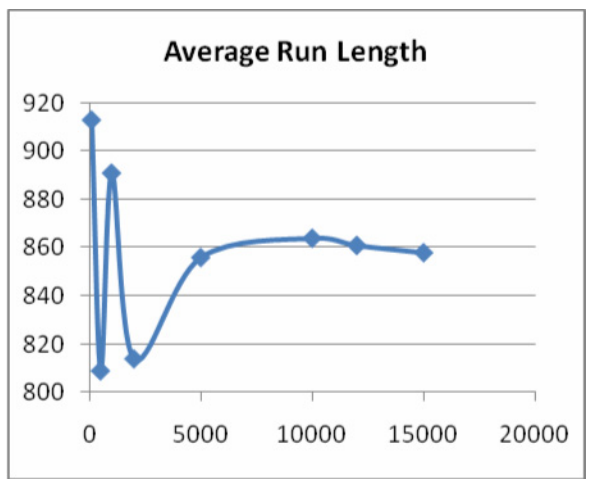


Figure 1. ARL at Different Sample Size

Thus the sample size (number of trials) was set at 5000 value to reduce the run time and minimize the variability from the true ARL

values. In this Figure, the values of the ARL represent the in-control average run length and were computed at $\mu_0 = 4$, A = 3.5, and $\lambda = 0.5$ as an example.

3.2. Simulation model validation

The value of the ARL at 5,000 random trials is compared to the output of the study of Borror et al. (1998) to ensure program validity. As a check, the in-control ARL in our study was compared to those of Borror et al. (1998). This was done at $\mu_0 = 5$, A = 3.5, and $\lambda = 0.1$, 0.2, 0.3, 0.4, and 0.5 as an example. The output of the comparison is shown below in Table (1) which proves that the computer simulation produce results that are very close (average deviation less than 5%) to those of Borror et al. (1998) which is considered satisfactory due to the fact that both studies involve approximate approach to the computation of the ARL values. Thus the simulation model is validated.

Table 1. Comparison with Borror et al. (1998)

λ	0.1	0.2	0.3	0.4	0.5
Borror et al.	3500	2040	1470	1140	950
The Current Simulation	3701	2156	1520	1216	978
Deviation	5.7%	5.6%	3.4%	6.6%	2.9%

It might be worth saying here that in this sub-section, only the in-control ARL is to be calculated. This is due to the fact that Borror et al. study is only focusing on the in-control ARL. However, in the rest of the paper, the computer simulation program computes both the in control and out of control ARL values at each combination of λ and A.

4. Design Tables

The output of the computer simulation program (Figure 2) is shown in Table (2) through Table (6). The first value in each cell is the in-control ARL at the specified combination of λ and A. The three other values in the cell represent the out-of-control ARL when the process average shifts to a bigger value than the process target by 1, 2, or 3. This process was repeated for different values (4, 5, 6, 7, and 10) of initial process target μ_0.

These designed tables can be used to find the values for λ and A that satisfy specific value of in control ARL with the minimum out of control ARL or specific value for out of control ARL with maximum value of in-control ARL.

5. Numerical illustration

Suppose that one wishes to design an EWMA control chart with an in-control ARL that is greater than or equal 350 units with the maximum sensitivity. Assume that the process mean $(\mu_0) = 6$ and this process mean might shift to $(\mu_1) = 7$.

Solution:

The maximum sensitivity means minimum out of control ARL, minimum ARL_1, ARL_0, $\geq$ 350. Looking at the Table of $(\mu_0) = 6$, at each combination of λ and L, the first value indicates ARL_0 and the Second Value Indicates ARL_1 (where process mean shifts to $(\mu_1) = 7$. Mark all the cells that contain ARL_0 greater than or equal

to 350, from these cells choose the one that has the minimum ARL_1. The solution is shown on Table (4) which reveals that λ and A should be kept at the value of 0.1 and 2.7 respectively

6. Concluding remarks

This paper presented an extension for the EWMA control charts that would be used for Poisson data when the design is based on a specified value for in-control and out-of-control ARLs.

The use of the tables presented in this paper in addition to the computer program will in-turn promotes the use of the EWMA control by practitioners in the future.

References

[1] Amin, R.W., Wolff, Hans, Besenfelder, Werner, and Baxley, Robert, "EWMA Control Charts for the Smallest and Largest Observations", 31, No. 2, (1999), 189-201.

[2] Aparisi, Francisco, Diaz, J. Carlos Garcia, "Design and Optimization of EWMA Control Charts for In-Control, Indifference, and Out of Control Regions", Computers and Operations Research, 34, (2007), 2096-2108.

[3] Borror, Connie M., Cham, Charles W., and Rigdon, Steven, E., "Poisson EWMA Control Charts", Journal of Quality Technology, 30(4), (1998), 352-361.

[4] Borror, Connie M., Montgomery, Douglas C., and Runger, George, C., "Robustness of the EWMA Control Chart to Non-normality", Journal of Quality Technology, 31(3), (1999), 309-316.

[5] Castagliola, P., "A new S2 – EWMA Control Chart For Monitoring the Process Variance", Quality and Reliability Engineering International, 21(8), (2005), 781-794.

[6] Crowder, S. V., "Design of Exponentially Weighted Moving Average Schemes", 21(3), (1989), 155-162.

[7] Crowder, Stephen V., "A Simple Method For Studying Run-Length Distributions of Exponentially Weighted Moving Average Charts", Technometrics, 29(4), (1987), 401-407.

[8] Lorenzen, T.J. and Vance, L.C., "The Economic Design of Control Charts: A Unified Approach", Technometrics, 28, (1986), 3-10.

[9] Lucas, J.M. and Saccucci M.S., "Exponentially Weighted Moving Control Shemes: Properties and Enhancements", Technometrics 32, (1990), 1-12.

[10] MacGregor, J., and Harris, T., "The Exponentially Weighted Moving Variance", Journal of Quality Technology, 25, (1993), 106-118.

[11] Maravelakis, Petros E., and Gastagliola, P., "An EWMA Chart for Monitoring the Process Standard Deviation When Parameters Are Estimated", Computational Statistics and Data Analysis, 53, (2009), 2653-2664.

[12] Montgomery, D.C., Statistical Quality Control. John Wiley and Sons Inc., (1991).

[13] Reynolds, M.R. and Stoumbos, Z.G., "Monitoring the Process Mean and Variance Using Individual Observations and Variable Sampling Intervals", Journal of Quality Technology, 33(2), (2001), 181-205`

[14] Roberts, S.W., "Control Charts Tests Based on Geometric Moving Averages", Technometrics 1, (1959), 239-250.

[15] Serel, D. A., "Economic Design of EWMA Control Charts Based on Loss Function", Mathematical and Computer Modelling, 49, (2009), 745-759.

[16] Serel, D. A. and Moskowitz, H., "Joint Economic Design of EWMA Control Charts For Mean and Variance", European Journal of Operational Research, 184, (2008), 157-168.

[17] Vargas, V.C., Lopes, Luis Felipe Dias, and Souza, Adriana Mendonca, "Comparative Study of the Performance of the CUSUM and EWMA Control Charts", Computers and Industrial Engineering, 46, (2004), 707-724.

```
%************** Author :Dr. Tamer A. Mohamed ************************
%************** EWMA Control Charts Program  ************************
clear all
close all
clc

% *******************Input Variables*********************************
me=4;
me1=4:7;
A= 2.5:0.1:3.5;
lum=0.1:0.1:0.9;
re=input('please Enter the Number of Random Trails=    ');

%******************Display Input Variables**************************
out=fopen('res5.txt','w');
fprintf(out,'*************************************************************');
fprintf(out,'\n');
fprintf(out,'********** Result of Control Charts Program **************');
fprintf(out,'\n');
fprintf(out,'*************************************************************');
fprintf(out,'\n');
fprintf(out,'    me       me1        lum         A      A_ARL ');
fprintf(out,'\n');

%************************Computing Control Limits***********************
for lum1=lum;
   for A1=A;
        for me10=me;
             for me11=me1;
k=1;
av       = me10;
upper     =av + A1*(sqrt(av*lum1/(2-lum1)));
lower     =av - A1*(sqrt(av*lum1/(2-lum1)));

%****************random sample generation****************

   r=1;

   while r<=re
        flag=1;
        loop_flag=1;
        j=0;
        trial=0;

        while loop_flag==1
             data = poissrnd(me11,1,1);

          [a b]=size(data);

             for i=1:a;
             if trial==0
             data1=me10;
             else
                  data1=lum1*data+(1-lum1)*s;
             end
                  s(i)= data1;

             end
                  data1(b+1)=s;

             c=0;
```

Figure 2. The Computer Simulation Program

```
    if (data1(b+1)>=upper)|| (data1(b+1)<=lower);
            loop_flag=0;
            trial=trial+1;
        else
            loop_flag=1;

            trial=trial+1;
        end
    if (data1(b+1)>=upper)|| (data1(b+1)<=lower);
            c=1;
    end

    end

    ARL(r)=trial;
    pro(r)=1/trial;
    A_ARL =(sum(ARL)/r);
      r=r+1;
end

disp('A_ARL =   ')
disp(A_ARL)
fprintf(out,'%4.0f  %4.0f   %1.1f  %1.1f  %4.3f',me10,me11,lum1,A1,A_ARL);
fprintf(out,'\n');
            end
        end
    end
end
fclose(out);
```

Figure 2. The Computer Simulation Program Cont.

Table 2. In-Control and Out-of-Control Average Run Lengths (μ_0 =4, μ_1 = 5, 6, and 7)

	λ								
A	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**
2.5	228.798	139.422	113.345	96.843	91.295	83.499	81.759	79.028	76.784
	22.999	20.786	20.083	19.154	19.429	19.618	19.323	20.824	20.352
	9.793	8.382	8.266	7.972	8.016	8.192	8.295	8.903	8.819
	6.469	5.463	5.176	4.963	4.82	4.866	5.124	5.352	5.277
2.6	288.135	188.684	147.091	123.443	113.46	105.995	98.838	92.019	105.39
	24.545	22.651	21.756	21.597	22.333	21.85	22.162	23.584	25.959
	10.336	8.979	8.626	8.506	8.897	8.707	9.159	9.563	10.363
	6.656	5.65	5.418	5.131	5.307	5.099	5.291	5.485	5.818
2.7	372.003	239.605	184.841	158.061	142.737	129.58	122.063	110.673	118.609
	26.931	25.055	24.239	24.519	24.761	24.481	25.701	26.507	30.401
	10.832	9.447	9.046	9.049	9.436	9.258	10.063	10.536	12.106
	6.904	5.914	5.474	5.393	5.444	5.305	5.688	5.82	6.361
2.8	463.132	305.984	235.915	200.237	175.595	160.008	147.998	138.787	123.909
	29.222	27.46	27.518	27.233	27.53	28.898	29.83	30.338	32.419
	11.164	10.031	9.638	9.594	10.037	10.607	10.703	11.368	13.052
	7.104	6.187	5.842	5.567	5.648	5.838	5.948	6.108	6.871
2.9	623.975	397.024	298.212	257.273	220.816	198.107	176.062	168	144.238
	31.243	30.608	30.96	31.478	31.651	33.763	33.912	35.292	34.854
	11.88	10.485	10.354	10.406	11.07	11.697	12.483	12.511	12.918
	7.361	6.332	6.085	6.029	5.937	6.291	6.536	6.539	6.907
3	816.758	508.24	401.264	317.305	274.82	241.977	226.769	212.033	206.52
	34.171	33.973	34.875	35.78	38.409	39.138	39.854	42.158	42.387
	12.187	11.19	11.077	11.374	11.933	12.492	13.489	14.745	14.886
	7.659	6.652	6.335	6.235	6.191	6.665	6.908	7.565	7.834
3.1	1087.709	671.907	489.468	403.151	350.426	308.528	278.365	259.138	293.09
	37.157	38.02	39.231	41.157	43.366	44.948	47.239	49.423	55.214
	12.944	12.059	11.849	12.555	13.049	13.995	14.957	16.431	18.126
	7.893	6.888	6.636	6.623	6.867	7.033	7.338	8.078	8.664
3.2	1427.134	893.318	636.113	513.645	422.681	376.93	344.846	311.74	343.345
	40.729	43.218	45.735	47.174	50.253	50.638	53.234	57.525	69.316
	13.564	12.649	12.793	13.358	14.35	15.411	16.377	17.779	21.446
	8.209	7.223	6.899	6.952	7.251	7.235	7.996	8.447	9.752
3.3	1956.606	1190.585	828.387	651.642	538.246	474.896	434.13	393.454	358.538
	45.07	48.667	51.933	53.434	58.65	61.462	64.118	66.965	73.127
	14.052	13.39	13.889	14.84	15.699	17.074	19.066	20.855	23.821
	8.515	7.459	7.348	7.465	7.619	8.003	8.757	9.164	11.056
3.4	2618.069	1553.392	1096.749	820.948	677.641	594.267	549.963	492.134	412.918
	47.965	54.688	58.322	63.689	66.562	72.463	77.85	79.096	78.249
	14.573	13.939	15.057	15.799	17.221	19.007	21.455	23.134	25.367
	8.759	7.741	7.687	7.684	8.106	8.792	9.677	10.11	11.369
3.5	3644.707	2011.725	1395.26	1063.664	856.202	752.404	680.886	622.677	617.324
	53.372	61.14	70.027	73.099	79.191	85.733	90.558	96.109	99.137
	15.385	15.155	16.095	17.282	19.893	21.183	24.327	26.588	29.452
	9.087	8.011	7.955	8.255	9.009	9.451	10.302	11.571	12.501

Table 3. In-Control and Out-of-Control Average Run Lengths (μ_0 =5, μ_1 = 6, 7,

	λ								
A	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**
2.5	221.752	141.083	113.109	97.355	89.113	86.753	79.587	81.193	74.968
	27.152	24.36	23.049	22.491	22.117	22.85	22.705	23.595	24.242
	11.194	9.55	9.141	9.224	9.196	9.684	9.666	10.425	11.213
	7.182	6.106	5.733	5.509	5.513	5.745	5.671	5.917	6.595
2.6	275.407	188.473	142.014	130.257	113.54	107.997	99.51	98.713	86.566
	28.451	27.132	26.066	25.345	25.296	25.928	26.624	26.759	26.344
	11.727	10.274	9.924	10.163	10.076	10.615	11.182	11.268	11.582
	7.42	6.384	6.001	5.999	5.798	5.966	6.297	6.231	6.594
2.7	369.691	242.236	183.618	166.714	146.529	134.952	128.433	125.737	118.986
	32.248	30.184	29.11	28.707	28.813	29.971	30.841	32.32	32.151
	12.315	10.927	10.718	10.967	10.647	11.403	11.902	13.127	12.963
	7.73	6.635	6.274	6.224	6.047	6.391	6.566	7.162	7.278
2.8	470.476	303.649	242.013	206.685	180.727	167.972	160.183	154.266	157.463
	34.91	33.877	32.738	33.351	34.13	34.595	35.902	38.087	41.84
	12.951	11.669	11.381	11.645	12.184	12.51	13.331	14.232	15.482
	8.019	6.979	6.623	6.586	6.541	6.821	7.086	7.709	8.13
2.9	634.1	398.033	313.008	258.558	233.742	209.41	198.994	179.757	181.012
	38.584	37.234	37.829	38.47	39.548	40.981	42.666	43.827	49.194
	13.659	12.611	12.43	12.986	13.067	14.407	14.653	16.091	18.414
	8.36	7.284	7.087	6.876	6.978	7.522	7.582	8.243	9.274
3	816.066	533.576	408.101	335.808	301.331	259.352	242.642	231.569	195.176
	41.878	42.539	42.952	43.927	46.404	47.869	50.503	52.219	52.616
	14.245	13.134	13.305	13.88	14.521	15.863	17.551	18.325	19.894
	8.658	7.52	7.336	7.309	7.416	8.01	8.585	8.799	10.045
3.1	1102.94	687.891	523.384	434.623	373.532	334.954	298.155	283.425	259.913
	45.149	47.333	49.437	51.581	54.324	54.59	58.712	61.468	59.815
	14.899	13.871	14.396	15.244	17.05	17.505	19.683	21.395	21.938
	8.915	7.98	7.709	7.759	8.24	8.576	9.352	10.29	10.137
3.2	1419.98	940.656	700.016	552.942	480.457	420.213	383.656	353.108	377.878
	51.013	54.61	57.391	59.874	64.249	65.057	68.386	73.795	79.119
	15.804	15.056	15.717	16.954	18.619	19.584	21.948	24.123	26.324
	9.296	8.259	8.174	8.369	8.917	9.557	9.984	10.975	11.597
3.3	1979.02	1235.16	897.289	702.761	594.576	527.707	471.975	446.02	461.095
	55.462	60.954	65.93	71.106	75.278	77.137	82.65	88.342	102.396
	16.43	16.296	16.885	18.466	20.266	22.87	24.628	27.875	31.625
	9.533	8.708	8.541	8.815	9.308	10.197	10.942	12.36	13.934
3.4	2737.85	1686.91	1210.47	927.357	766.889	656.546	610.41	559.638	505.627
	61.859	71.594	79.083	84.362	90.144	93.412	96.539	106.948	112.146
	17.39	17.257	18.881	20.741	23.1	25.098	28.421	31.632	38.028
	9.962	9.126	8.9	9.295	10.158	11.427	11.916	13.519	16.463
3.5	3701.27	2156.21	1519.64	1216.42	977.687	834.197	747.428	705.601	614.052
	67.893	83.711	92.964	99.764	104.579	114.933	119.284	124.331	126.139
	18.106	18.635	20.577	23.203	25.849	29.382	33.383	36.435	39.649
	10.364	9.456	9.593	10.19	11.256	11.967	13.198	14.453	16.86

Table 4. In-Control and Out-of-Control Average Run Lengths (μ_0 =6, μ_1 = 7, 8, and 9)

	λ								
A	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**
2.5	224.789	148.363	114.314	100.916	91.972	88.087	82.519	78.716	79.854
	30.748	28.276	25.962	25.727	25.189	25.358	26.757	26.793	28.409
	12.287	11.124	10.436	10.686	10.623	11.078	11.174	12.083	12.549
	7.825	6.844	6.301	6.21	6.374	6.445	6.644	7.087	7.146
2.6	283.188	187.064	149.311	128.833	116.21	108.796	103.461	98.866	107.233
	34.012	30.715	30.146	29.273	28.855	29.782	29.949	30.969	34.849
	12.896	11.623	11.448	11.391	11.82	11.906	12.227	13.302	14.683
	8.188	7.035	6.731	6.506	6.654	6.768	7.121	7.399	8.053
2.7	362.096	238.661	188.487	166.037	149.594	136.323	132.606	122.288	117.456
	37.222	34.551	34.293	33.569	34.152	34.503	35.532	36.193	38.876
	13.579	12.31	12.348	12.44	13.04	13.649	13.971	14.878	16.407
	8.477	7.406	7.06	7.006	7.131	7.341	7.651	7.977	8.964
2.8	489.488	306.601	249.827	215.011	188.211	174.885	169.354	157.808	147.216
	40.881	38.945	39.597	38.95	39.959	39.748	41.377	43.828	42.607
	14.557	13.401	13.404	13.647	14.525	14.654	16.141	16.945	17.54
	8.799	7.773	7.463	7.398	7.661	7.902	8.447	8.644	9.237
2.9	631.925	409.158	325.873	271.67	245.051	221.385	209.492	200.196	200.881
	44.614	44.449	44.308	46.033	47.035	46.564	49.33	52.511	55.064
	15.111	14.108	14.392	15.041	16.066	16.968	18.118	19.728	20.766
	9.125	7.991	7.852	7.926	8.272	8.488	8.978	9.9	10.544
3	832.542	529.524	434.653	360.934	306.4	278.75	257.422	247.148	254.31
	50.505	49.404	50.439	53.763	54.434	56.501	59.704	62.717	69.112
	16.135	15.181	15.25	16.249	17.775	18.927	20.479	22.587	25.294
	9.494	8.493	8.297	8.529	8.941	9.566	9.813	11.274	11.845
3.1	1108.51	725.566	568.2	451.68	403.402	363.802	325.231	309.737	284.771
	54.746	56.756	59.184	62.05	64.995	67.557	71.162	72.931	78.831
	16.891	16.52	17.096	18.378	19.565	21.361	23.225	25.946	29.025
	9.974	8.821	8.754	9.161	9.421	10.478	10.919	11.759	13.653
3.2	1505.36	990.184	736.972	609.315	518.072	458.601	414.997	374.419	341.072
	59.805	66.169	70.202	74.277	77.749	80.561	83.819	88.434	89.266
	17.96	17.793	18.553	20.295	22.241	24.115	27.783	29.416	32.064
	10.305	9.392	9.154	9.504	10.445	11.136	12.389	13.456	15.083
3.3	2039.69	1310.74	976.419	790.422	673.352	576.292	544.662	496.599	508.641
	67.23	75.808	83.165	89.751	92.765	96.288	100.326	107.376	114.912
	19.047	19.239	20.549	23.195	25.777	27.474	31.842	35.777	38.262
	10.795	9.825	9.948	10.486	11.425	12.402	14.061	15.064	16.963
3.4	2816.69	1768.24	1300.74	1016.03	860.287	729.435	662.285	615.926	632.657
	77.974	90.625	98.3	107.18	110.475	119.114	123.919	129.749	143.735
	20.064	20.587	23.03	25.478	28.852	32.709	36.752	41.023	46.418
	11.162	10.347	10.491	11.326	12.31	13.816	15.391	17.649	19.549
3.5	3765.63	2389.78	1747.79	1394.34	1123.13	941.621	856.708	796.499	725.514
	85.981	106.321	122.019	129.786	133.731	141.451	149.652	161.941	166.575
	21.307	22.322	25.098	29.614	33.1	37.245	42.104	48.903	54.505
	11.537	10.82	11.167	12.18	13.4	14.892	16.818	19.345	23.163

Table 5. In-Control and Out-of-Control Average Run Lengths (μ_0 =7, μ_1 = 8, 9, and 10)

	λ								
A	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**
2.5	220.507	145.552	113.59	97.599	91.215	85.001	81.668	79.451	76.31
	34.385	30.713	29.927	28.335	27.978	28.611	28.652	29.038	29.872
	13.641	12.175	11.692	11.69	12.24	12.32	12.671	13.155	14.288
	8.53	7.227	6.983	6.952	7.007	7.068	7.277	7.397	8.395
2.6	286.327	185.631	146.756	128.603	115.484	108.221	104.859	102.194	98.302
	38.667	34.502	32.811	32.296	32.395	33.545	34.484	35.551	34.216
	14.422	12.751	12.625	12.387	13.128	13.466	14.404	15.243	15.357
	8.879	7.7	7.324	7.199	7.511	7.708	8.127	8.624	8.59
2.7	362.703	241.137	191.019	169.168	155.425	138.426	134.22	131.679	130.889
	41.975	39.161	38.181	37.996	38.38	38.257	38.856	41.352	43.013
	15.205	13.83	14.046	13.964	14.56	15.567	16.264	17.396	18.551
	9.184	8.025	7.814	7.614	8.013	8.29	8.691	9.453	9.888
2.8	484.16	319.033	248.038	215.013	192.666	181.369	172.636	164.13	166.871
	46.335	43.961	44.607	44.472	45.435	45.691	47.533	49.321	54.135
	16.109	15.126	15.086	15.485	16.19	17.395	18.211	20.036	22.294
	9.734	8.435	8.21	8.409	8.677	9.242	9.47	10.545	11.285
2.9	623.911	410.455	335.592	276.998	251.698	234.523	216.389	213.103	189.386
	51.224	52.134	52.351	51.876	52.502	56.148	56.328	60.409	61.048
	17.121	16.286	16.53	17.313	18.294	19.451	21.318	22.226	25.06
	10	8.963	8.766	9.127	9.038	10.001	10.729	11.466	12.625
3	837.045	548.986	444.379	369.3	321.179	290.042	263.256	262.177	244.184
	57.504	58.516	60.664	60.984	62.766	66.356	67.339	71.426	70.419
	18.01	17.407	17.984	19.398	20.932	21.801	24.324	26.427	27.149
	10.448	9.393	9.452	9.625	10.039	10.76	12.166	12.855	13.572
3.1	1123.41	731.298	576.156	488.175	410.505	379.01	343.203	329.89	336.793
	63.905	67.291	72.017	72.725	76.054	78.237	82.622	86.271	93.583
	19.392	18.802	19.876	21.754	23.584	25.008	27.823	31.26	34.431
	10.907	9.888	9.813	10.313	10.891	11.771	13.039	14.291	15.688
3.2	1502.2	1006.87	760.246	629.4	540.143	481.179	453.741	412.437	408.346
	70.744	78.291	84.974	88.548	91.719	90.658	97.904	102.993	113.746
	20.206	20.496	22.172	23.558	26.502	29.681	32.101	35.709	41.02
	11.449	10.28	10.594	11.239	12.252	13.175	14.58	16.823	18.619
3.3	2089.03	1336.42	1042.24	837.322	700.67	606.725	559.002	529.659	471.915
	81.196	93.166	98.563	104.058	110.731	115.607	117.715	126.443	126.692
	21.778	22.544	24.457	27.272	30.613	33.908	37.952	42.912	46.4
	11.802	11.013	11.299	11.942	13.184	14.243	16.248	18.126	21.104
3.4	2711.06	1908.21	1406.16	1083.73	927.748	822.745	716.427	667.442	638.723
	91.976	108.936	120.937	127.569	134.461	139.401	147.245	154.893	156.831
	22.897	24.801	27.485	30.928	35.578	38.415	44.244	49.998	54.456
	12.374	11.759	12.281	13.429	14.445	16.129	18.107	20.747	23.709
3.5	3892.74	2488.27	1838.95	1454.75	1181.25	1018.67	923.509	863.934	857.861
	104.484	129.719	143.999	154.044	168.067	169.854	176.049	188.331	204.341
	24.455	26.537	31.107	35.674	40.422	48.008	52.68	59.188	66.598
	12.868	12.423	13.104	14.424	15.819	18.686	20.624	23.833	27.761

Table 6. In-Control and Out-of-Control Average Run Lengths (μ_0 =10, μ_1 = 11, 12, and 13)

	λ								
A	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**
2.5	220.188	142.506	114.297	102.93	92.713	88.399	83.93	82.335	77.093
	44.221	39.801	37.595	36.618	36.153	35.124	36.362	36.222	36.245
	16.96	15.28	15.348	15.674	15.835	16.47	16.965	18.432	18.454
	10.214	8.937	8.824	8.623	8.793	9.29	9.753	10.202	10.726
2.6	283.27	188.383	149.621	133.714	123.522	113.425	108.839	104.278	105.715
	50.948	45.325	44.011	42.171	42.387	41.328	42.882	43.849	46.716
	18.044	16.988	16.625	17.319	17.274	18.21	19.812	20.862	21.647
	10.816	9.563	9.136	9.507	9.553	9.929	10.701	11.71	12.353
2.7	371.983	239.023	193.788	175.828	159.184	145.414	139.904	134.492	130.877
	55.903	52.555	52.44	50.06	50.62	52.81	53.055	55.066	54.027
	19.516	18.476	18.131	18.978	20.116	21.289	22.328	24.551	25.983
	11.501	10.248	9.924	10.133	10.53	11.563	11.591	12.832	14.31
2.8	487.468	312.925	258.677	226.296	208.643	194.133	180.504	176.628	166.982
	63.534	61.531	60.549	60.97	61.096	62.493	64.071	64.039	67.653
	20.855	20.097	20.83	22.117	22.852	24.657	25.374	28.151	30.072
	11.895	10.855	10.829	10.946	11.518	12.531	13.603	15.031	15.591
2.9	634.408	405.082	352.688	299.058	268.879	257.366	237.005	227.497	225.57
	71.318	70.997	72.222	71.998	75.293	76.048	76.675	78.51	82.978
	22.433	21.665	22.953	24.649	26.391	28.745	30.747	32.563	36.234
	12.539	11.426	11.602	12	12.96	13.93	15.325	16.534	18.364
3.0	843.746	556.861	465.636	403.121	361.792	332.531	313.335	288.793	284.552
	83.008	84.666	86.439	88.163	91.511	92.305	93.064	99.515	100.127
	23.553	24.186	25.903	27.495	30.467	32.808	35.815	39.727	44.273
	13.201	12.231	12.58	13.172	14.348	15.363	17.131	18.505	22.24
3.1	1118.993	755.06	623.982	538.127	490.2	430.435	413.932	373.973	347.523
	92.761	99.557	103.48	103.885	110.282	113.63	116.304	121.697	121.269
	25.514	26.363	29.432	32.01	36.256	38.303	42.725	46.795	49.519
	14.034	12.984	13.673	14.697	15.991	17.658	19.609	21.79	25.212
3.2	1554.043	1048.23	852.557	743.606	648.712	583.495	521.582	491.541	484.873
	104.905	121.649	127.177	129.137	138.414	139.327	147.984	148.826	159.257
	27.999	29.643	32.637	37.345	42.34	45.338	50.336	53.9	62.424
	14.554	13.859	14.848	16.255	17.769	19.823	21.85	25.196	28.978
3.3	2033.41	1450.888	1144.284	993.337	863.431	767.708	700.113	627.875	597.651
	123.93	139.438	155.158	162.003	168.222	169.49	180.998	184.305	192.074
	29.591	33.655	37.223	43.505	49.321	54.576	60.281	66.064	78.601
	15.286	14.709	16.196	17.829	20.389	22.762	26.29	30.127	34.765
3.4	2760.196	1986.135	1624.829	1336.157	1162.147	1009.177	916.937	797.323	725.847
	143.351	171.593	194.084	201.653	206.344	215.499	227.541	234.548	232.068
	32.14	37.226	43.266	51.291	56.983	65.496	73.57	79.568	87.407
	16.122	16.356	17.709	20.311	23.719	26.871	30.558	34.372	40.775
3.5	4006.781	2740.737	2207.727	1848.628	1547.722	1342.37	1206.599	1078.97	1056.045
	169.83	211.991	232.853	249.643	259.519	281.598	288.943	304.442	312.021
	35.322	41.306	50.297	60.045	68.054	79.669	90.039	100.897	112.686
	16.775	17.311	19.444	22.606	26.047	30.219	34.977	41.636	47.297

The Journal of Management and Engineering Integration
Vol. 3, No. 1, Summer, 2010

Factors that Influence Usage of Knowledge Management by Information Technology Professionals at Institutions of Higher Education

Ricardo A. Calix, Sri Abhishikth Mallepudi, Gerald M. Knapp and Isabelina Nahmens
Louisiana State University
rcalix1@lsu.edu, smalle@lsu.edu, gknapp@lsu.edu, nahmens@lsu.edu

Abstract

Information Technology (IT) staff are often distributed across various departments within an organization and have limited opportunity to share knowledge. Because the same problems may arise at different departments within an organization, Knowledge Management (KM) can make a significant contribution to reducing workload, minimizing cost and improving productivity. The purpose of this study was to explore the factors that influence usage of KM methodologies and technologies by IT professionals. A survey was used to collect information regarding the types of KM technologies currently in use, overall perceptions about KM, and usage of rewards to encourage KM. Fifty-one IT professionals at Louisiana State University (LSU) were surveyed. The results of the survey showed that unstructured data implementations of KM technologies (wiki's, blogs, etc) are used more than structured data implementations (i.e. databases). Of the fifty-one respondents, 59% used unstructured implementations, 24% used structured implementations, and 17% used nothing. Results from the survey also indicated that factors related to workload (Q2) have strong correlations to reward attitudes. Workload was correlated with salary increases (Q7), pay based on effort (Q9), and teamwork (Q12). Based on these results, guidelines for application of KM in IT organizations are proposed.

1. Introduction

The information technology practice in industry and in academia is a knowledge intensive profession that requires the application of current technological skills to support various departments within a given organization. Solving IT related problems requires specific knowledge of old and new technologies and practices, as well as the sharing and dissemination of information throughout the organization. Because the same problems may arise in different departments within an organization, approaches such as knowledge management can make a significant contribution to reducing workload, minimizing cost, and improving productivity as argued by Choo and Papoutsakis [10, 11].

Challenges of KM

According to Sinotte [3], motivation and trust are two of the key challenges that knowledge management faces. Another challenge is the lack of proper methods to measure the relation between knowledge management and performance [6, 18]. In Pedersen [14], the researchers argue that incentive mechanisms should be used to encourage the use of KM. This study tries to address these issues by asking the following questions:

- What types of technologies are currently used for Knowledge Management?
- What factors influence the adoption of KM methodologies and tools?
- Are KM methodologies encouraged at institutions of higher learning?
- Is workload a significant factor with relation to KM perceptions?

- Are KM efforts rewarded? If rewards are used, what are they?
- What are the overall perceptions IT professionals have towards KM?
- Do position and experience affect KM perceptions?
- Is there any relation between knowledge sharing factors and reward attitudes?

A survey was used to collect information regarding the types of KM technologies currently in use, overall perceptions about KM, and usage of rewards to encourage KM. Fifty-one IT professionals at Louisiana State University (LSU) were surveyed.

2. Literature review

According to Evans and Prusak [2, 7], Knowledge Management (KM) "involves the process of identifying, capturing, organizing, and using knowledge assets to create and sustain competitive advantage. Knowledge assets are accumulated intellectual resources such as information, ideas, learning, understanding, memory, insights, cognitive and technical skills, and capabilities."

There are many implementations of knowledge management. These can vary in the level of methodology and technology used. One common theme, however, in most KM implementations is that they all involve large amounts of information that are continuously growing and which need to be managed properly.

Who Uses Knowledge Management?

Most organizations can benefit from knowledge management. In Yang et al. [13], the authors argue that knowledge has a tendency to flow through the organization and that it is up to management to ensure that effective methodologies and technologies are in place to facilitate this flow. In particular, service oriented organizations can benefit from KM because their assets are accumulated intellectual knowledge about a particular industry which can be passed on to clients in the form of consulting services.

Service oriented organizations such as law firms, health care institutions, consulting firms, and universities view knowledge as a source of competitive advantage. Therefore, protecting and retaining knowledge can be very valuable. Examples of these types of organizations include Accenture, Google, McKinsey, the big four accounting firms KPMG, E&Y, Deloitte, PWC, and universities such as Louisiana State University (LSU). KPMG, for instance, has implemented Kworld [17] as a knowledge base which allows the organization to leverage its talents across different consulting practices and different locations. LSU uses the GROK system [16] for its IT services practice to disseminate information across the university.

Knowledge Management (KM) in Academia

KM in higher education is beneficial because it can increase efficiency and quality in learning, help to determine new ways to save money, and be more productive.

The IT practice in higher education is dedicated to maintaining and upgrading the IT infrastructure in a university [8]. Each department or school has its own IT staff which is responsible for its own systems. In general, most of the problems that IT personnel face are common across all departments or schools and can, therefore, be solved more efficiently through knowledge sharing.

How is Knowledge Stored?

Knowledge can be stored as structured or unstructured data [5]. Structured data implementations are those that rely on databases. Unstructured data implementations consist of loosely structured data written in a natural language format such as news paper articles, reports, websites, and blogs.

The current solutions to storing and accessing data are database management systems and information retrieval systems.

According to [3], digital libraries are a very important option which can be used to store knowledge. In Jurafsky [4], the authors propose the use of natural language processing, machine learning techniques and artificial intelligence to develop expert systems for KM. In general, organizations will make these systems available through the use of intranets or the internet [12].

Some of the most important types of stored knowledge include: (1) previous solutions to specific problems, (2) knowledge directories of who knows what (such as the Spree system [12]), and (3) expert systems that infer knowledge [3].

Knowledge Sharing and Reward Attitudes

Previous studies on the relation between knowledge sharing and reward attitudes have been conducted at other service oriented institutions such as law firms. One such study was conducted by Khandelwal [1]. In Yang [13], the authors tried to evaluate the effect of knowledge sharing in an organization and whether sharing can be obstructed. Their study found that regardless of payoff schemes, people will ultimately share their knowledge [13].

In [20], the authors conducted a survey study on the critical success factors (CSFs) for adopting KM in small to medium size companies. They ranked 11 factors that have been identified in the literature for KM adoption. Their results indicate that "management leadership and support" ranks most important for implementing KM. IT infrastructure ranks 8 and motivation aids, infrastructure, and performance measurement rank 9, 10, and 11, respectively.

In [19], the authors conducted an empirical study of Knowledge Management usage and incentives. They concluded that rewards are important (either monetary or professional development) to encourage usage and contribution to Electronic Knowledge Repositories.

Issues with Knowledge Sharing

In [18], the authors argue that there are several problems related to knowledge sharing. Among them are: high failure rate of IT project implementations, work effort, contributor's loss of power, etc. The loss of power issue is very important because it provides contributors with a powerful incentive not to share knowledge (e.g. a worker that shares knowledge is less valuable and more likely to lose their job). This means that rewards need to be put in place to directly offset the cost of sharing knowledge.

The authors in [18] argue that there is no conclusive evidence regarding the role of rewards in knowledge sharing. To address this issue, they conducted a study in a resource planning vendor to model the knowledge sharing problem and its relationship to rewards. They concluded that rewards should be used with proper evaluation mechanisms, and supported by information technology and organizational policy. Their conclusions are consistent with the guidelines proposed in this paper based on the survey results.

3. Methodology

A survey was used to conduct this study at Louisiana State University (LSU). The data was collected from a sample of fifty-one IT professionals at LSU.

A survey used in [1] was adapted and extended to conduct this study. The survey is divided into four sections. The first section addressed knowledge sharing factors, the second section addressed reward attitudes, the third section addressed the subject's overall impression of knowledge management and technologies, and the fourth section was used to collect user characteristics.

For most of the questions the survey used a Likert scale which ranged from 1 to 5. The scale is as follows:

1=Totally agree 2=agree 3=neither agree or disagree 4=disagree 5 = totally disagree

This study includes seventeen questions which can be seen in the following table.

Table 1. Survey questions

Knowledge Sharing Factors	
Q1	IT professionals are encouraged to share with others what they have learned from their recent projects
Q2	Senior staff are too busy to reflect on their experiences and share them
Q3	The organization has a well-organized system for sharing knowledge (e.g. about clients, managing projects, new approaches) within departments or practice areas
Q4	The firm has a well-organized system for sharing knowledge (e.g. about clients, managing projects, new approaches) across departments or practice areas
Q5	There is an expectation that IT personnel or their teams will have to take regular turns to provide a reflection on learning experiences
Q6	Sharing knowledge systematically is part of the organization's culture
Reward Attitudes	
Q7	IT staff salary increases in the organization are based on ability and how well he/she does his/her work?
Q8	Promotion of an employee in the organization is based on ability and how well he/she does his/her work
Q9	IT professionals are fairly rewarded for the amount of effort they put in
Q10	The interesting nature of the work compensates for long hours and a stressful workload?
Q11	The team as a whole is rewarded for good work
Q12	Teamwork in the organization is fully recognized and rewarded
Other	
Q13	Do you think mistakes are repeated because previous solutions to IT problems are not retained and stored?
Q14	Based on your experience, what do you think would be the most efficient way to store knowledge?
Q15	Do you currently use some type of knowledge management tool/methodology to support your work? If so, what?
Q16	Implementing a new knowledge management approach could increase your workload because of the added task of loading/updating information to the knowledge base?
Q17	Knowledge management tools can be valuable to my work.
IT professional background	
Position in the organization, Department, Years with the organization, Total years of IT experience.	

Analysis of Variance (ANOVA) and Spearman correlation analysis were conducted using SAS for survey analysis [9]. The correlation analysis is used to test the relation between knowledge sharing factors and reward attitudes. ANOVA tests are performed to determine if perceptions of Q17 are the same for all positions and/or levels of experience.

4. Results

Results from ANOVA and Spearman tests indicated that the relationship between reward attitudes and knowledge sharing factors is not strong for all cases (Tables 3 and 4). These results differ from those of [1] where their findings indicated stronger correlations between knowledge sharing and reward attitudes. This may suggest that knowledge management at LSU does not have the level of support that it does in other industries such as in [1]. The results also indicate that employee experience or position does not significantly affect KM perceptions (tables 6, 7, and 8). Respondent characteristics can be seen in table 2.

Table 2. Breakdown of respondents per position and years of experience

	Experience (< 10 years)	Experience (> 10 years)
Manager	2	8
Analyst	39	2

The analysis showed that workload (Q2) was a factor that presented a significant correlation with reward attitudes. Workload was correlated with salary increases (Q7: p-value 0.001), pay based on effort (Q9: p-value 0.02), and teamwork (Q12: p-value 0.05). The p-value is defined as the probability of obtaining the same or more extreme data assuming the null hypothesis of no effect [15]. Correlation is considered significant at $p<=0.05$. See Table 3.

A significant negative correlation was found between Q2 and Q9 with a correlation coefficient of -0.33. This can be interpreted by saying that: "senior staff members are too busy to reflect on their experiences and share them, and that this situation is negatively correlated to

the extent that IT professionals are fairly rewarded for the amount of effort they put in."

Table 3. Significant correlations between knowledge sharing and reward attitudes

		Q7	Q8	Q9	Q12
Q1	Corr.				0.28
	P-value				0.04
Q2	Corr.	0.42		-0.33	0.28
	P-value	0.00		0.02	0.05
Q3	Corr.			0.34	
	P-value			0.01	
Q6	Corr.		0.45		
	P-value		0.00		

The correlation analysis was performed between knowledge sharing factors (Q1-Q6) and reward attitudes (Q7-Q12). Additionally, a correlation analysis between overall KM perception (Q17) and factors and rewards was also performed (see Table 4).

Table 4. Significant correlations between overall KM perceptions and other factors

		Q1	Q8
Q17	Corr.	0.38	0.33
	P-value	0.01	0.02
Q13	Corr.	0.29	
	P-value	0.04	

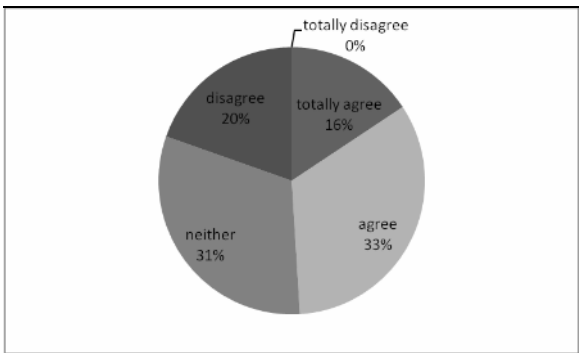


Figure 1. Knowledge sharing perceptions: "Do you think mistakes are repeated because previous solutions to IT problems are not retained and stored?"

Survey results also indicated that of the fifty-one IT professionals surveyed, 49% agree that mistakes are repeated because knowledge is not retained and shared, 20% disagree, and 31% were undecided (Figure 1). Based on the mean score for Q17 of 2.02, it can be seen that users have an overall fair perception of KM (Table 5).

Table 5. Mean response per question

Q	Mean	Std Dev	Min	Max
q1	2.12	0.93	1.00	5.00
q2	2.88	1.09	1.00	5.00
q3	2.45	0.90	1.00	5.00
q4	2.71	0.88	1.00	5.00
q5	2.76	1.01	1.00	5.00
q6	2.31	1.01	1.00	4.00
q7	2.45	0.94	1.00	4.00
q8	2.20	0.87	1.00	4.00
q9	2.45	0.88	1.00	5.00
q10	2.88	1.01	1.00	5.00
q11	2.51	1.12	1.00	5.00
q12	2.55	0.83	1.00	4.00
q13	2.55	0.99	1.00	4.00
q16	1.94	0.76	1.00	4.00
q17	2.02	0.65	1.00	3.00

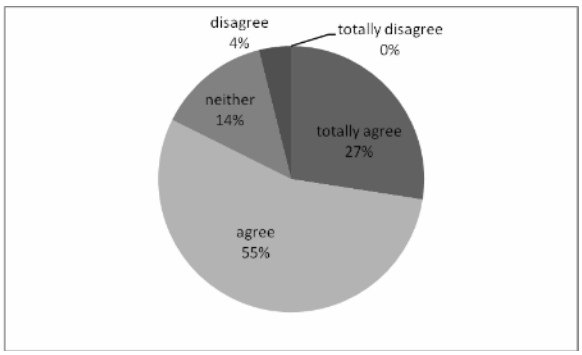


Figure 2. KM to workload perceptions: "Implementing a new KM approach could increase your workload because of the added task of loading/updating information to the knowledge base?"

With regards to workload from added KM responsibilities, results of the survey indicated

that 55% of respondents think that KM methodologies can increase their workload (Figure 2).

On the issue of KM technologies, 59% of respondents used unstructured implementations, 24% used structured implementations, and 17% used nothing (see figure 4). Additionally, more than half of all respondents think unstructured data implementations are more efficient ways of storing knowledge (see figure 3).

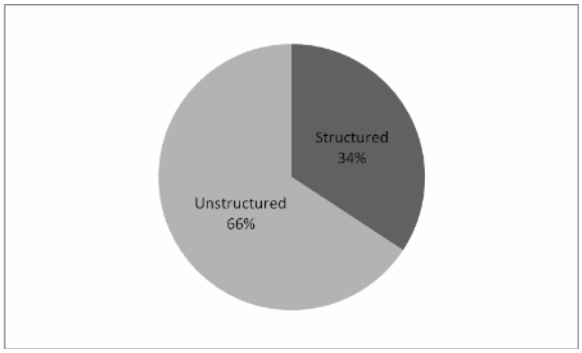


Figure 3. Perceptions of knowledge storage: "Based on your experience, what do you think would be the most efficient way to store knowledge?"

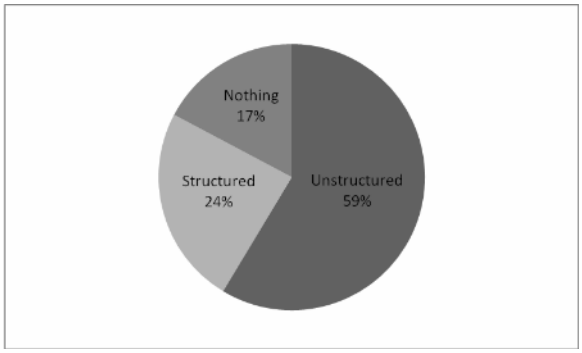


Figure 4. KM technologies currently in use: "Do you currently use some type of KM tool/method to support your work?"

Analysis

To address the question: "*Is there any relation between knowledge sharing factors and reward attitudes?"* a Spearman correlation test was performed. The Spearman correlation test is a nonparametric measure of how well a function can explain the relation between two variables. It is used when both variables consists of ranks such as those implemented when using a Likert scale. As an example of the findings of this study, it can be seen from Table 3 that Q2 and Q12 are related because their p-value is less than or equal to 0.05. The relation between Q2 and Q12 can be expressed in words as: "Senior staff are too busy to reflect on their experiences and share them is positively correlated to the extent that teamwork in the organization is fully recognized and rewarded."

To determine if experience and position within an organization affect KM perceptions in IT professionals, Analysis of Variance tests were conducted. Results of the tests indicate that there is no difference across levels of experience or position. In general across all levels, the null and alternative hypotheses were as follows:

- H0: Perceptions of Q17 are the same for all positions or levels of experience
- H1: At least one position or level of experience differs

Here Table 6 and Table 7 show based on the p-value that perceptions of KM across all positions are not significantly different and that the main effects for each independent variable are not significant.

Table 6. ANOVA (Q17 vs. multiple positions in the organization)

Source	DF	Sum of Squares	Mean Square	F Value	Pr> F
Model	1	0.08	0.08	0.19	0.67
Error	49	20.90	0.43		
Corrected Total		50.00	20.98		

Spearman correlation test

The Spearman test was conducted to compare the correlation between knowledge sharing factors and reward attitudes. The null hypothesis for the Spearman correlation test is as follows:

- Ho: The ranks of one variable do not covary with the ranks of the other variables.

Table 7. ANOVA (Q17 vs. two independent variables: position and IT Experience)

Source	DF	Type I SS	Mean Square	F Value	Pr > F
position	1.00	0.08	0.08	0.18	0.67
IT experience	1.00	0.01	0.01	0.03	0.87
Source	DF	Type III SS	Mean Square	F Value	Pr > F
position	1.00	0.011	0.011	0.03	0.87
IT experience	1.00	0.011	0.011	0.03	0.87

Based on the results it can be concluded that the null hypothesis does not hold for some knowledge sharing factors and reward attitudes. This means that these variables are correlated. In particular, there is correlation between the following questions: Q1 & Q12, Q3 & Q9, Q2 & Q7, Q2 & Q9, and Q2 & Q12.

Additionally, another important aspect is that there is a relationship between knowledge sharing attitudes, reward attitudes, and overall perception of the value of knowledge management. This can be seen by the relationships between: Q1 & Q17 and Q8 & Q17. The Spearman test was also conducted to compare the correlation between position, IT experience, and Q17.

Table 8. Correlation between position, IT experience, and Q17

		q17	position	IT experience
q17	Corr.	1.00	-0.06	0.07
	P-value		0.67	0.61
position	Corr.	-0.06	1.00	-0.62
	P-value	0.67		<.0001
IT experience	Corr.	0.07	-0.62	1.00
	P-value	0.61	<.0001	

5. Limitations

The study had a limited sample of 51 respondents. Additionally, this study had a limited scope since it only considered KM in academia, without considering other types of industries.

6. Conclusions and recommendations

Based on a sample of 51 IT professionals, this study found that (1) experience and position within an organization may not affect KM perceptions in IT professionals and (2) that workload is a significant factor that needs to be addressed when implementing a KM system.

Similarly to [19], results from this study identified a need for encouragement to use KM. Therefore, this study recommends that incentives should be used to encourage the use of KM under similar settings (e.g. IT professionals in academia). This is consistent with the assertion in [13] that management should facilitate the tendency of knowledge to flow through the organization. Incentives will offset other factors like workload from obstructing that flow.

The results of the study also indicate that most of the information stored in KM solutions is stored as unstructured data. This presents a justification and opportunity for new research in the fields of multimedia semantic analysis, NLP and Information Retrieval. Finally, some strategies for knowledge sharing in IT departments or organizations in an academic setting are presented.

Strategies for knowledge sharing in IT departments/organizations:

- Implement reward structures to encourage teamwork and sharing of experiences from recent projects (Q1 & Q12).
- Provide fair reward based on effort put in by IT staff and encourage them to reflect on experience and share knowledge (Q2 & Q9).
- Link performance evaluations to amount of effort put in and for knowledge sharing within departments (Q3 & Q9).
- Encourage IT senior staff to share their knowledge and mentor other members of the

staff. Link performance of mentees to senior staff rewards (Q7 & Q2).

- Promote sharing of experiences by senior staff as part of a good teamwork effort and reward appropriately (Q2 & Q12).
- Establish a systematic KM organizational culture as a knowledge sharing tool to add value to the work performed by all employees (Q1 & Q17).

7. Future work

This study can be extended in size and scope by considering larger samples across different industries. Research in techniques such as Natural Language Processing (NLP) and Multimedia Semantic Analysis to automate the process of loading and updating knowledge bases could be very productive. Other possibilities for future work are: studying the factors or problems with unstructured KM, comparing tools available for KM, studying ways to quantify cost to benefit ratio of KM, measuring how workload could be reduced by KM. Finally, metrics for KM usage based on Factor Analysis could be useful.

8. References

[1] V. Khandelwal, P. Gottschalk, "A Knowledge Management Survey of Australian Law Firms", University of Western Sydney, Technical Report, CIT/5/2003.

[2] J.R. Evans, W.M. Lindsay, "The Management and Control of Quality", South-Western College Pub, 6 edition, 2004.

[3] M. Sinotte, "Exploration of the Field of Knowledge Management for the Library and information Professional", Libri, vol. 54, no. 3, pp. 190–198, 2004.

[4] D. Jurafsky, J. Martin, "Speech and Language Processing", 2nd ed., New Jersey: Prentice Hall, 2008.

[5] C. Manning, P. Raghavan, H. Schütze "Introduction to Information Retrieval", Cambridge University Press, 1 edition, July 7, 2008.

[6] S. Shane, "KM does the job but ROI remains vague survey finds", Knowledge management, 27(7):1 & 8, 2001.

[7] L. Prusak, "Where did knowledge management come from?", IBM Systems journal 40(4): 1002–1007, 2001.

[8] H. Koch, D. Paradice, C. Bonsug, Y. Guo, "An investigation of knowledge management within a university IT group", Information Resources Management Journal, Vol. 15, No. 1, pp. 13-21, 2002.

[9] A. Gravely, "Your guide to survey research using the SAS System", SAS Publishing; 1 edition, 1998.

[10] C. W. Choo, "The Knowing organization: How organizations use information to construct meaning, create knowledge, and make decisions", Oxford University Press, 1998.

[11] H. Papoutsakis and R. Salvador Valles, "Linking Knowledge Management and Information Technology to Business Performance: A Literature Review and a Proposed Model", Journal of Knowledge Management Practice, Vol. 7, No. 1, 2006.

[12] F. Metze, C. Bauckhage, T. Alpcan "The "Spree" Expert Finding System", IEEE International Conference on Semantic Computing, pp: 551-558, 2007.

[13] H. Yang, T. Wu, "Knowledge sharing in an organization – Share or not?", IEEE International Conference on Computing and Informatics, 2006.

[14] M. Pedersen, M. Larsen, "Robustness of a Distributed Knowledge Management Model", IEEE Proceedings of the 37th Hawaii International Conference on Systems Sciences, 2004.

[15] J. Gibbons, J. Pratt, "P-value: Interpretation and Methodology", The American Statistician, Vol. 29, No. 1, 1975.

[16] Website: http://grok.lsu.edu/, Date: 3/22/10

[17] Website: http://www.kpmgcampus.com/why kpmg /knowledgesharing.shtml, Date: 3/22/10

[18] X. Zhang, Z. Chen, D. Vogel, M. Yuan, C. Guo, "Knowledge-Sharing Reward Dynamics in Knowledge Management Systems: Game Theory–Based Empirical Validation", Human Factors and Ergonomics in Manufacturing & Service Industries, 20 (2): 103–122

[19] A. Kankanhalli, B. Tan, K. Wei, "Understanding Seeking from electronic knowledge repositories: An empirical study", Journal of the American Society for Information Science and Technology, 56(11), 1156–1166.

[20] K. Wong, E. Aspinwall, "An empirical study of the important factors for knowledge-management adoption in the SME sector", Journal of Knowledge Management, Vol. 9 Iss: 3, pp.64 – 82, 2005

The Journal of Management and Engineering Integration
Vol. 3, No. 1, Summer, 2010

A Comparative Study on HCN and CO_2 Gas Lasers

Jeng-Nan Juang and R. Radharamanan
Mercer University
juang_jn@mercer.edu; radharaman_r@mercer.edu

Abstract

This paper presents a comparative study on HCN and CO_2 lasers. It also provides insight into the operating principles and distinctive differences of these two gas laser. The far-infrared (FIR) submillimeter wave laser is especially useful in applications requiring good spatial resolution. Its ability to obtain spatial resolution makes it possible to scan across the plasma. Hydrogen Cyanide (HCN) is the FIR gas laser that has been utilized in such applications as plasma diagnosis in nuclear fusion. The HCN gas laser operates at a wavelength of 337 μm as compared to the CO_2 gas laser which operates at a wavelength of 10.6 μm. CO_2 lasers on the other hand, produce the highest output power known among all gas lasers. CO_2 gas lasers can be discharged by a number of different methods. The focus of this study is to compare these two types of gas lasers and their primary application areas since HCN laser is more stable compared to other lasers and the output of power of the CO_2 laser is much higher than other lasers.

With a gas laser, the active mediums are either enclosed in a tube or glass envelope. In the experiments, a Pyrex tube is commonly used that varies from a fraction of a meter to several meters. The tube length is very sensitive to the output power, as had been discussed in detail by several authors. Laser beams from gas lasers are available at a variety of wavelengths, including the emission of infrared and ultraviolet light.

1. Introduction

This paper presents a comparative study on HCN (wave length 337 μm) and CO_2 (wave length 10.6 μm) gas lasers. The study is focused on the operating principles and distinctive differences of these two gas lasers [1].

There are three basic elements to a laser system. There is an active medium which may be a collection of atoms, molecules, ions or semiconducting crystal. There is the excitation (or pumping) process which excites these atoms (molecule etc.) into higher quantum-mechanical energy levels. Finally, a source of feedback elements serves as an oscillator (or amplifier).

The active medium can be any type of material that absorbs and releases energy. An energy source is required to discharge the medium into a stage of absorbing and releasing of energy. A laser tube (or cavity) is required to contain the photons that are generated by discharged atoms. A set of gold or copper coated plates or mirrors are required for directing the photons. These plates can be installed internal or external to the laser cavity. In the actual lasing process, the laser material is placed into the container, where a source is used to stimulate the active material into the emission of light waves. The laser cavity must be terminated at each end with a window. These windows are necessary to minimize optical losses. In directing the laser beam, the laser cavity is constructed with two mirrors (copper or gold coated plates). One of the mirrors must have a perfect reflectance of 100% and the other can have a slightly lower reflectance of 90% [2].

In the discharge process, the atoms are discharged into their higher quantum-mechanical energy levels. This pumping process must create a condition of population inversion, where more atoms are in the higher levels of the quantum-mechanical states than the lower levels. This is referred to as "Metastable State" or intermediate energy level, where atoms pause for awhile prior to returning to the ground state.

2. Background: CO_2 and HCN lasers

The development of these gas laser systems is based upon the detailed knowledge of the electron distribution function. The excitation cross section, energy transfer rates, and quenching cross section from all the excited levels are all essential elements to a laser system.

The main development of infrared lasers is CO_2 lasers. The first Continuous Wave (CW) laser in action in a CO_2 gas discharge was achieved by Patel in 1964. As for Far-Infrared (FIR) lasers, it has been the long-standing, and an increasingly used technique for plasma diagnostics. The use of FIR lasers have been extended into the field of astronomy and metrology [3].

The CW FIR source exists for wavelengths in the 100 μm to 500 μm range, and this light emission is emitted in one steady beam, usually of relatively low power.

Patel's research, on CW output power of CO_2 laser at 10.6 μm wave length, has been raised from milliwatts-power level to 10 W in 1965. A kW/m of CW power, at 10.6 μm in cross flow, with transversely excited CO_2 laser, was achieved by Tiffany, Targ, and Foster in early 1970 [4]. Today, the research on CW CO_2 gas laser has produced an output power of thousands of watts. The drawback with the high output power of CO_2 laser is the thermal instability. Modes of thermal instability can occur for conditions typical of molecular laser discharges. Also, the thermal instability can be caused by the high power density levels in the laser discharges and the absence of effective dissipation mechanisms, as well as the volume-dominated molecular laser discharges. The instability can also occur as a consequence of negative-ion production, which is commonly referred to as electron attachment-induced ionization instability, or simply attachment instability [5].

Because of the occurrence of instabilities in the tokamak plasma, a far-infrared (FIR) submillimeter wave laser, with a good spatial resolution, is especially useful. This makes it possible to scan across the plasma. Ideally, such a source should be tunable, but an acceptable alternative would be a source with a number of narrow-line and at various points in the FIR.

Ion temperature measurement in a tokamak plasma machine is most difficult, but is required. A high power submillimeter laser can be a unique tool for performing this kind of measurement.

Ion components and electrons of "Thomas scattering" are determined by α,

$$\alpha = [\lambda_o / (4\pi\lambda_d \sin(\theta/2))] \qquad (1)$$

where: θ is the scattering angle and λ_d is the Debye length. The electron dominates in the scattering when α is less than one. When α is greater than 1, this implies that the ion scattering component dominates.

In order that the 10.6 μm CO_2 laser radiation can be useful for this measurement of ion temperature, the scattering angle would have to be less than 1 degree. Measurement is extremely difficult when dealing with such small angles.

The initial studies on the waveguide characteristics of longitudinally excited submillimeter-wave HCN lasers were performed by Schwaller et al. in 1967. Schotzau and Kncubuhl in 1974 demonstrated chemical

excitation and suppression of higher dielectric-waveguide modes in a 337 μm HCN laser. This could be achieved by variation of the composition of the laser-active gas mixtures CH_4: N_2, C_2H_6:N_2, C_2H_4:N_2, C_2H_2:N_2, and H_2 without any changing of other laser parameters.

Characteristic designs of pulsed and CW longitudinally excited waveguide 337 μm HCN laser and their performance have been reported by Turner and Pochler in 1971, Sharp and Wetherell in 1972, and Bicanic and Dymanus in 1976. According to the knowledge gained from experiments, visits to national laboratories, literature review, and communications, the maximum performance of this type of laser is about 0.3 W CW power, and 0.1 KW pulsed power.

3. Comparison of CO_2 and HCN lasers

Wavelength and efficiency comparison for HCN and CO_2 lasers: HCN is a submillimeter gas laser that operates at the wavelength of 337 μm, as opposed to the CO_2 gas laser that operates at 10.6 μm wavelength (Table 1).

Table 1. Wavelength and efficiency of CO_2 and HCN lasers

Laser Type	Wavelength	Efficiency
CO_2	10.6 μm	5-15%
HCN	337 μm	3%

Experiments have revealed that HCN laser inherent strongest transition at 337 μm. It is at this particular transition that the highest pulsed and CW output power have been reported with thousands of watts of pulsed output power, and in excess of 200mW of CW power.

In terms of efficiency, HCN efficiency is much lower, by the ratio of the far-infrared to the infared wavelength, not to mention the efficiency of the infrared pump laser.

Early reports of CW laser action involving the CO_2 gas discharged were around 10.4 μm and 9.4 μm made by Patel. Because of its many transitions in the 9.3 and 10.6 μm region, it is useful for the linear and nonlinear excitation of molecule vibration frequencies. The CO_2 laser can be operated at efficiencies in the 5 to 15 % range [5].

3.1. HCN and CO_2 gas lasers applications

HCN laser applications: HCN lasers, having a strong oscillation line at 337 μm, have been used as a powerful oscillator at this wavelength. They have also been regarded as a powerful signal source along with the development of applications on submillimeter waves.

The HCN laser is of particular importance when applied to plasma diagnosis in nuclear fusion. There is a vast increase in the pursuit of astronomy in this part of spectrum and in the use of far-infrared in metrology. A high power submillimeter HCN laser can be used as a unique tool for performing this kind of measurement.

CO_2 laser applications: Because of the CO_2 laser flexibility for pumping techniques, it is the most efficient laser system and produces the most output power. CO_2 gas lasers have many diversified applications. They have been used for many high-power applications such as welding, weapons, fusion, and spectroscopy.

In addition, there has been increasing interest in CO_2 lasers for a variety of military applications. These applications include range findings, laser radars, laser beam rider missile guidance, and laser communications [6].

3.2. HCN and CO_2 lasers: DC-excited versus RF-excited

A research experiment from Masao Makiuchi and Mitsuo Kawamura has revealed that the output power for the RF-excited discharge is larger than that of the DC-excited discharge. Furthermore, the input power, as was discovered, is smaller in RF-excited discharge as opposed to the DC-excited type that required a higher input power [2] (Table 2).

Table 2. Comparison of input and output power: DC-excited vs. RF-excited

	Input Power	Output Power
RF-Excited Discharge	Low	High
DC-Excited Discharge	High	Low

There are a number of reasons that the RF-excited discharged is preferred over the DC-excited discharge. First, the RF-excited discharge laser tube is simple, where no electrodes are required. Secondly, no resistor is needed for stability of the discharge. Third, because it is electrode-less, this reduces the maintenance requirements. Finally, the efficiency of the laser tube is higher, in that there is no power dissipation in the cathode dark space.

RF excitation can achieve maximum efficiency when the experimental conditions are arranged, such that the highest possible fraction of the power generated by the RF power supply is dissipated in the laser discharge. To achieve maximum power transfer, the impedance of the laser structure including the discharge, should be matched to the generator. This is accomplished when the impedance, seen by the source looking into the load, is a complex conjugate of impedance seen by the load looking into the power generator. Other advantages when utilizing RF excitation operations are:

1. The elimination of the high voltage source.
2. Offers a laser output with high electrical conversion efficiency.
3. RF excitation tubes are less expensive.
4. The amount of artificial cooling is less than low efficiency operation.

3.3. HCN and CO_2 discharge operation

HCN pulsed operation: The operational characteristics of pulsed excitation are dependent on the discharge conditions. The conditions are a combination of gas pressure and compositions of the gas; flow rate; pulsed energy input; length and amplitude of the current excitation pulse; and wavelength of the laser emission. The power to be obtained is dependent on the resonator geometry, and output coupling as well as the discharge conditions that affect the basic operating characteristics. The results from experiments further indicate that the output power is dependent on either the gas pressure or the flow rate.

From the Sharp and Wetherell, Kon et al, the oscillation occurs over the pressure range of 0.2 to 1.2 torr, with the range of optimum located between 0.6 and 0.8 torr, and the optimum flow rates are between 1.1 and 2.01cc/min, with a discharging tube of 15 cm bore.

With the same composition of CH_4 and N_2 mixture with 1:1 ratio, the output power of the pulsed 337 μm HCN laser is affected by the discharge voltage and charge capacity. At high input energy of 90 J and a charge voltage of 16 KV applied across a discharge length of 3.4 m, and the total pressure of 0.7 torr, the output power maximizes at 10 kV [4].

HCN CW operation: CW operation characteristics of the HCN laser are much less involved than those observed in the pulsed operation. The output power and the gain have been found to be dependent on the composition of the excited gas, gas pressure, and discharge current.

Experiments have revealed that the best mixture for high gain and high output power, coupled with cleanliness of operation, is with $CH_4 + N_2$ mixture [4].

CO_2 pulsed operation: The output power of pulsed operation CO_2 lasers is considerably higher than those obtained from a CW CO_2 laser. Power in excess of megawatts can be easily obtained in pulsed operation of CO_2 lasers. The number of researchers has indicated that higher pressure rates results in higher output power [7].

CO_2 CW operation: Much discussion and emphasis of the 10.6 μm CO_2 laser have been on CW operation at low gas pressure, up to approximately 10 torr of total pressure, and on processes that are operative at low pressure, which bring about population inversion.

A significant achievement in the CW CO_2 laser was made by Tiffany, Targ, and Foster. In their research, they determined that the maximum CW 10.6 μm output power that could be obtained from 1 meter of discharge was limited heating of the gas to about 200W. However, by using electrical excitation transverse to the laser axis and rapidly flowing the CO_2-N_2-He laser mixture across the discharge region, Tiffany et al were able to demonstrate that more than 1 kW could be obtained from 1 m of discharge at a 13 kW of electrical discharge input power. With a relatively high pressure of 10 to 120 torr, high-speed flowing CO_2 lasers, in which the gas flow at several hundred meters per seconds, a high CW output of 140 W was obtained from a tube that is 10 cm long [8].

3.4. HCN and CO_2 gas lasers additives

Addition of N_2 and He to CO_2 laser: Addition of N_2 allowed Patel to increase the average power of his device from the 1 mW level to 11.9 W with a 3% conversion efficiency of dc electrical discharge power into coherent output power.

It was suggested that He was more important than N_2 because higher powers were obtained with CO_2-He mixtures than with CO_2-N_2 mixtures. It was suggested and proposed by Moeller and Rigden that He serves as a buffer for the CO_2 molecules, thus reducing the number of excited molecules diffusing to the wall and becoming quenched [9].

Patel also suggested the role of He might be to transfer energy from the He metastable levels to the N_2 molecule. It is later believed that the role of He is to cool the gas by aiding in the depopulation of a lower CO_2 vibration energy level that acts as a bottleneck for the effective transferring of molecules down to the ground state without significantly interfering with the upper laser level. He also has an equal important role in maintaining the energy distribution of electrons within the discharge in the proper range for more efficient excitation of CO_2 molecules [10].

Additives to HCN laser: Oscillation in HCN has been obtained under both CW and pulse discharged conditions in mixtures of a whole multitude of chemical components. Such mixtures include: CH_4+NH_3; CH_4+N_2; CH_3CN; $(CH_3)_2NH$; HCN or ICN (pure impurities); and CD_3CN; CH_2H_5CH; $(C_2H_5)_2O$+N_2; CH_3OH+N_2; C_2H_5OH and $(C_3H_7)_2O$+N_2. A lot of these mixtures have a serious drawback in that tar-like deposits can be caused by the polymerization of some of the reaction products of the discharge, quickly degrading the discharge tube. The best mixture found among these is, the CH_4+N_2 combination. As a result, its use gives the cleanest discharge products, and has been used in experiments to get high output power (of the 337 μm line) under pulsed and CW conditions.

A study for improving the CW power has revealed that the dymethylamine-ammonia 1:1, C_6H_5N $(CH_3)_2$-NH_3 has the highest output power among forty different gas mixtures; each containing vapors of one, two or more of the following constituents: ammonia, HN_3, dimethylamine, $C_6H_5N(CH)_2$, methane, CH_4, nitrogen, N_2, carbon monoxide, CO, ethylene, C_2H_4, acetylene, C_2H_2, and diethylether, $(C_2H_5)_2O$ respectively.

The working pressure for the laser tube were approximately 1 torr and the flow rate of 111 cc/min of air when reduced to standard temperature. The external temperature of Duran 50 laser tube (2.5 mm think walls), measured by means of a bimetal thermometer placed under the protective buckler was 105°. This experiment produced the most significant result, which

increased by 30% as compared to the power levels accessible with NH_3-CH_4, 4:1 blend.

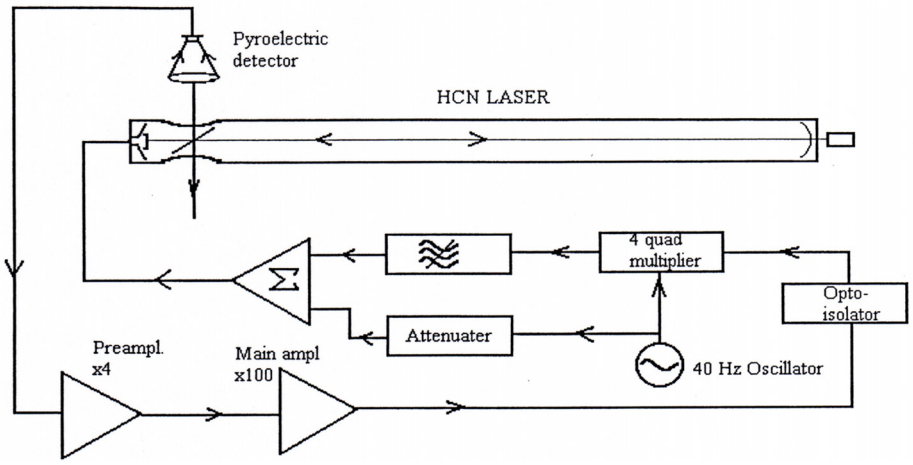


Figure 1. Block diagram of the servo loop

3.5. Stability of HCN and CO2 gas lasers

Methods to stabilize HCN gas laser: In 1982, a servo system was developed for cavity stabilization of a far infrared HCN laser operating at 337 μm (Figure 1). This technique does not require an accurate temperature control of the laser cavity.

The laser cavity is stabilized by locking to the peak mode peak using cavity modulation technique. This was done by using a hemispherical laser cavity configuration with the lighter 3" diameter plane mirror supported in the cone of a 5" loudspeaker. The servo system is used to power the loudspeaker to correct the changes in the cavity length. A small AC modulating voltage is applied to the speaker to obtain cavity modulation. The modulation voltage applied to the loudspeaker causes modulation of the laser output. This detected modulation is amplified and applied to a phase sensitive detector, along with the modulating signal, which acts as the reference. The correction voltage generated by the phase sensitive detector is fed back to the loudspeaker along with the modulating signal.

Frequency stability of CO2 gas laser: Having a homogeneous line width and an operating pressure above 5.2 torrs, the CO_2 laser is well-suited for relatively high-power, single-frequency and high-frequency stability applications. Frequency-stable He-Ne lasers typically operate in the tens of microwatts range whereas frequency-stable CO_2 lasers operate in the range of fractions of a watt, up to mega-watts. The success of frequency stability, coupled with high power and efficiency, excellent atmospheric propagation characteristics and the ease of heterodyning because of the relatively long (i.e., 10 μm) wavelength, make the CO_2 laser the prime candidate for "radar-like" application.

3.6. Characteristics of HCN/CO_2 gas lasers

HCN gas laser oscillation characteristics: Among the thirty-eight laser lines ranging from 12.85 μm to 773.3μm, it has been reported and allocated to HCN, DCN (deuterium cyanide), and HCN^{15}. Twenty-eight are believed to be HCN laser lines; six DCN lines; and four HCN^{15} lines [4]. The most efficient line is located about 337 μm, with the highest peak power and CW pulsed output power as indicated on Table 3.

HCN lasers have a strong oscillation line at 337 μm wavelength that is why it has been

regarded as a power signal source along with the recent development of application research on submillimeter wave.

Table 3. Strongest laser lines of HCN laser between 12.85 and 774 μm

Measured wavelength λ_{vac} (μm)	Vibrational Transition	Rotational Transition	Power	
			Pulsed (W)	*CW* (mW)
98.693	-		0.8	-
116.132	-		0.5	-
126.164	(12^20)-$(05^10)^a$	R(26)	3	-
128.629	(12^20)-$(05^10)^a$	R(25)	9	0.2
130.834	(12^00)-$(05^10)^a$	R(25)	4	-
134.932	(12^00)-$(05^10)^a$	R(24)	0.8	-
309.908	(11^10)-$(11^10)^b$	R(10)	0.4	-
310.908	(11^10)-$(04^00)^b$	R(10)	140	3
336.579	(11^10)-$(04^00)^b$	R(9)	1000s	260
372.547	(04^00)-$(04^00)^b$	R(8)	0.6	-

Effective CO_2 oscillation of a CW laser: The Co_2 gas laser oscillation under CW conditions was first reported in 1964. Shortly after, Patel reported CW and pulsed oscillation on 14 lines in the CO_2 in the wave length range from 10.5135 to 10.7880 μm, and seven lines in the vacuum wavelength (9.5691 to 10.7880 μm). The strongest of all oscillation was found to be at 10.6 μm, with highest power level output.

3.7. Capacitively versus inductively coupled RF-excited CW HCN gas laser

In an experiment with a capacitively coupled, RF excited, CW HCN submillimeter wave laser was investigated by Kawamura, Okabyashi and Fukuyama, on a 260 mW continuous wave at 337 μm using a 1 m length and 55 mm in diameter.

Three electrodeless discharges were investigated in this experiment: L-Coupled, TERF, and LERF (Figure 2). In the case of C-coupled, RF energy-supply system can be easily achieved, with a sufficient RF power supply.

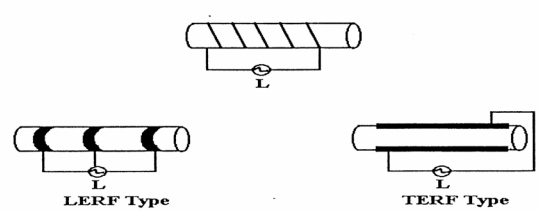


Figure 2. Coupling types

As referenced previously, a capacitively coupled RF-excited CW HCN laser and an inductively coupled RF-excited CW HCN laser generated output power levels of 260 mW and 160 mW, respectively (Table 4) [2].

For military applications, CO_2 lasers have high electrical efficiency. They are relatively compact and are scalable to large powers and/or pulse energies (CW power from 1 W to 10 kW

or high pulse energies from 1 mJ to 10 kJ has been demonstrated).

In the performance of conventional rangefinders applications, the existing mini-TEA CO_2 laser is adequate for these applications. They have output power of the order of 250 kW to 800 kW.

Table 4. Principal gas lasers and their characteristics

Type	λ	Pumping	Efficiency	Power	Specify energy density	Uses
He, Ne	632.8nm	Discharge	<0.1%	1-10mW	10^{-3}J/liter	Surveying, code scanners, alignment, memory readout
Ar^+ (Argon Ion)	Several	Discharge	<0.1%	10mW-20W	10^{-3}J/liter	Photochemistry, dye laser pumps, spectroscopy
Iodine	1.32 μm	Photon	0.5%		30 J/liter	Possible fusion laser, high energy
CO_2	**9.4 μm 10.6 μm**	**Discharge Chemical Aerodynamic E-beam Sustainer**	**5-15%**	**>100kW**	**15 J/liter**	**High-power applications, welding, weapons, fusion, etc. also spectroscopy**
HCN	**337 μm**	**RF Discharge**	**3%**	**260mW**		**Measurement and Plasma Diagnosis**

Furthermore, cavity-dumped pulses from a 45 cm long device have obtained 2 kW output power, and with a frequency up to 100 kHz.

4. Conclusions

The CO_2 laser field has grown tremendously since its beginning, and is continuing to expand at such a fast rate that it was not possible to cover all aspects in this study. Thus this study is an attempt to make correlation to the HCN laser. As a result, they both share some basic operational characteristics in area of the RF-discharge, versus DC-discharge, and pulse, versus CW operations. With the realization of additive N_2 in various experiments, both systems experienced a greater output power and cleaner discharges with less residue deposit.

As compared to the CO_2 laser field, the HCN laser field is relatively limited. In today's research on submillimeter wave lasers, the main focus is on the HCN laser and its operational characteristics to obtain higher output power. In this region, HCN has the strongest oscillation and has the most efficiency. Compared to the improved CW's output power and stability, the HCN is much better for the existing applications and is a potential candidate for future applications in the fields of astronomy and metrology.

5. References

[1] J. Juang and R. Radharamanan, "Experimental Investigation on Radio Frequency Excited Compact Continues Wave Hydrogen Cyanide Gas Laser", The Journal of Management and Engineering Integration, Winter 2009/10, Vol. 2, No. 2, pp. 75-80.

[2] J. Juang, "The RF-Excited Compact CW HCN Waveguide Gas Laser", Research Report, Tennessee Oak Ridge National Laboratory, 1986.

[3] N. G. Douglas, "Submillimeter Wavelength Laser", Kaptyn Sterrewacht, Rijksuniversityit te Groningen, NL-9301 KA Roden, Netherlans Feb. 1980.

[4] C. S. Willetts, "Introduction to Gas Laser", Pergamon Press Inc., Fairview Park, Elmford, NY, 1974.

[5] Bekefi, G., "Principal of Laser Plasma", John Wiley & Sons, Inc., New York, 1976.

[6] J. D. Evans, and E. V. Locks, "Comparison of Techniques for Long-Range Laser Ranging", SPIE, 1989, Vol. 1042, pp. 42-54.

[7] S. Ya. Tochitsky, C. Filip, R. Narang, C. E. Clayton, K. A. Marsh, and C. Joshi, "Light coupling between two parallel CO_2-laser written long-period fiber gratings" Optics Letters, 2001, Vol. 26, Issue 11, pp. 813-815.

[8] X. S. Zhang, H J Baker and D R Hall, "Flowing gas operation of a planar discharge RF excited CO_2 laser" Journal of Physics D: Applied Physics Create an alert RSS this journal, 1993, Vol. 26, No. 3, pp. 359-366.

[9] C. K. N. Patel, "American Institute of Physics", Reprinted from Applied Physics Letter, December 1965, Vol. 7(II), pp. 290-292.

[10] A. Kasapi, G. Y. Yin, and M. Jain, "Pulsed Ti: Sapphire Laser Seeded off the Gain Peak" Applied Optics, 1999-2004 (1996), Vol. 35, Issue 12, pp. 1999-2004.

Service Optimization in the Fast Food Industry: The Case of the Pizza Delivery Service

Ahmad D. Rahal
University of Arkansas-Fort Smith
arahal@uafortsmith.edu

Nabeel Yousef
University of Central Florida
nyousef@mail.ucf.edu

Abstract

In the fast food delivery services, firms compete on their abilities to deliver product value along with fast service to minimize the waiting time experienced by their customers. This holds especially true during tough economic climates where consumers experience spending power contractions.

The recent *economic* downturn and the fierce competitive nature of their business, has raised concerns about the current abilities of a local pizza delivery franchisee to retain its customers. In an effort to protect their customer base and improve their service level, the store manager sought the authors' assistance to assess and recommend the necessary improvements to minimize the mean order delivery time to about twenty five (25) minutes, and reduce the call loss rate (busy signal experienced by customer) during peak business hours to less than one (1) percent of the total call.

System data analysis and queuing theoretical models were used and calibrated to balance between the service level requirement and the efficient utilization of the system. It was determined that five (5) phone lines would be sufficient to achieve the call loss objective of 1%, while a staffing level of nineteen (19) drivers instead of the fifteen (15) currently scheduled, would minimize the mean delivery time to the intended goal of about 25 minutes.

1. Introduction

Customer satisfaction, customer loyalty, and service quality have been shown to be positively related to a firm's performance (Parasuraman et al. 1985, Buzzell and Gale, 1987). An increase in customer satisfaction reduces the risk of customer defection (Anderson and Sullivan 1993), positively influence customer repurchase intention (Bolton and Drew 1991), leads to greater customer retention (Mittal and Kamakura 2001) and loyalty (Anderson and Sullivan 1993; Bearden and Teel 1983; Bolton 1998; Bolton and Drew 1991; Boulding et al. 1993; Fornell 1992; Fornell et al. 1996; Fornell and Wernerfelt, 1987; Mittal, Ross, and Baldasare 1998; Oliver 1980; Patterson, Johnson, and Spreng 1997; Rust and Zahorik 1993).

When deciding where to spend their hard earned dollars, customers are looking for better values, and are increasingly demanding reliable and responsive supplier with fast deliveries services. This view holds especially true for the Pizza delivery business where customers are bombarded with offers and promises of fast delivery.

The success of a pizza delivery business depends on its ability to provide a fast and a high level quality service to satisfy and retain its customers. This can be accomplished by minimizing their customers' waiting time while maintaining efficient scheduling and utilization of its human resources (delivery drivers).

For a local pizza delivery franchisee, Fridays between 5:00 and 8:00 PM represent the weekly peak business hours where the order arrival rate is sometimes more than double the regular daily order rate. To improve customer satisfaction and retention, the authors' were asked to assess the current staffing levels and the number of the existing phone lines, and recommend the necessary improvements to minimize the delivery time to a mean of twenty five (25) minutes, and the call loss rate (busy signal) to less than one (1) percent.

1.1 Operational Process of Pizza Anonymous

For a pizza delivery business, customer orders are tracked throughout the process using a computerized system as described below:

I. Customers' orders are keyed in by operators where a record of an arrival time t_g is generated.
II. Orders are then routed to the pizza maker and topper, and then to the loader who loads the pizza in the conveyor oven and logs the time of loading t_l.
III. A fixed eight (8) minutes trip in the conveyor oven is required to cook the pizza pie, after which it will be picked up by the oven operator who slices and places it in the hot holding rack 's queue to be picked up by the delivery driver at time t_p
IV. Upon returning to the store, the driver logs in his arrival time t_a, and the duration of the delivery time is assumed to equal half of the time difference between the pickup time t_p and his return arrival time t_a.

2. Customer Orders Arrival Process

Analysis of the company's archived data yielded that for the time frame in question (Fridays 5:00 to 8:00 PM) fluctuations in the intensity of the incoming orders of most Fridays were within the five (5) percent margin concluding that a randomly chosen data can be assumed to be a good representative of the system.

The analysis of a randomly chosen data yielded that for that for the three hours period, 154 delivery orders were received (see Figure 1) where each order may consist of one or multiple pizza pies. It took a total of 9.16 minutes for an order to reach the delivery waiting queue (an average of 1.16 minutes to load the order in the oven and a fixed 8 minutes to bake).

Statistics for the delivery queue clearly shows that about 77% of the deliveries exceeded the twenty five minutes limit set by management as shown in Figure 2.

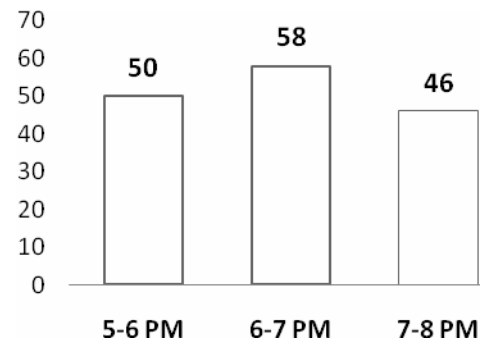


Figure 1. Hourly orders arrivals Rate

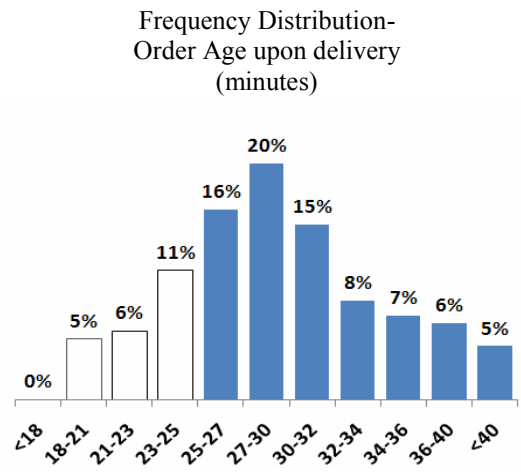


Figure 2. Orders' age frequency distribution.

2.1 Goodness of Fit Tests

Although Poisson can be assumed to be a plausible model for the customer orders arrival process, the calibration of the queuing models requires the confirmation of such an assumption, hence the use of the graphical and statistical goodness of fit tests (Hall, 1991).

a. Graphical Goodness of FIT Tests

The data set was plotted and examined for patterns, where the absence of a pattern is generally an indication that the arrival process follows a Poisson distribution. Due to small deviation above and below the diagonal line in a fairly random fashion, the

arrivals can be assumed to be stationary (Figure 3).

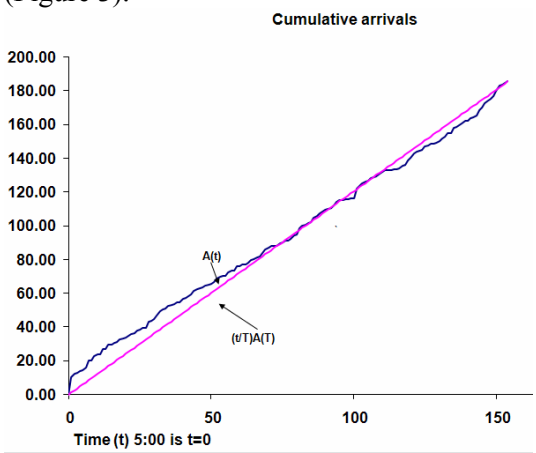


Figure 3. Cumulative arrivals of orders

The inter-arrival times (Figure 4) were randomly scattered about the average inter-arrival time, and did not show any sign of an existing cyclical pattern.

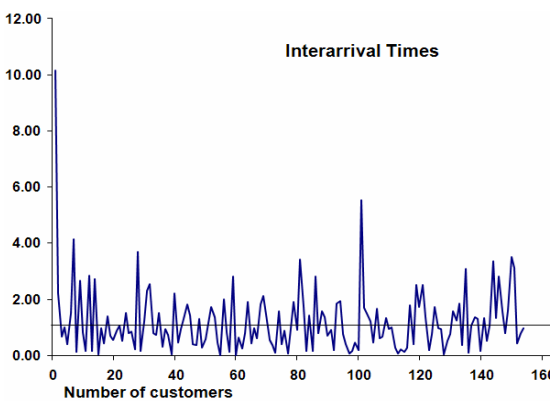


Figure 4. Inter-arrival times (minutes) vs. # of orders

Testing the data points for independence, the X_n, vs. X_{n-1} for n=2, 3, etc… (Figure 5) plot; where X_{n-1} represents the number of arrivals "n" minus 1; failed to detect any visible patterns thus concluding that orders arrivals were independents of each other.

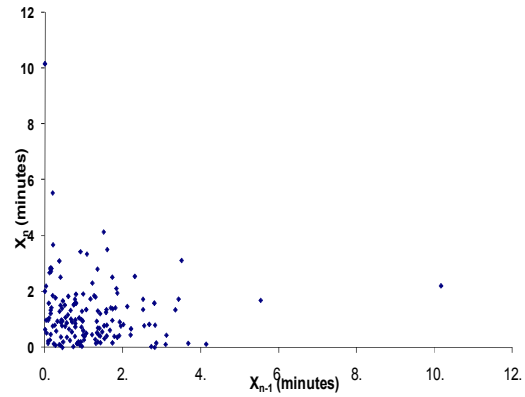


Figure 5. Paired Successive Inter-arrival times (minutes)

For a Poisson process, as is the case in Figure 6, no visible patterns of the deviation should exist between the empirical probability distribution for the inter-arrival times and the theoretical exponential distribution with λ =A (T)/T where T is the length of the observed interval.

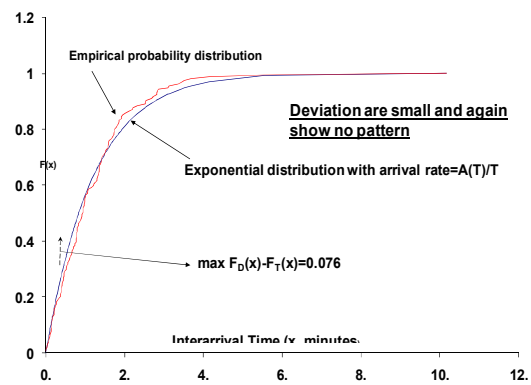


Figure 6. Inter-arrival Time (X, minutes)

As a conclusion, the use of the graphical tests did not provide any supporting evidence reason to believe that the arrival process does not follow the Poisson distribution, hence the assumption that the arrival process conforms to the Poison distribution.

b. Statistical Goodness of Fit Tests

To alleviate any doubts that might still exist about the validity that the arrival process follows the Poisson assumption, the

hypotheses of the inter-arrival times having exponential probability distribution and dependency were tested.

I. Hypothesis 1: The inter-arrival times have an exponential probability distribution

A Kolmogorov-Smirnov (K-S) test based on the maximum deviation between the empirical distribution function FD(x) and the theoretical distribution function FT(x) was used (see Figure 6) to test whether or not the data set conforms to the theoretical probability distribution (Mendenhall, and Sincich 2006). A small deviation would indicate the validity of the model, while a large deviation would negate such an assumption (Hall, 1991). Data comparison for the two distributions, showed that the maximum deviation over the entire range of values, including X equal "0" between the two distributions [FD(x) - FT(x)] to be 0.076 concluding the hypothesis would not be rejected at the 1% confidence level, hence the inter-arrival times can be assumed to have an exponential probability distribution.

II. Hypothesis 2: The inter-arrival times are independent

The hypothesis states that correlation between independent events X_n and X_{n-1} must have a correlation coefficient of zero, although the opposite may not hold true. To test such a hypothesis, the correlation coefficient $r_{xy;}$ defined as the ratio of covariance to the product of the standard deviations in the below equaled -0.123.

$$r_{xy} = \frac{\sum_{n=1}^{N}(X_n - \overline{X})(Y_n - \overline{Y})}{(N-1)S_x S_y} = -\ 0.123$$

For random variables to be independent and have a normal distribution, the statistic t (see equation below) with N-2 degrees of freedom would be used to approximate for the exponential case (Hall 91).

$$t = \frac{r_{xy}\sqrt{N-2}}{\sqrt{1-r_{xy}^{2}}}$$

With 152 Degrees of freedom (154 data points minus 2) the t-statistics of 1.61 would be rejected at the 1% confidence level, thus concluding that the inter-arrival times are most likely independent.

III. Hypothesis 3: Arrival times conform to the uniform distribution

For a stationary arrival process, the unordered arrival times should conform to the uniform distribution (Hall, 1991) where the empirical probability distribution should approximately be a straight line (see figure 3). Using Kolmogorov-Smirnov (K-S) statistic (see equation below), the hypothesis could not be rejected at the 1% level of significance, hence concluding that the process is stationary, and the unordered arrival times conform to the uniform distribution.

$$D = \max\left|\frac{A(t)}{A(T)} - t/T\right|$$

IV. Quick Test

For an exponential distribution, a quick test is to check for the value of the coefficient of variation (ratio of Standard deviation to mean) which should be equal to one (1). For the data set, the mean of the inter-arrival time was calculated to be equal to 1.212 minutes while the standard deviation was calculated to be 1.196 minutes, resulting in a coefficient of variation of 0.986 or nearly one. Hence there is a good reason to believe that the inter-arrival time is exponential.

V. Parameter estimation

"The Poisson process is not a single model but actually a family of models, defined by different values of the parameter λ. The model is complete when the value of the arrival rate λ is estimated"

(Hall, 1991). The method of moments (MOM) is used to estimate the value of λ, where the average arrival rate $\lambda = \frac{A(T)}{T}$ was determined to equal 0.855 per minute (calculated as the numbers of received orders "154" over the duration of the data set "180 minutes") or 51.33 per hour, where A (T) represents the total number of data points that occurred within the study period.

Using the equation listed below,

$$\lambda \pm Z_{\alpha/2} \frac{\lambda}{\sqrt{n}} \Rightarrow$$

And a confidence level of 95%, the Confidence interval was determined to be

$$0.855 \pm 1.96 \frac{0.855}{\sqrt{154}} \Rightarrow 0.72 \le \frac{1}{\lambda} \le 0.99$$

2.2 Data Analysis

Analysis of the data set shows that for that day a total of fifteen (15) drivers were scheduled with an average each was dispatched with 1.92 orders (where each order may consist of several pizza pies), traveled 10.59 miles, with a mean service time of 32.40 minutes per dispatch (driving time per delivery).

When drivers are dispatched with more than one order, the actual delivery time of each of the orders is not known, however the mean can be approximated by dividing the average round trip time per dispatch (32.40 minutes) over the number of order per dispatch (1.92 orders/dispatch) yielding a mean service time per order of 16.87 minutes or an hourly service rate of $1/\mu$ of 3.56. Statistics concerning system primitives are tabulated in table 1

Table 1. System Statistics

Order Average Time (minutes)	Mean Time (min)
To reach the delivery Q	9.16
Wait in the delivery queue	10.58
Order age upon Pickup	20.14
Driving Time	16.87
Average Time for order delivery	37.01

Having determined the statistics of the system, the goal now is to eliminate or minimize the waiting time in the delivery queue from the 10.58 minutes the system is currently experiencing.

2.3 Modeling

As a result of the goodness of fit testing, the following assumptions can be made:

- The queuing system is said to operate in steady state
- The probability distribution for the queue length is time independent
- Arrival process is stationary (the arrival process has to be stationary for the system to enter steady state)
- Arrival process is a Poisson process
- Service time is exponential
- The service process is time independent
- Infinite queue capacity
- Queue discipline of FCFS (first-come, first-serve)

The M/M/m/∞ was used to model the queue, with Poisson arrivals, an independent exponential service times, with "m" servers working in parallel and operating at the same service rate μ, where only one customer can be handled at one time. Assuming also that the orders' arrival rate is not influenced by the state of the system, hence the performance measures of the queue would then be determined using R.W. Hall's derivations (Hall, 1991) where:

I. The expected number of orders in the queue is given by

$$\mathbf{L_q} = \frac{\rho^{m+1}/m}{m!(1-\frac{\rho}{m})^2} P_0$$

Where P_0; the probability that all the servers will be idle; is given by

$$P_0 = \frac{1}{\sum_{n=0}^{m-1} \rho^n / n! + \frac{\rho^m}{m!} * \frac{1}{1-\rho/m}} \quad (\rho/m < 1)$$

where the absolute system utilization rate $\rho(t) = \frac{\lambda(t)}{\mu(t)}$ is defined as the order arrival rate over the order service rate

II. The expected number of orders in the system is given by $\mathbf{L_s = L_q + \rho}$

III. The expected waiting time in the queue is $\mathbf{W_q = L_q / \lambda}$

IV. The expected waiting time in the system is $\mathbf{W_s = L_s / \lambda}$

Applying the queuing models to the currently existing system, it was found that fifteen (15) drivers operating at a utilization rate of about 93% would be the required minimum for the system to operate, and the average time spent by an order in the system was found to be about thirty six (36) minutes. In other words, given the current operating conditions of the system, it takes on average about 36 minutes for the customer to receive his order (Figure 8, row 1).

2.4 Sensitivity Analysis

Sensitivity analysis was performed to determine the effect of the number of drivers on the measures of performance (Table 2 and Figures 9-12).

Table 2. Model Statistics

m	Rate	Ls	Lq	Ws (minutes)	Tq	Ws (minutes)	Wq (minutes)
15	96.74%	39.98	25.47	46.44	9.16	55.60	10.58
16	90.69%	20.51	6.00	23.83	9.16	32.99	6.97
17	85.36%	17.02	2.51	19.77	9.16	28.93	2.92
18	80.61%	15.73	1.21	18.27	9.16	27.43	1.42
19	76.38%	15.14	0.63	17.58	9.16	26.74	0.73
20	72.55%	14.84	0.33	17.06	9.16	26.22	0.00

The system shows great improvements to the measures of performance when nineteen drivers (19) instead of the fifteen (15) currently scheduled. For example, the average waiting time of an order in the queue (W_q) drops by almost 95% (from 10.58 minutes to about 0.73 minutes, while the utilization rate of the servers would decrease from 93% to about 76%. No meaningful improvements in the performance levels were achieved when the more than 19 drivers were scheduled (see Figures 7-9).

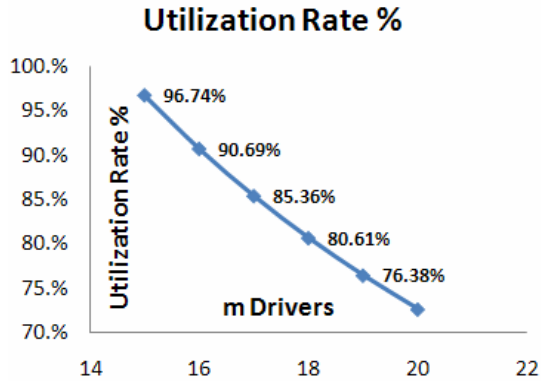


Figure 7. # of Drivers vs. utilization rate

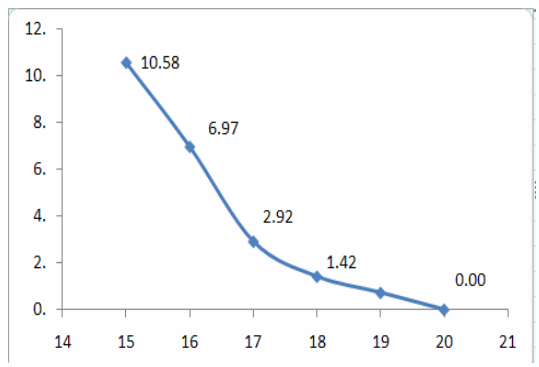


Figure 8. Order's waiting time in the queue (minutes) vs. # of Drivers

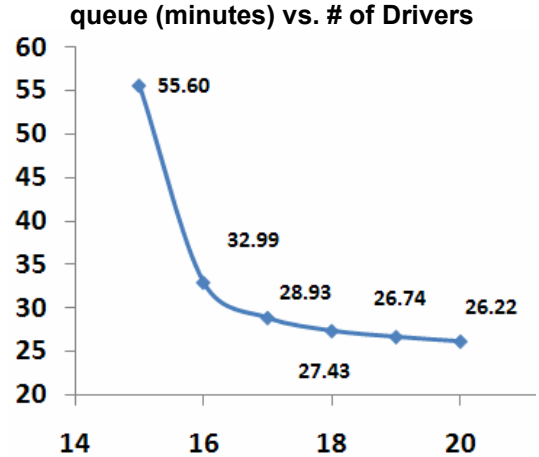


Figure 9. Order's Duration in the System (minutes) vs. # of Drivers

3. Call Loss Rate Analysis

As is the case for most businesses, calls might sometimes be either placed on hold hence forming a queue, or might sometimes get a busy signal. To determine the proportion of calls lost, Erlang's loss formula (Brockmeyer et al. 1948) with an M/G/m/0 queuing system with "**m**" servers and a buffer size of zero (customer will get a busy signal when all lines are busy), with general service time and Poisson arrivals can be used. Erlang's loss formula is given by

$$\text{P(lost calls)} = P_m = \frac{r^m / m!}{\sum_{i=0}^{m} r^i / i!}$$

where "**r**" is defined as the intensity rate representing the ratio of the arrival rate of customer orders to their service rate.

Using the selected data for the case in question, it was determined that customer calls were arriving at an average rate "**λ**" of 53.67 calls per hour, and were being served at a rate "**μ**" of 61.43 per hour, hence an intensity rate "**r**" of 0.874.

Applying Erlang's loss formula for different scenarios of available phone lines, it was determined that five (5) phone lines would be sufficient to help achieve the objective of a call loss rate of less than 1% (see Figure 10).

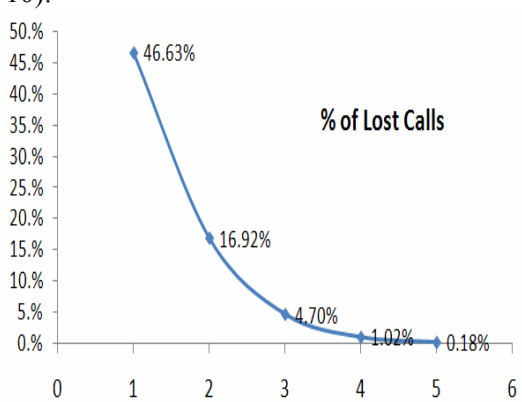


Figure 10. Call loss rate vs. # of existing phone lines

4. Recommendations

Because no meaningful improvements in the performance levels were achieved when more than 19 drivers were scheduled, it is recommended that scheduling a total of nineteen (19) drivers would be sufficient to help the management achieve the objective of minimizing the mean order delivery time to about twenty five (25) minutes.

By minimizing the customer waiting time by almost 10 minutes, it is believed that the increase in customer satisfaction would surely offset the increased labor cost of about $120 (4 additional drivers at $10/hr for 3 hours) and would help management achieve its objective. In addition, the five (5) phone lines currently being used by the company should be sufficient to achieve the management objective of a call loss rate of one (1) percent.

References

[1] E. Anderson, and M. Sullivan, "The Antecedents and Consequences of Customer Satisfaction for Firms", Marketing Science, Vol. 12, No. 2 (1993), pp. 125-143.

[2] W. Bearden, and E. Teel, "Selected Determinants of Consumer Satisfaction and Complaint Reports", JMR, Journal of Marketing Research, Vol. 20, No. 1 (1983), pp. 21-28.

[3] R. Bolton, "A Dynamic Model of the Duration of the Customer's Relationship with a Continuous Service Provider: The Role of Satisfaction", Marketing Science, Vol.17, No.1 (1998), pp. 45-60.

[4] R. Bolton, and J. Drew, "A Multistage Model of Customers' Assessments of Service Quality and Value", Journal of Consumer Research, Vol. 17, No. 4 (1991), pp. 375-384.

[5] W. Boulding, K. Ajay, R. Staelin, and V. Zeithaml, "A Dynamic Process Model of

Service Quality: From Expectations to Behavioral Intentions", JMR, Journal of Marketing Research, Vol. 30, No. 1 (1993), pp. 7-27.

[6] E. Brockmeyer, Transactions of the Danish Academy of Technical Sciences, No.2 (1948).

[7] R. Buzzel, and T. Bradley, Principles: Linking Strategy to Performance. Free Press, 1987.

[8] C. Fornell, and Wernerfelt B., "Defensive Marketing Strategy by Customer Complaint Management: A Theoretical Analysis." JMR, Journal of Marketing Research Vol. 24, No. 4 (1987), pp. 337-346

[9] R. Hall, Queuing Methods for Services and Manufacturing. Englewood Cliffs, NJ: Prentice-Hall, 1991.

[10] V. Mittal, and W. Kamakura, "Satisfaction, Repurchase Intent, and Repurchase Behavior: Investigating the Moderating Effect of Customer Characteristics." JMR, Journal of Marketing Research, Vol. 38, No. 1 (2001), pp. 131-142.

[11] V. Mittal, W. Ross, and P. Baldasare, "The Asymmetric Impact of Negative and Positive Attribute-Level Performance on overall Satisfaction and Repurchase Intentions." Journal of Marketing Vol. 62, No. 1 (1998), pp. 33-47.

[11] R. Oliver, "A Cognitive Model of the Antecedents and Consequences of Satisfaction Decisions." JMR, Journal of Marketing Research Vol. 17, No. 4 (1980), pp. 401-417

[12] Parasuraman, A. "A Conceptual Model of Service Quality and its Implications for Future Research." Journal of Marketing Vol. 49, No. 4 (1985), pp. 41-50.

The Journal of Management and Engineering Integration
Vol. 3, No. 1, Summer, 2010

Life Cycle Cost Estimator as a Decision Tool for Bridge Deck Selections

Sidharta Sahirman
Jenderal Soedirman University, Indonesia
ssahirma@mix.wvu.edu

Robert C. Creese and Hota V.S. GangaRao
West Virginia University
Robert.Creese@mail.wvu.edu, Hota.Gangarao@mail.wvu.edu

Abstract

Comparison of infrastructural projects that have high investment costs and a long life expectancy requires a life cycle costing (LCC) methodology, because there are costs beyond the initial construction costs that should be considered. LCC is the comprehensive way to compare long term projects, which lead the United States Federal Highway Administration (FHWA), encourages states to use LCC when determining which roadway projects to fund.

Regardless of the importance of this analysis, there has been little research conducted for estimating life cycle cost of Fiber Reinforced Polymer (FRP) bridge decks, the new material for future bridges. The LCC estimator – Excel®-based software is developed to help decision makers in analyzing this alternative as opposed to conventional bridge decks. The main function of the tool is to quantify the costs associated with FRP systems as well as ones for Steel Reinforced Concrete (SRC) systems for their entire life cycles. West Virginia Katy Truss bridge deck project is used to demonstrate the usefulness of the software in the decision making.

1. Introduction

West Virginia District of Transportation (WVDOT) under cooperation with West Virginia University (WVU) were considering using new bridge deck material for both new bridge projects and bridge deck replacement projects. Two alternatives being considered were concrete deck and FRP deck.

FRP deck has a lot of advantages to offer, including increased structure service life, reduced maintenance time and costs, reduced field installation cost and time, and greater vehicle loads on repaired structured due to its light weight. It is true that more than a decade ago it had economic and technical barrier as suggested by Ehlen and Marshall in 1996. However, the technical barrier has been eliminated in the last few years, so the only criterion for bridge deck selection is cost. Hence the question becomes: Are FRP bridge decks competitive on a cost basis with SRC decks?

2. Fiber Reinforced Polymer for Construction

An FRP composite is defined as a combination of a polymer matrix (either a thermoplastic or thermoset resin, such as polyester, isopolyester, vinyl ester, epoxy, or phenolic) and a reinforcing agent, such as glass, carbon, aramid, or other reinforcing material so

that there is a sufficient aspect ratio (length to thickness) to provide a discernable reinforcing function in one or more directions. The fiber is the critical constituent in composites and occupies 30-70% of the composite matrix volume.

FRP deck has been used as bridge material for more than two decades. In 1986, the world's first highway bridge using composite reinforcing tendons was built in Germany. The first all composites pedestrian bridge was installed in 1992 in Aberfeldy, Scotland. In the US, the first FRP-reinforced concrete bridge deck was built in 1996 at McKinleyville, WV followed by the first all-composite vehicular bridge deck in Russell, KS in 1996. Up to date, there are more than 80 FRP bridge decks in US; 13 of them are in West Virginia.

3. Bridge Deck Life Cycle Cost

There are two main approaches one could use for cost analysis: Initial cost analysis and life cycle cost analysis. Initial cost is basically a subset of life cycle cost. Life cycle cost is sum of initial and all recurring costs over the full life span or a specified period of a system. It includes purchase price, installation cost, operating costs, maintenance and repair costs, and salvage value at the end of its useful life. This technique can be used for many reasons, but its primary uses are for comparison of competing alternatives, long range planning and budgeting, selection among competing contractors, decisions about replacement of aging equipments.

Comparison of infrastructure projects that have high investment cost and a long life expectancy require life cycle cost (LCC), because there are significant costs beyond the initial costs. That's why LCC is recommended by FHWA for determining which roadway projects to fund.

Five important studies performed on LCC of FRP bridge decks are those by Lopez-Anido [3], Ehlen [2], Nathan and Onyekmeluwe [4], Roychoudury and Creese [6], and Nishizaki et al. [5]. In the above models, the service life of an FRP bridge deck is either fixed or has to be estimated by the user. It is understandable that one makes assumptions about values according to what one believes, because currently the FRP bridge deck service life is unknown. However, since service life plays an important role in LCC estimation, it is important to get the best possible estimate for FRP bridge deck service life. Furthermore, those models do not take into account the substructure cost reductions from using FRP as bridge deck material. Inclusion of this cost will provides us a better comparison between the two bridge deck systems. Additionally, Ehlen [2] and Lopez-Anido [3] used a preset maintenance/repair schedule based on concrete deck maintenance/repair schedule to obtain maintenance/repair costs. FRP bridge deck cost estimation model is not available in all of the above mentioned research. All of those concerns are addressed in the proposed model.

LCC of bridge deck consists of Initial Costs, Maintenance/Anticipated Repair Costs, and Disposal Costs. Each includes two major components: Agency Costs and User Costs. Agency Costs include all direct costs, i.e. material costs, labor costs, transportation costs. User Costs are costs associated with lost time for the drivers of the vehicles, higher vehicle operation costs and increased accident rates. These costs can be sizable, depending on the total installation time as well as expected delay time. The expected delay time is a function of average daily traffic and length of the affected road work. The expected delay time multiplied by the value of user's time reflects the cost for driver as result of losing his production time; its' multiplication with vehicle operation cost per unit time reflects increased vehicle operation cost [1]. Bridge deck life cycle cost components are depicted in Figure 1.

4. Life Cycle Cost Estimator

To ease the way one obtain life cycle cost of FRP bridge deck, LCC cost estimator, Excel® based program was developed. The unique modules included in the program are service life prediction which based on factor method and FRP deck cost estimates which based on learning curve application. Additionally, the advantage of FRP to enable higher load rate for existing bridges or to require less expensive substructure for new bridges is also considered in the model, this special feature was named 'cost saving from reduced substructure' cost.

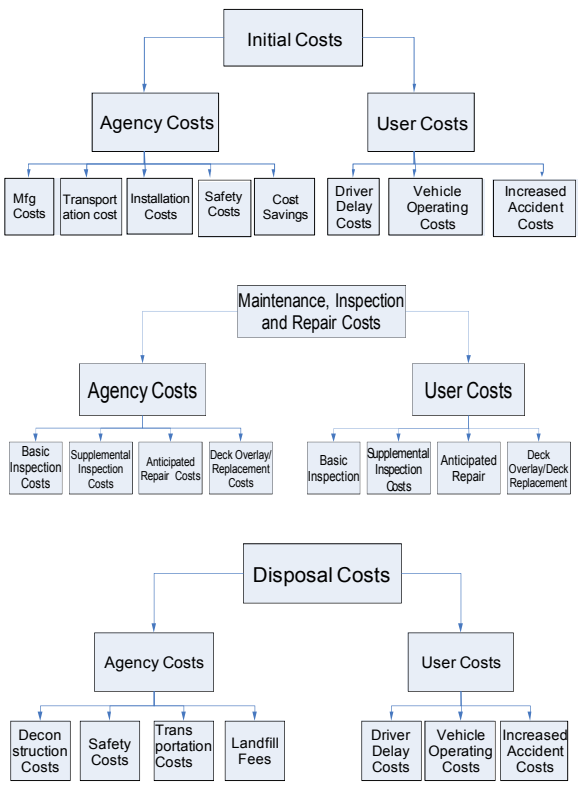


Figure 1. Life cycle cost components

The model requires 25 inputs as given in Table 1. All of input costs are converted to the installation year using Consumer Price Index. Based on those inputs, the LCC Estimator determines total life cycle costs as well as the annual average costs of the bridge deck alternatives during the service life.

Table 1. Required inputs

Item	Description	Symbol	Units
1	Length of bridge deck	DL	ft
2	Width of bridge deck	DW	ft
3	Project Site		
4	FRP Manufacturer		
5	Thickness of FRP bridge deck	DT	inch
6	Self weight of FRP bridge deck	DS	lb/ft ^{2}
7	Thickness of SRC bridge deck	DTC	inch
8	Self weight of SRC bridge deck	DSC	lb/ft ^{2}
9	SRC bridge deck cost	CC	$/ft ^{2}
10	Year of installation	YR	years
11	Normal traffic speed	NS	miles/hour
12	Traffic speed during bridge deck construction	CS	miles/hour
13	Hourly time of driver	HC	$/hr
14	Hourly vehicle operating cost	VC	$/hr
15	Normal accident rate	NA	mile
16	Accident rate during bridge work	CA	mile
17	Average cost per accident	AC	$/accident
18	Self weight of wearing surface	WS	lb/ft ^{2}
19	Type of wearing surface	WST	
20	Labor cost	LC	$ /hr
21	Average daily traffic	ADT	vehicles/day
22	Freeze Thaw Cycles	FTC	
23	Discount rate	DR	%
24	Study Period	SP	years
25	Reference Service Life	RSL	years

Since FRP bridge decks and SRC bridge decks have different life spans, the equivalent annual cost or annuity method was chosen as it can be applied for any combination of two different service lives. This approach was performed by determining a fixed study period based on the life of the bridge superstructure. For medium bridges, the study period can range from 50 -70 years while for large bridges the study period can be 100 years. Based on the given study period, the life cycle cost per square foot of the two alternate bridge decks are calculated and compared. The life cycle cost model is shown in Figure 2.

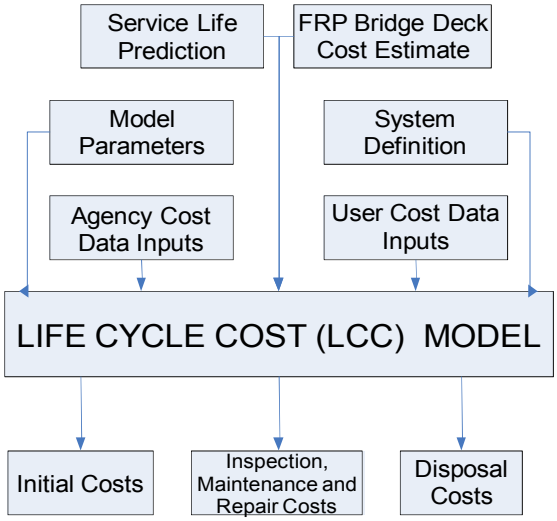


Figure 2. Life cycle cost model

5. Bridge Deck Selections using Life Cycle Cost Estimator

Katy Truss bridge deck project is chosen to illustrate the helpfulness of LCC estimator in the decision making. The bridge which located in Marion County, WV is a single span bridge with steel substructure. It is located in Marion County, WV with 91.3 feet in length and 14.3 feet in width. Two choices being considered are FRP deck and SRC deck.

The FRP deck manufacturing cost for this bridge deck based on the learning curve module is $67.5/sq ft in year 2000 and $57.7 in year 2006. The service life of FRP deck is estimated to be 60 years, while SRC deck is 30 years. Life cycle cost components under default scenario (deck installed in year 2000 with Average Daily Traffic (ADT) of 700 vehicles/day) is given in Table 2. Overall, the LCC of FRP and SRC under default scenario are $67.6/sq ft and $62.1/sq ft respectively. Inspection/repair costs accounted for 13 percent of FRP life-cycle cost and 35 percent of SRC cost, while disposal costs accounted for one percent of FRP life cycle cost and nine percent of SRC cost as illustrated in Figure 3. Hence, FRP is not financially viable under the given inputs. This phenomenon emphasizes the important role of the FRP bridge deck manufacturing cost in determining the economic viability of a FRP bridge deck. Cost savings from superstructure when FRP chosen, reduced maintenance frequency as well as FRP's longer service life are the three important inputs that might balance this out to make FRP competitive to SRC in most cases, but not for this case. This shows that FRP decks can't compete with SRC deck when the ratio between FRP deck and SRC deck is too high. The outputs of LCC estimator for various FRP deck manufacturing costs are shown in Figure 4. For Katy Truss Bridge, FRP is more economical than SRC if FRP deck costs less than $61.8/sq ft.

In addition, the model showed that there is a significant relationship between ADT and LCC cost [3]. The higher the ADT he more favorable the FRP deck becomes. The main reason for the phenomenon is the higher the ADT, the larger the difference between an FRP deck relative to an SRC deck. Based on the life cycle cost distribution, the main benefit of using FRP deck was in the user cost during construction and replacement. The higher the ADT, the higher the user cost is. Since SRC deck initial and disposal activities consume more time, it results in higher sensitivity to ADT as shown by a steeper slope. Therefore, higher ADT is more beneficial for FRP deck. Figure 4 shows total life cycle costs for the Katy Truss bridge deck under different ADT. For this case study bridge, increasing ADT to 3,000 enables the FRP deck to be competitive to SRC deck.

It can be concluded that FRP deck viability is a function of its service life, maintenance/anticipated repair schedule, ADT, and initial price ratio (substructure saving included). Overall, the results suggest that a $61/sq ft FRP deck is a viable alternative to $30/sq ft SRC deck.

Table 2. Life cycle cost (default values)

Bridge Deck Project	Katy Truss	Katy Truss
Yr	2000	2000
Bridge Deck Alternative	FRP	SRC
ADT (vehicles/day)	700	700
Study Period (yr)	60	60
Service Life (yr)	60	30
Deck Cost ($/sq ft)	$67.5	$26.5
Initial Costs		
Agency Costs		
Total Agency Costs	$69.3	$33.6
User Costs		
Total User Costs	$0.2	$1.3
Structural Savings		
Total Structural Savings	$11.4	$0.0
Total Initial Costs	$58.1	$34.9
Maintenance/Repair Costs		
Agency Costs		
Inspection/Repair	$6.9	$6.6
Deck Overlay or Replacement	$1.3	$13.9
Total Agency Costs	$8.2	$20.5
User Costs		
Total User Costs	$0.9	$1.5
Total Maintenance/Repair Costs	$9.1	$21.9
Disposal Costs		
Agency Costs		
Total Agency Costs	$0.3	$5.1
User Costs		
Total User Costs	$0.0	$0.1
Total Disposal Costs	$0.3	$5.2
LIFE CYCLE COST		
Total Ageny Costs	$66.4	$59.1
Total User Costs	$1.2	$2.9
Total Life Cycle Costs	**$67.6**	**$62.0**

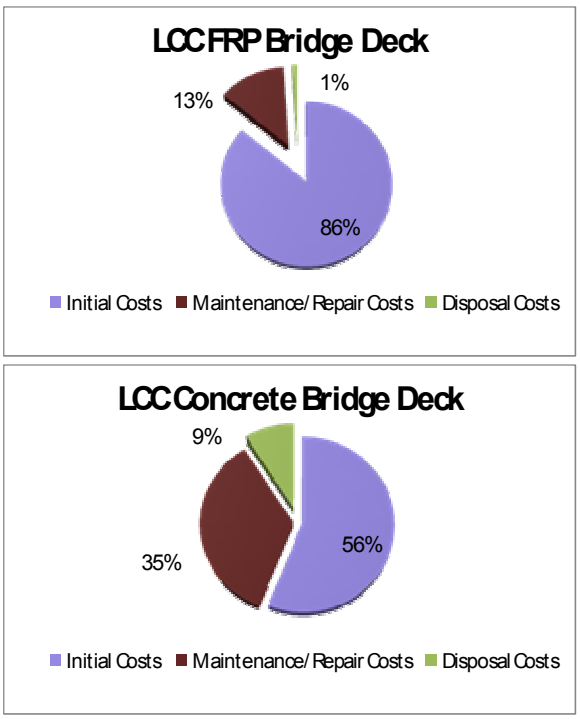


Figure 3. Life cycle components for two alternatives

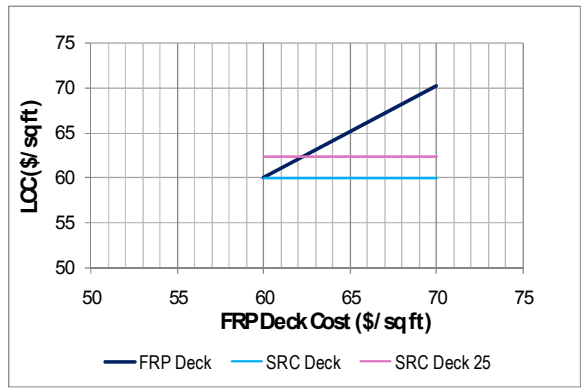


Figure 4. Manufacturing cost and LCC

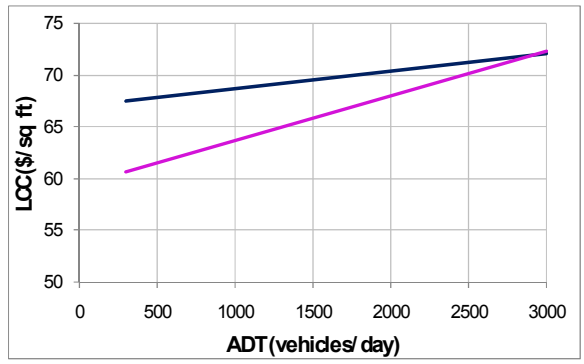


Figure 5. ADT and LCC

6. Conclusions

Life cycle cost estimator – the Excel® based tool has features that are not covered in existing models: (1) Estimating FRP deck cost using Learning curve method, (2) Taking into account cost savings from reduced substructure costs, and (3) Estimating FRP bridge deck service life using factor method. The software provides critical information in the decision making of bridge deck selections.

The case study utilizing life cycle estimator shows that FRP deck viability if a function of its service life, manufacturing cost and ADT. The higher the ADT, the more competitive the FRP deck becomes.

7. References

[1] Ehlen, M.A. and H.E. Marshall, "The Economics of New-Technology Materials: A case study of FRP Bridge Decking", NISTIR 5864, US Department of Commerce, 1996, pp. 1-10.

[2] Ehlen, M.A., "Life-Cycle Costs of Fiber-Reinforced-Polymer Bridge Decks". Journal of Materials in Civil Engineering August 1999: 224-230.

[3] Lopez-Anido, R, "Life Cycle Cost Evaluation of FRP Composite Bridge Deck", Technical Report. Morgantown: WV University Research Corporation, 1998.

[4] Nathan, T.R. and O.U. Onyemelukwe, "Comparison of Bridge Deck Alternatives Using Life-Cycle Costs", Proceedings of the 8th ASCE Specialty Conference on Probabilistic Mechanics and Structural Reliability, July 24-26, 2000, Notre Dame, Indiana.

[5] Nishizaki, I., N.Takeda, Y.Ishizuka and T.Shimomura, "A Case Study of Life Cycle Cost Based on a Real FRP Bridge". Proceedings of the Third International Conference on FRP Composites in Civil Engineering, December 13-15, 2006, Miami, Florida, 99-102.

[6] Roychoudhury, P. and R.C. Creese, "Cost Analysis of FRP Bridge Decks", AACE International Transactions 2001, EST 02:1-8.

[7] Sahirman, S., R.C. Creese, and H. GangaRao, "Preliminary Study of Life-cycle Cost Model for Fiber Reinforced Polymer Bridge Deck", Advances in Life-Cycle Cost Analysis and Design of Civil Infrastructure Systems, Proceedings of the 4th International Workshop on Life-Cycle Cost Analysis and Design of Civil Infrastructures Systems, Cocoa Beach, Florida, May 8-11, 2005.

8. Acknowledgements/Disclaimer

The U.S. Department of Transportation, Federal Highway Administration sponsored this work under the Center of Excellence Project (Contract No. DTFH61-04-R-00032). The contents of this report reflect the views of the authors who are responsible for the facts and accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of either the State or the Federal Highway Administration. This report does not constitute a standard, specification or regulation. Trade or manufacturers names that may appear herein are cited only because they are considered essential to the objectives of this report. The United States Government does not endorse products or manufacturers.

The Journal of Management and Engineering Integration
Vol. 3, No. 1, Summer, 2010

Business Decision to Adopt a New Technology: SMEs Struggle to Increase Performance

Michael Stoica
Washburn University

David Price
Griffith University

Abstract

Adoption of new technology is critical for small and medium-sized businesses (SMEs). The adoption process is dependent upon organizational culture of the business, its strategy and maturity, as well its size and the surrounding environment and the industry in which the business is operating. A model has been developed that incorporates all these variables. The model is tested using hierarchical regression. The methodology, results, and managerial implications are discussed.

Introduction

Adoption of new technology by businesses is a dynamic process, not well understood, with complexities that draw from the business environment, the organizational culture of the firm, its business strategy, and many other dimensions [1]. The adoption process is fast and as new (information) technology becomes cheaper and more accessible, the competitive advantage early adopters have at the beginning, dissipates quickly [2]. Information and communication technology, such as e-commerce and m-commerce, has been widely embraced by businesses both small and large. Across a wide range of industries firms have adopted wireless business initiatives to better manage their internal processes as well as their interfaces with the environment. By understanding the determinants of business performance firms can better formulate their business models and comprehend how a technology, such as the Internet, impacts the firm as well as how companies can exploit new technology [3].

Businesses adopt innovations to gain competitive advantages or capabilities. The literature in this area has defined specific characteristics and conditions that facilitate the adoption process. These include functional differentiation [3], administrative intensity [4], external and internal communication, and vertical integration [5]. Environmental factors that have been identified as antecedents of innovation adoption are: (a) competition, the fear of being left behind, the perceived need to keep-up [6]; (b) consumers and suppliers, that demand the adoption of a new technology that they perceive will reduce their costs, or increase their benefits.

Intensity of New Technology Adoption

Mobile Commerce applications are evolving with a wide variety of procedures in use. Applications have been developed for existing mobile platforms including laptop PCs, PDA's, probes (wireless sensors), telephone handsets, telephone headsets, mobile multimedia, wireless routers, GPS, and specialty pagers. Mobile applications have focused on: (a) delivering existing Internet services to the mobile customer; (b) using location sensing to deliver location based information; (c) using location sensing for tracking (fleet services, automobiles, pets, etc.); (d) using broadband to deliver mobile entertainment content (music, games, etc.).

Application coverage is not universal and is constrained by the size, location and technology used by the mobile infrastructure provider. Low-End Applications (LEA) include email, web browsing and information services. High-End Applications (HEA) include transactions, inventory management, supplier-buyer relationship, interactivity, etc. SMEs can choose, and several have already chosen, between adopting LEA or HEA. In time those who adopted LEA will move to the more sophisticated and more profitable HEA. High intensity adoption is determined by HEA. Low intensity adoption will characterize firms that adopt only LEA.

Information search frequency is represented by the number of times managers receive data about the environment [7]. Fahey, King and Narayanan (1981) [8] observed that managers could obtain information along a continuum ranging from irregular to continuous gathering. The irregular approach is a reactive, spot behavior that involves external cues that force management into action. By contrast, continuous scanning is a perpetual, systematic and proactive approach to search relevant environmental information. Some researchers suggested that the level of environmental uncertainty and the availability of resources within an organization affect the extent of scanning activities undertaken by managers [9, 10]. Product innovation and technological change are related to information search [11]. Mole, Ghobadian, O'Reagan and Liu (2004) [12] analyzed the intelligence gathering in small and mid-sized businesses and the impact on performance. Innovative businesses are less profitable since they spend time and energy for technology improvement. Egan, Clancy and O'Toole (2003) [1] arrived at the same conclusion for the case of e-commerce. Hence it is hypothesized that:

H1-1. Rate of product innovation is negatively associated with performance. The relationship is weaker for HEA.

H1-2. Rate of technological change is negatively associated with performance. The relationship is weaker for HEA.

H1-3. Stability of buyer-supplier relationships is positively associated with performance. The relationship is stronger for LEA.

Managers routinely use information search to reduce uncertainty. When the environment has remained relatively stable or changed in a slow cycle, managers at SMEs will commit fewer resources to information search, which will lead to a longer cycle time. Research suggests that higher environmental turbulence generally elicits more frequent scanning efforts [13]. Under such conditions, SME decision makers need more information to define problems, while generating and evaluating alternative solutions [14].

It is hypothesized that for all SMEs (both LEA adopters and HEA adopters) a firm's strategy affects its performance [15, 16]. Strategies are more desirable if they create or exploit asymmetries among firms and contexts (isolation from rivals, low cost provider). As mentioned by Barrett and Weinstein (1998) [17] and Landstrom (1999) [18], substantial literature on new venture performance does exist although it has more in common with psychology and economics than with strategic management research. Schumpeter (1934) [19] and later Birch (1987) [20] and Reagan (1985) [21] analyzed the determinants of new venture performance. Strategy configuration (such as isolation from rivals, low pricing strategy, brand loyalty) and links between strategy, structure, and performance have been studied by Ward, Bickford and Leong (1996) [22], Shane (1996) [23] and Castrogiovanni (1996) [24]. Strategy and organizational adaptation and their influence on performance were the topic of several other studies among such as Miller, Lant, Milliken and Korn (1996) [25], Jennings and Seaman (1994) [26]. Findings show that businesses with particular levels of adaptation (speed of response, brand loyalty) tend to have specific strategy-structure arrangements, which influence performance. Firms with an optimum strategy-structure match tend to have higher performance. Size is thought to have an effect and thus, the case of SMEs is of particular importance. Therefore it is hypothesized that:

H2-1. The more the firm is isolated from its rivals the higher will be the small business performance. The relationship will be stronger for LEA businesses than for HEA businesses.

H2-2. The extraction of temporary profits will negatively affect business performance. The relationship will be stronger for HEA businesses.

H2-3. Low pricing strategy is negatively related to business performance. The relationship will be weaker for the LEA businesses than for the HEA businesses.

H2-4. The higher the brand loyalty the higher the performance (stronger for HEA).

H2-5. Speed of response and performance are directly related (stronger for HEA).

Methodology

A survey was administered to over 900 businesses in part of the rural Midwest to better understand the struggle local companies have with new technology adoption, especially mobile commerce. The questions asked relate to wireless technology devices used, applications implemented, or planned for the next two years, motivation to use wireless technology, and changes needed to expand their usage of wireless business. Other questions refer to the market and business environment (including government regulations and components of economic readiness as defined previously), its degree of turbulence, their business strategy, and their business performance.

The sampling frame was given by a list of businesses provided by the state Chamber of Commerce. The survey was sent to 987 small and medium-sized firms and received a 21.3 percent response rate. Most firms polled use cellular phones (89.7 percent). There are many methods by which the performance of a venture might be measured [27], for this research an objective measure utilizing return on sales was utilized.

Hierarchical regression was the method used to test the two sets of hypotheses. In addition a mediation testing procedure (see Figure 1) was conducted for the first set of hypotheses. The analysis was conducted in accordance with Baron and Kenny (1986) [28].

The basic objective of statistical mediation is to show that a portion of the independent variable's effect on the dependent variable is 'captured' or 'carried' through the mediator variable [29]. If the entire effect is captured, the independent variable is said to be fully mediated. Fully mediated effects are rare and a more common result is partial mediation, whereby the mediator accounts for some but not all, of the independent variable's impact on the dependent variable.

Mediation is implied when the independent variable's estimated coefficient in the first-step equation is significantly different from its estimated coefficient in the third-step equation. That is, by introduction of the mediating variable into the explanatory model accounts for some or all of the independent variable's effect.

Therefore, this research tests if adoption intensity is a mediator of the relationship between product innovation, technological change and the stability of the Buyer-Seller relationship.

Results

Wireless networks are used only by early adopters. This research indicated that over 20 percent of all businesses have no application implemented. However, 8.5 percent will not implement any type of application for the next two years. While emailing/calendaring and web access are widely used (over one third of businesses), information services are used only by early adopters. Wireless operations management and interactive transactions are just emerging as options for business use.

Table 1 reports the means of the variables and their pairwise correlations. Tables 2 and 3 provide the results from hierarchical regression analysis with return on sales and growth rate as the dependent variable. Overall in both cases, incremental R-squares were all statistically significant, suggesting that after controlling for the effect of years in business and size, industrial business strategy variables contribute significant explanations of performance variance.

The hypotheses are partially confirmed. H1 stated that change and innovation are negatively associated with performance. Indeed product innovation is negatively associated with performance for HEA firms (see Table 2 and 3, industry structure variables). There is no significant association between innovation and performance for LEA firms. However, there is a significant positive relationship between technological change and performance for HEA firms. The stability of buyer-supplier relationships has, as expected, a positive influence over performance for both HEA and LEA firms.

H2 stated a relationship between the firm's strategy and its performance. The results partially support the hypothesis. It seems that only in the LEA firm case, successfully managing to isolate firm from rivals (H2-1) will

increase its return on sales (Table 2). However, extracting temporary profits (H2-2) seems to be a non-lucrative strategy for both HEA and LEA businesses. Achieving a low cost position (H2-3) shows no impact on performance, but building brand loyalty (H2-4) will significantly increase the firm's return on sales for all firms.

Strategic approaches could be clustered into two categories: Defending strategies and Proactive strategies. Isolating a firm from rivals, extraction of temporary profits and low cost achievements belong to the first category. Speed of response and brand loyalty belong to the second. High end application adopters (HEA) use a proactive strategic approach and have a long term orientation. LEA adopters behave differently; they embrace a defending strategy and work on a short term perspective. This offers a partial explanation on the differences between HEA and LEA firms.

The intensity of the adoption process was considered the mediator for the above relationships. The analysis was conducted in accordance with Baron and Kenny (1986) [28] and the results are shown in Table 4. Results indicate the intensity of adoption is a mediator for all three variables.

The hypothesis testing results presented in Table 4 indicate that the addition of product innovation, technological change and the stability of buyer-seller relationship as mediator variables do enhance the relationship between adoption intensity and firm performance. In other words, a mediator is the mechanism through which a predictor variable influences an outcome variable [28, 29, 30].

Conclusions

The findings of the testing performed above confirmed Chrisman's et. al (1998) theory that any model of SME performance should recognize the critical nature of strategy. Performance of businesses will largely depend on the competitive strategies within the chosen industry. Partial confirmation was also obtained on the innovation-performance relationship. As mentioned previously, HEA firms are considered to pursue a proactive strategy, while LEA firms behave more as defenders.

It is evident that high-end adopters are not defenders. The proactive strategy is more market oriented, relying on brand loyalty and speed of response. These are two important components of marketing orientation as defined in the MARKOR model. The results show an implicit relationship between marketing orientation and high-end adoption of wireless applications. Anecdotal evidence suggests that corporate and government subsidies can have a substantial impact on the adoption process. Future research in this area will attempt to quantify this effect. The intensity of adoption appears to have a positive effect on firm performance, suggesting that firms should adopt these industry structure variables at a higher rate.

References

[1] Egan, T., Clancy, S., and O'Toole, T. 2003. The integration of e-commerce tools into the business processes of SMEs, Irish Journal of Management, 12 (1), pp. 139-153.

[2] Johnston, H.R., and Vitale, M.R., 1988. Creating competitive advantage with interorganizational information systems, MIS Quarterly, 12 (2), pp. 153-165.

[3] Afuah, A., and C. Tucci, 2001. Internet Business Models and Strategies, McGraw-Hill, Boston, MA.

[4] Wu, F., Mahajan, V., Balasubramanian, S., 2003. An analysis of e-business adoption and its impact on performance. Journal of the Academy of Marketing Science, 31 (4), pp. 425-447.

[5] Chrisman, J., Bauerschmidt, A., and Hofer, C., 1998. The determinants of new venture performance. Entrepreneurship: Theory and Practice, 23 (1), pp. 122-36.

[6] Morace, F., and Chrometzka, L., 2001. The new dynamics of the wired era: Net flow and creative pollination. Presentation at the M-Conference: Seizing the Mobile Advantage. Rotterdam, School of Management, Rotterdam, Holland, January 19-20.

[7] Hambrick, D. C., 1981. Environment, strategy, and power within top management teams. Administrative Science Quarterly, 26 (2), pp. 253-275.

[8] Fahey, L., King, W. R., and Narayanan, V. K. 1981. Environmental scanning and forecasting in strategic planning: The state of the art. Long Range Planning, 14 (1), pp. 32-39.

[9] Boyd, B., and Fulk, J. 1996. Executive practices and perceived uncertainty: A multidimensional model. Journal of Management, 22 (1), pp. 1-21.

[10] Milliken, F. J., 1987. Three types of uncertainty about the environment: State, effect, and response uncertainty. Academy of Management Review, 12, pp. 133-143.

[11] Weiber, R., and Kollman, T., 1998. Competitive advantages in virtual markets - perspective of 'information-based' marketing in cyberspace. European Journal of Marketing, 32 (7/8), pp. 603-615.

[12] Mole, K., Ghobadian, A., O'Reagan, N., and Liu, J., 2004. The use and development of soft process technologies within UK manufacturing SMEs: An empirical assessment using logit models. Journal of Small Business Management, 42 (3), pp. 303-324.

[13] Chakravarthy, B., 1997. A new strategy framework for coping with turbulence. Sloan Management Review, 38 (2), pp. 69-83.

[14] Elenkov, D., 1997. Strategic uncertainty and environmental scanning: The case for institutional influences on scanning behavior. Strategic Management Journal, 18 (4), pp. 287–302.

[15] Black, J., and Boal, K., 1994. Strategic resources: Traits, configurations and paths to sustainable competitive advantage. Strategic Management Journal, 15, pp. 131-48.

[16] Sandberg, W.R., 1986. New venture performance: The role of strategy and industry structure. Lexington, MA: DC Health and Company.

[17] Barrett, H., and Weinstein, A., 1998. The effect of market orientation and organizational flexibility on corporate entrepreneurship. Entrepreneurship: Theory & Practice, 23 (1), pp. 57-70.

[18] Landstrom, H., 1999. The roots of entrepreneurial research. New England Journal of Entrepreneurship, 2 (2), pp. 9-21.

[19] Schumpeter, J.A., 1934. The theory of economic development. Cambridge, MA: Harvard University Press.

[20] Birch, D.L., 1987. Job creation in America: How our smallest companies put the most people to work. New York: Free Press.

[21] Reagan, R., 1985. Why this is an entrepreneurial age. Journal of Business Venturing, 1, pp. 1-4.

[22] Ward, P.T., Bickford, D.J., and Leong, G.K., 1996. Configurations of manufacturing strategy, business strategy, environment, and structure. Journal of Management, 22, pp. 597–626.

[23] Shane, S., 1996. Explaining variation in rates of entrepreneur-ship in the United States: 1899-1988. Journal of Management, 22, pp. 747-781.

[24] Castrogiovanni, G., 1996. Pre-startup planning and survival of small businesses: Theoretical linkages. Journal of Management, 22 (6), pp. 801-822.

[25] Miller, D., Lant, T.K., Milliken, F.J., and Korn, H.J., 1996. The evolution of strategic simplicity: Exploring two models of strategic adaptation. Journal of Management, 22, pp. 863-887.

[26] Jennings, D., and Seaman, S.L., 1994. High and low levels of organizational adaptation: An empirical analysis of strategy, structure, and performance. Strategic Management Journal, 15, (6), pp. 459-475.

[27] Dollinger, M. J. 1984. Measuring effectiveness in entrepreneurial organizations. International Small Business Journal, 5(1), pp. 10-20.

[28] Baron, R.M., and Kenny, D.A., 1986. The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. Journal of Personality and Social Psychology, 51, pp. 1173-1182.

[29] Holmbreck, G.N. 1997. Toward terminological, conceptual, and statistical clarity in the study of mediators and moderators: Examples from the child-clinical and pediatric psychology literatures. Journal of Consulting and Clinical Psychology, 65, pp. 599-610.

[30] James, L.R., and Brett, J.M. 1984. Mediators, moderators, and tests for mediation. Journal of Applied Psychology, 69, pp. 307-321.

Table 1. Descriptive Statistics and Pearson Correlation

	Mean	Standard Deviation	Years in business	Size	Product innovation	Technological change	Stability of buyer-supplier relationships	Intensity of rivalry	Isolate from rivals	Extract temporary profits	Achieving a low cost position	Build brand loyalty	Speed of response
Years in business.	10.3	3.83	1.000										
Size	4.92	1.21	.310**	1.000									
Product innovation	2.89	.98	-.017	.208*	1.000								
Technological change	3.34	.87	-.200	.222	.390**	1.000							
Stability of buyer-supplier relationships	3.33	.99	.129	.027	-.210*	-.102	1.000						
Intensity of rivalry	2.43	.66	-.038	.140*	.132	.099	.201	1.000					
Isolate from rivals	4.15	2.59	.004	-.005	.029.	-.068	-.059	.197**	1.000				
Extract temporary profits	2.77	2.38	-.124*	-.079	-.022	.034	.302*	.077	.121	1.000			
Achieving a low cost position	5.34	2.53	.1082	.208**	.093	.142	-.034	.066	-.115	.192***	1.000		
Build brand loyalty	6.19	2.31	.181**	-.032	.107	.078	.206	-.020	.001	-.242***	.009	1.000	
Speed of response	6.75	2.15	.055	-.136*	-.211*	.230*	.045	.115	-.034	-.081	-.078	.024	1.000

*** Correlation is significant at the 0.01 level (2-tailed).
** Correlation is significant at the 0.05 level (2-tailed).
* Correlation is significant at the 0.10 level (2-tailed).

Table 2. Hierarchical Regression Results

HEA Firms. Dependent Variable: Return on Sales

	Return on Sales					
	Model 1		Model 2		Model 3	
Control Variables	***β***	***t***	***β***	***t***	***β***	***t***
Years in business	.111	1.353	.061	.800	.191	-.113
Size	-.019	-.297	.049	.702	.011	.097
Industry Structure Variables						
Product innovation			-.293	-2.441***	-.222	-2.632***
Technological change			.274	2.399**	.189	1.788*
Stability of buyer-seller relationships			.129	1.911*	.189	1.699*
Business Strategies						
Isolate from rivals					.101	1.099
Extract temporary profits					-.243	-2.585**
Achieving a low cost position					-.099	-1.189
Build brand loyalty					.34	1.840*
Speed of response					.48	2.583***
Adj. R^2	.004		.170		.229	
F value	1.111		2.445***		2.988***	
Delta R^2			.166		.059	
F change	1.111		3.39***		2.901**	

*** alpha<.01, **alpha<.05, * alpha<.1

Table 3. Hierarchical Regression Results

LEA Firms. Dependent Variable: Return on Sales

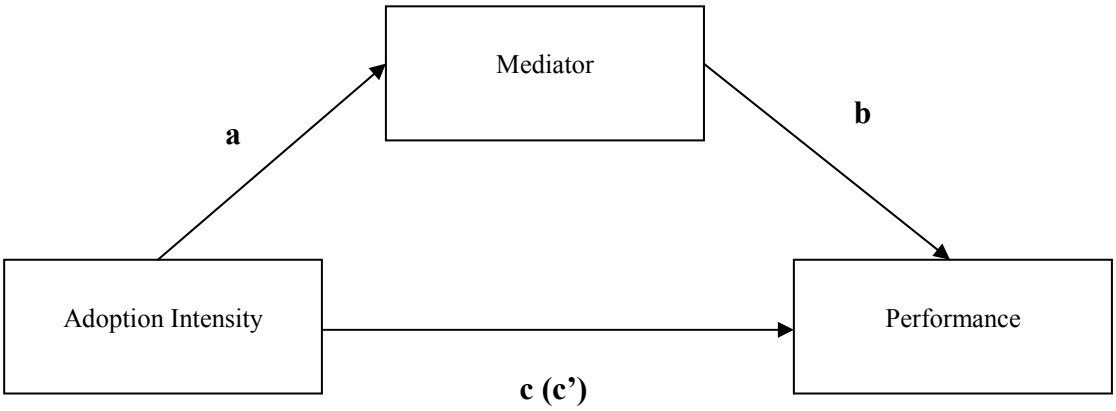
	Return on Sales					
	Model 1		Model 2		Model 3	
Control Variables	***β***	***t***	***β***	***t***	***β***	***t***
Years in business	.099	1.291	.051	.891	.167	-.092
Size	.087	.767	.098	1.000	.032	.089
Industry Structure Variables						
Product innovation			-.142	-.675	-.302	-.2348
Technological change			.333	1.001	.189	1.903*
Stability of buyer-seller relationships			.801	2.666**	.105	1.777*
Business Strategies						
Isolate from rivals					.200	1.629*
Extract temporary profits					-.137	-1.634*
Achieving a low cost position					-.122	-.022
Build brand loyalty					.099	1.333*
Speed of response					.043	-.408
Adj. R^2	.021		.147		.244	
F value	1.002		1.666**		2.732**	
Delta R^2			.126		.097	
F change	1.002		2.73***		2.809**	

*** alpha<.01, **alpha<.05, * alpha<.1

Table 4. Mediation Results

	Dependent variable	**Independent variable**	**Adj.-R square**	**β coefficients**	**ANOVA F-statistics**
Mediation I	Performance	Adoption Intensity Product Innovation	.18	.43* .01	26.02***
	Adoption Intensity	Product Innovation	.00	.01	
	Performance	Product Innovation	.00	.02	
	Performance	Interaction	.03	.08	
Mediation II	Performance	Adoption Intensity Technological Change	.23	.61* .22*	21.84***
	Adoption Intensity	Technological Change	.15	-.19	
	Performance	Technological Change	.16	.13*	
	Performance	Interaction	.10	.08	
Mediation III	Performance	Adoption Intensity Stability of Buyer-Seller Relationships	.25	.53* .21*	19.02**
	Adoption Intensity	Stability of Buyer-Seller Relationships	.18	.11	
	Performance	Stability of Buyer-Seller Relationships	.17	.16*	
	Performance	Interaction	.11	.17*	

Figure 1. The Mediation Model



Taguchi Analysis for Plasma Etching Optimization

J. S. Sutterfield, Tiffany A. Williams and Christopher R. Strong
School of Business and Industry
Florida A&M University

Abstract

One of the most powerful techniques used in experimental method is that of Factorial Analysis. This technique relies upon an experimental construct called the orthogonal array. In this paper we employ Taguchi methods with orthogonal arrays to perform Factorial Analysis aimed at optimizing a plasma etching process for integrated circuit components.

1. Introduction

Factorial Analysis as employed in experimental method is a very powerful and frequently used method for performing experimental analysis and arriving at valuable conclusions. Factorial Analysis relies upon a mathematical device known as orthogonal arrays. These were first studied by the great Swiss mathematician and physicist, Leonhard Euler (1707-1783), and were originally viewed as a type of mathematical recreation (Taguchi, 1988). Since Euler's time, they have been studied extensively by Joseph Leonard Walsh (1895-1973) and others as a part of the general investigation of orthogonal functions (Taguchi, 1988).

Factorial Analysis has been in use for upward of 70 years and was extended considerably through the efforts of Frank Yates (1937, 1970). Excellent discussions of the topic may be found in such works as Box, *et al* (1978), Cochran and Cox (1957), Fisher (1966) and Montgomery (2008). On the application and adaptation of Factorial Analysis to specific problems, one should see such works as John (1971, 1972), Margolin (1967, 1969), Plackett and Burman (1946). Taguchi (1988) has extensive discussions and examples of the various approaches to Factorial Analysis.

In factorial analysis, several experimental conditions, known as control factors, are systematically varied to determine how some factor of interest, known as the response factor, is affected by changes in the control factors. It must be emphasized that the control factors are changed systematically: This is why orthogonal arrays are vitally important in Factorial Analysis. An orthogonal array actually provides a template for conducting a factorial analysis. In orthogonal arrays, each column is, mathematically speaking, orthogonal to all others in the array. The experimental runs are randomized, which means assigning a random number to each row in the orthogonal array. An example of an orthogonal array is shown in Figure 1.

Figure 1: Example Orthogonal Array

No.	Run Number	1 A	2 B	3 AxB	4 C	5 AxC	6 BxC	7 e	Response Variable
1	4	1	1	1	1	1	1	1	RV_1
2	8	1	1	1	2	2	2	2	RV_2
3	3	1	2	2	1	1	2	2	RV_3
4	6	1	2	2	2	2	1	1	RV_4
5	1	2	1	2	1	2	1	2	RV_5
6	7	2	1	2	2	1	2	1	RV_6
7	5	2	2	1	1	2	2	1	RV_7
8	2	2	2	1	2	1	1	2	RV_8

In the above orthogonal array, three control factors, "A," "B," and "C" are examined along with their interactions. These interactions are identified as "AxB," "AxC" and "BxC." The "AxBxC" is not shown in above array, but it could be adapted to include this interaction. The next to last column, labeled "e," is where the random experimental error is to be found. The last column, labeled "Response Variable," contains the measured values for the response vari-

able. Further, the level of each control factor for a given experimental run is set by the corresponding entry in the orthogonal array. For example, in run number "6," factor "A" is set at its lower level, factor "B" at its upper level and factor "C" at its upper level. Moreover, the upper levels of all three interactions are to be found in this row.

In general, orthogonal arrays such as that in Figure 1 are identified by the number of the rows in the array, the number of factors and levels to be examined, or both. Thus, the orthogonal array in Figure 1 can be described as an L_8, referring to the number of rows; a 2^3 indicating that three factors, along with their interactions, are to be examined at two levels each; or as L_8 (2^7) indicating both. Again, if four factors, along with their interactions, were to be examined, each at two levels, an orthogonal array of sixteen rows and fifteen columns would be required: Such an orthogonal array would be designated as an L_{16} (2^{15}).

Orthogonal arrays have the curious property that if the numbers of the columns for any two of the principal factors are added together, the number of the column is obtained in which the interaction results for those two factors is found. For example, if factors "A" and "B" were of interest, their corresponding columns are "1" and "2." If "1" and "2" are added together the result is "3," meaning that "3" is the number of the column in which the interaction of "A" and "B," "AxB," is to be found. The reader may verify from Figure 1 that similar results will be found for the factor combinations of "A" and "C," as well as for "B" and "C." The application of orthogonal arrays does not limit Factorial Analysis to two factor levels: Orthogonal arrays have been developed three or more levels and for very large numbers of factors. As an example the designation L_{27} (3^{13}) would indicate an array with thirteen factors to be tested at three levels.

2. Methodology

To perform a Factorial Analysis, it is necessary first to decide upon those quantities (control factors) though to affect the phenomenon (response factor) to be investigated. One then decides upon the number of levels of each factor to be investigated, along with the range in which each control factor is believed to exert the greatest effect. Since experiments frequently involve a large number of factors, they are expensive to perform. Consequently, it is fairly commonplace to reduce experimental cost by performing initial experiments at two levels to identify those factors having the greatest effect, along with range at which of values in which they produce the greatest effect, as well as those having little to no effect. Once this is done subsequent experiments can be performed with fewer control factors using levels within the range found previously to have the greatest effect. It is worth pointing out that if it were desirable to investigate factors at differing levels, that it possible to do so. For example, if it were desired to investigate thirteen (13) different factors, say factor A at four (4) levels, factor B at three (3) levels, and the remaining six (6) factors at two (2) levels each, it would be possible to adapt a standard orthogonal array to accommodate such an investigation. It would be identified as an L_{16} (4^1 x 3^1 x 2^6) array. There are a vast number of arrays and adaptations: There are upward of 800 variations and adaptations of L_{16} alone (Taguchi, 1988).

Once the factors and their levels have been determined, and a satisfactory orthogonal array selected, the array itself becomes a template for conducting the experiment. The rows are *randomized* to conform to one of the canons of experimental method (Montgomery, *et al*, 2008), and the experiment executed according to this randomization. The experiment analyzed in the following section using Taguchi methods was conducted in this way.

3. Application of methodology

The experiment analyzed in this paper was originally performed by Yin and Jillie (1987) to determine the levels of those factors that optimize the plasma etching process for the production of silicone wafers used in integrated circuits. A schematic of a typical such process is shown in Figure 1. However, the data from the original experiment was not analyzed with analysis of variance methods. The motivation for this paper was to perform an analysis variance on the original data to arrive at conclusions for optimizing the manufacturing process.

Figure 1: A typical plasma etching process

This experiment was originally performed using a different experimental format and analytical technique. To make possible this analysis using Taguchi methods, it was necessary for our purposes to reformulate the original experiment into an L_{16} (2^{15}) orthogonal array. The reason for choosing this array was that it was desired to examine the interactions of any three of the four control factors being investigated. Thus, the L_{16} (2^{15}) array makes it possible to examine the variation of the four principal control factors, all possible interactions between any two control factors, all possible interactions among three control factors, with one remaining column for assignment to experimental error. This larger array made it possible to analyze interactions among control factors. The purpose of the original experiment was to analyze an etching process to determine which of four control variables, or combinations of control variables, most affected the depth of the etch achieved by the process. The factors which were thought most to affect the process are shown below in Table 2:

Table 2: Factors involved in plasma etching process

			Factor levels	
Factor	Units	Factor	Level 1	Level 2
Applied power	Watts	A	275	325
C_2F_6 Gas flow	SCm^3/min	B	125	200
Reactor chamber pressure	mTorr	C	450	550
Cathode gap	Cm	D	0.8	1.2

This experiment was aimed at analyzing a nitride plasma etching process, used to manufacture electronic wafers, to determine which control factors, along with possible factor interactions, were most prominent is the etching process. The etch rate was the response variable, and was measured in Angstroms per minute, designated as Å per min. (1 Å = 10^{-8}cm). Since it was desired to check all interactions up to an including third-level, an L_{16} (2^{15}) orthogonal array was selected: That is to say an array having 16 rows and 15 columns, where the last column was used to capture experimental error. The assignment of the control factors, along with the randomization of the runs is shown in Table 3.

Table 3: Orthogonal array for plasma-etching experiment

In Table 3, the first column indicates the number of the row in the orthogonal array. The second column indicates the order in which the combination of factors in the row was run. Columns 1, 2, 4 and 8 are the columns for control factors "A," "B," "C," and "D," respectively. The remaining columns through 14 are those where the levels of the interactions of these control factors are to be found. Column 15 is where

the random errors for the experiment are found. Column 16 is where the levels of the response factor are found, and column 17 contains the coded data for the response variable, obtained by deducting a value of 750 for the working mean. An experimental run is conducted by setting each control factor to the particular level indicated by the number for that control factor in the row.

The total variation is obtained by summing the squares of the coded values, and deducting the square of the sum of these divided by 16, the number of coded values. This is illustrated in the following calculation:

$$S_T = X_1^2 + X_2^2 + X_3^2 + \ldots + X_n^2 - \frac{(CF)^2}{n}$$

Then substituting the coded data for the response variable from Table 3 …

$$S_T = (-200)^2 + (-81)^2 + \ldots + (313)^2 + (-21)^2 - \frac{[-200 - 81 + \ldots + 313 + (-21)]^2}{16}$$

$$S_T = 531,421 Å^2$$

The effect for a given control factor is obtained by summing the values of the response factor for the "1s" in a given column, summing the values of the response factor for the "2s" in the column, taking the difference between the two sums, and squaring it. The result of this calculation is the variation for the effect is known as the variation. For a 2^n orthogonal array, the variation for any factor may be written as …

$$S = \frac{[(\Sigma RV_2) - (\Sigma RV_1)]^2}{n}$$

where …

RV_2 – the value of the response variable at the high level of the control factor in question

RV_1 – the value of the response variable at the low level of the control factor in question

n – the number of experiments performed

This computation is illustrated for control factor "A" as follows:

Σ(coded values corresponding with "1s" in column for A) = 1,433

Σ(coded values corresponding with "2s" in column for A) = -1,016

$$S_A = \frac{[(1,433 - (-1,016)]^2}{16}$$

The variations for the remaining control factors and their interactions are calculated similarly. The results for these calculations are shown in Table 4, the analysis of variance.

Table 4: Analysis of variance

Now it is evident from the Table 3 that only control factors "A," "D," and "AxD" are significant. Then, we combine the remaining, non-significant control factor variations, along with their interaction variations, with the error term. Further, the values of each of the three significant control factors are the gross variation. This means that each of these three factors contains the amount of one experimental error variation per degree of freedom. Since each of the significant control factors has one degree of freedom, each also contains one the amount of one experimental error variation. Thus, in order to obtain the desired net variation, we deduct the amount one experimental error variation. This adjustment results in the values for net variation shown in the column labeled ***S'***. Now, having combined the insignificant control factors and their interactions with the experimental error term, and having adjusted the gross variation

values to obtain net variation, the revised analysis of variation is obtained as shown in Table 5.

Table 5: Revised analysis of variance

Source	*F*	*S*	*V*	$F_0(95\%)$	*S′*	ρ
A	1	374,850	374,850	215.66	373,112	0.70
D	1	41,311	41,311	23.77	39,572	0.07
A&D	1	94,403	94,403	54.31	92,664	0.17
(e)	12	20,858	1,738	–––––	26,072	0.05
Total	15	531,421			531,421	100.00%

Now, the $F_{1,12}$ statistic at a 95% level of significance is a value of 4.75, and it may be seen from Table 4 that the least of the F_0 values, that for control factor **D,** is more than five times greater than the $F_{1,12}$ statistic. Further, $F_{1,12}$ statistic at a 99% level of significance is a value of 9.33. Again, the value for control factor **D** in Table 4 is more than 2.5 times greater than that for the $F_{1,12}$ statistic. Thus, the significance levels for two of the control factors in Table 4, as well as their interaction, are well beyond the 99% significance level.

We turn next to an estimation of the confidence levels for each of the mean of the response variable. In order to do this, it is necessary to estimate the average effects of levels of the above control variables. Then …

$$\overline{A_1} = 750 + \left(\frac{-1,016}{8}\right) = 623.0\text{Å}$$

$$\overline{A_2} = 750 + \left(\frac{1,433}{8}\right) = 929.1\text{Å}$$

$$\overline{D_1} = 750 + \left(\frac{615}{8}\right) = 826.9\text{Å}$$

$$\overline{D_2} = 750 + \left(\frac{-198}{8}\right) = 725.2\text{Å}$$

$$\overline{A_1D_1} = 750 + \left(\frac{-612}{4}\right) = 597.0\text{Å}$$

$$\overline{A_1D_2} = 750 + \left(\frac{-404}{4}\right) = 649.0\text{Å}$$

$$\overline{A_2D_1} = 750 + \left(\frac{1,227}{4}\right) = 1,056.7\text{Å}$$

$$\overline{A_2D_2} = 750 + \left(\frac{206}{4}\right) = 801.5\text{Å}$$

Thus, the desired etching effect is maximized when the control variables take on the values $\overline{A_2}$, $\overline{D_1}$ and $\overline{A_2D_1}$. Next, we wish to estimate the mean effect, $\hat{\mu}$, at optimal conditions. This mean is estimated as …

$$\hat{\mu} = \overline{A_2} + \overline{D_1} + \overline{A_2D_1} - 2 * \overline{T}$$

$$\hat{\mu} = 929.1 + 826.9 + 1,056.1 - 2 * 776$$

$$\hat{\mu} = 1,260.6 \text{ Å}$$

The general expression for the confidence limits of a statistical quantity is …

$$\mu = \hat{\mu} \pm \sqrt{F_n^m \, x \, V_e \, x \, \frac{1}{n_e}}$$

where …

$\hat{\mu}$ – the estimate of the true mean at optimal conditions obtained from the data

F_n^m – the F statistic at the significance level corresponding to the desired confidence limit with "m" degrees of freedom in the numerator and "n" degrees of freedom in the denominator

V_e – the experimental error variation

n_e – the effective number of experimental repetitions

Now, it will be recalled that the two control factors, as well as their interaction, were found to be significant at well above the 99% level. Now, using our estimate of the true mean as a value of 1,260.6, an F (99%) statistic of

9.33, V_e equal to 1,738, as in **Table 4** above, and an n_e equal to 8, for the calculation we have …

$$\mu = 1,260.6 \pm \sqrt{9.33 x V_e x \frac{1}{8}}$$

μ = 1,260.6 ± 45 Å/min.

Thus, the mean effect of the response variables, the average rate at which the plasma etching process is achieved, lies between 1,215.6 and 1,305.6 Å/min. with a significance level of 0.01.

4. Conclusions

The foregoing experimental analysis was a reformulation of a rather simpler factorial experiment performed several years ago in which the variations neither among significant control factors nor their interactions *were* investigated. As a matter of fact, the original work specifically states that a design was chosen which did not contemplate interactions. In the instant analysis it was desired to investigate interactions among control factors, and consequently it was necessary to use the original data in a recast orthogonal array, an L_{16} (2^{15}), in which the expanded array made it possible to study all possible control factor interactions up to and including third order (AxBxC). Whereas the original experiment was performed using classical methods, the reformulated experiment was analyzed using Taguchi methods. The authors prefer Dr. Taguchi's approach because it greatly facilitates the analysis of control factor interactions. As a result of examining the control factor interactions, it was determined that not only factors A (applied power) and D (cathode gap) were significant, but also their interaction, AxD.

In extending the original analysis the foundation has been laid for further experimentation on those control factors which were found significant. This experimentation would be aimed at determining the optimal values of these factors to maximize the etching rate.

Finally, it is important to make mention of the fact that as the number of factors in an experiment becomes large, the number of repetitions (rows) necessary to test all of the combinations of factors increases exponentially. Thus, it becomes expensive and unwieldy to execute such as experiment. I this case an experimental design known as the partial or fractional factorial may be employed to decrease the number of replications necessary for the experiment. For a discussion of the fractional factorial design, see Box, *et al* (1978), Cochran and Cox (1957), or Montgomery (2008).

5. References

[1] Box, G. E., Hunter, J. S., & Hunter, W. G. (1978). *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building* (1 ed.). New York: Wiley-Interscience, pg. 15.

[2] Cochran, W. and Cox, G. (1957). *Experimental Design. Second edition.*(Wiley Series in Probability and Mathematical Statistics)*. New York, NY: Wiley.

[3] Fisher, R. A. (1966). *The Design of Experiments, Eighth Edition*. New York: Hafner Publishing Company.

[4] John, P. (1971). *Statistical Design & Analysis Experiment*. New York: The MacMillan Company.

[5] Margolin, B. H. (1967). "Systematic Methods of Analyzing 2^n3^m Factorial Experiments with Applications," *Technometrics*, vol. 9, pgs. 245-260.

[6] Montgomery, D. C. (2008). *Design and Analysis of Experiments, 7th Ed.*. Main: John Wiley & Sons Inc, Somerset, New Jersey, U.S.A., pgs. 414-415.

[7] Plackett, R. L. and J. P. Burman (1946). "The design of Multi-factorial Experiments," *Biometrika*, vol, 33, pgs. 305-325.

[8] Taguchi, G. (1987). *The System of Experimental Design: Engineering Methods to Optimize Quality and Minimize Cost.* Tokyo: Quality Resources, Kraus International Publication, White Plains, NY, vol. 1.

[9] Yates, F. (1937). "The design and analysis of factorial experiments," *Imperial Bureau of Soil Science. Technical Communication*. London: Imperial Bureau of Soil Science, technical comment No. 35.

[10] Yates, F. (1970). "Selected Papers," Hafner (Macmillan), New York.

[11] Yin, G.Z., and Jillie, D.W. (1987). "Orthogonal Design for Process Optimization and its Application in Plasma Etching," *Solid State Technology*, pgs. 127-132.

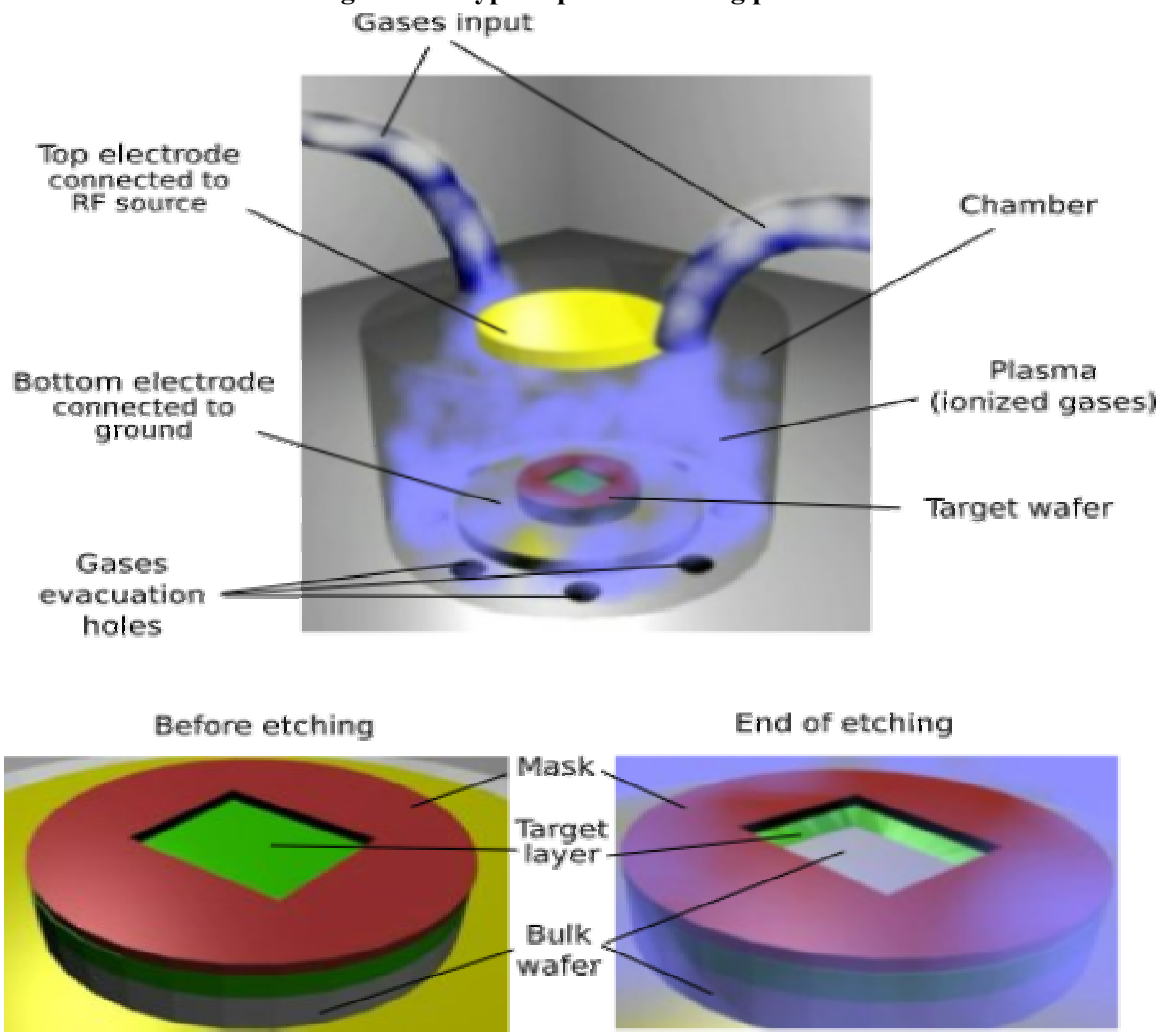
Table 3: Orthogonal array for plasma-etching experiment

No.	Run Number	1 A	2 B	3 AxB	4 C	5 AxC	6 BxC	7 AxBxC	8 D	9 AxD	10 BxD	11 AxBxD	12 CxD	13 AxCxD	14 BxCxD	15 e	Etch Rate (Å/min)	Coded Data
1	13	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	550	-200
2	8	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	669	-81
3	12	1	1	1	2	2	2	2	1	1	1	1	2	2	2	2	604	-146
4	9	1	1	1	2	2	2	2	2	2	2	2	1	1	1	1	650	-100
5	4	1	2	2	1	1	2	2	1	1	2	2	1	1	2	2	633	-117
6	15	1	2	2	1	1	2	2	2	2	1	1	2	2	1	1	642	-108
7	16	1	2	2	2	2	1	1	1	1	2	2	2	2	1	1	601	-149
8	3	1	2	2	2	2	1	1	2	2	1	1	1	1	2	2	635	-115
9	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1,037	287
10	14	2	1	2	1	2	1	2	2	1	2	1	2	1	2	1	749	-1
11	5	2	1	2	2	1	2	1	1	2	1	2	2	1	2	1	1,052	302
12	10	2	1	2	2	1	2	1	2	1	2	1	1	2	1	2	868	118
13	11	2	2	1	1	2	2	1	1	2	2	1	1	2	2	1	1,075	325
14	2	2	2	1	1	2	2	1	2	1	1	2	2	1	1	2	860	110
15	7	2	2	1	2	1	1	2	1	2	2	1	2	1	1	2	1,063	313
16	6	2	2	1	2	1	1	2	2	1	1	2	1	2	2	1	729	-21

Table 4: Analysis of variance

Source	*F*	*S*	*V*	$F_0(99\%)$	*S′*	ρ
A	1	374,850	374,850	215.66	373,112	0.70
B	1	218	218	———	———	———
C	1	11	11	———	———	———
D	1	41,311	41,311	23.77	39,572	0.07
AxB	1	18	18	———	———	———
AxC	1	2	2	———	———	———
AxD	1	94,403	94,403	54.31	92,664	0.17
BxC	1	7,700	7,700	———	———	———
BxD	1	2,475	2,475	———	———	———
CxD	1	248	248	———	———	———
AxBxC	1	2,576	2,576	———	———	———
AxBxD	1	127	127	———	———	———
AxCxD	1	68	68	———	———	———
BxCxD	1	977	977	———	———	———
E	1	6,440	6,440	———	———	———
(e)	12	20,858	1,738	———	26,072	0.06
Total	15	531,421			531,421	100.00%

Figure 1: A typical plasma etching process



The Journal of Management and Engineering Integration
Vol. 3, No. 1, Summer, 2010

A Resource-based Perspective of Porter's Five Forces and Total Quality Management: Their Role as Dependents for Competitive Advantage

Jason C. Van Pelt
California State University, Stanislaus
jvanpelt@csustan.edu

Abstract

Competitive advantage has been the subject of many examinations throughout the years. The studies however have failed to recognize the relationship between Porter's five forces model, the resource-based view of the firm, and total quality management in regards to creating a competitive advantage. This paper explains a method that firms can make use of to create a sustainable competitive advantage. The theory explores the relationship of the three theories (i.e. Porter's five forces, resource-based view, and total quality management) and explains that if each theory is used in conjunction with the others a competitive advantage can result.

1. Introduction

The choice of strategy is a vital decision in business. Despite the duration and position of an organization in its respective industry, a constant requirement exists to capitalize on the weaknesses of others. This study presents a method that will guide an organization through a series of components, each building on the one preceding it, which will develop an advantage that capitalizes on identified factors within the industry. This theory is performed in three steps.

The examination of an organization's industry will allow a broad-based strategy to be established based on Porter's five forces framework. From this broad-based strategy, the resource-based view (RBV) of the firm is incorporated which allows the organization to determine those resources that can be optimized based on strengths and opportunities as well as other organization's weaknesses which builds on the broad-based strategy decided earlier. This allows companies to exploit the competitive resources that are underutilized in effect creating a barricade around the organization. This barricade leads to the securitization of a greater piece of the market, increased revenue, and an improvement in overall financial stability. Once the resources are identified it is now possible to incorporate total quality management (TQM) initiatives that provide the action steps necessary to obtain a competitive advantage based upon those identified resources as well as the broad-based company strategy.

As the first component of this theory, Porter's five forces model will be discussed. Emphasis is placed on the environmental factors that provide the foundation for determining which of the three broad-based strategies will be chosen. Secondly, this paper will show how the resource-based view is an essential component to this theory from a resource utilization standpoint. This paper will discuss the requirements that the resource must meet. As the final component of this theory, TQM will be discussed as it pertains to the action steps required to complete an advantage for the organization. In the final section, we will show how each component of the theory acts as a foundation for the subsequent components of the theory.

2. Porter's five forces model

Since the theory's development in the 1970s, Michael Porter's five forces model has been an essential tool to analyze and determine a company-wide strategy. The tool facilitates the analysis of the industry-wide external factors that effect the organization's competitive position and profitability of the industry [11]. The factors include threat of new entrants, threat of substitute products, power of buyers and suppliers, and competitive rivalry. Once the assessment of threats in the industry is

completed, an organization can then determine which broad-based strategical path provides the most advantages for the organization to pursue. If an organization does not have a clear company-wide strategy, it stands to lose valuable market share and revenue [20]. The strategies identified in Porter's model include differentiation, low-cost producer, and focus strategies.

A differentiation strategy permits organizations with products that are thought of as holding the same value to distinguish themselves. Substitute products are a distinctive form of differentiation. During this strategy, an organization strives to provide an increase in value, which results in a colossal threat to existing products in the marketplace. A substitute product limits the growth of the industry and sets a ceiling on prices, thus limiting profitability [11]. Substitute products also encourage shopping, bundling of goods and services, and causes emotional purchasing decisions in its customers [26]. For these reasons, it is extremely advantageous for an organization to eliminate the threat of substitute products in their industry. "This incentive to differentiate products from those of competitors stems from the fact that sufficiently novel products may create monopoly power and thus economic profit" [9].

The attainment of the cost-leadership strategy is largely dependent upon delivering the customer's expected level of value while ensuring the company is at an adequate level of profitability through the attempt to achieve the lowest per-unit cost of production [6]. When an organization obtains a low-cost level of production relative to its competitors, it is often difficult for others in the industry to make the necessary adjustments in their operations to ensure continued competitiveness. On the other hand, an organization can deter this threat through its anticipation and react in a way that creates this strategy for themselves.

The focus strategy outlined by Porter [20] is based on organizations focusing on a niche segment within their industry. Through this focus, they achieve an advantage in a highly profitable segment in their industry. This strategy cannot be employed concurrently with the other strategies. "The benefits of optimizing the firm's strategy for a particular segment (focus) cannot be gained if a firm is simultaneously serving a broad range of segments (low-cost leadership and differentiation) [20].

3. Resource-based view of the firm

The resource-based view (RBV) of the firm declares that a competitive advantage can be achieved by deploying valuable resources and capabilities [3] [17] [23]. In a competitive environment an organization's unique resources and capabilities provide them the opportunity to develop a competitive advantage so long as certain conditions are met. The conditions within RVB that are required in order to establish a competitive advantage are preceding limits to competition, resource heterogeneity, imperfect resource mobility, and subsequent limits to competition [17]. These broad conditions are based on the distinct resources currently available to the firm and assists in identifying which new resources can be developed from gaps within the industry that will result in a competitive advantage.

This process involves the examination of RVB competitive conditions to determine if an organization can obtain a competitive advantage through the deploying of resources in a more efficiently or effectively manner. The examination of internal and external environmental factors is considered. These internal factors are strengths and weaknesses that come from the management, employees, equipment, and processes within the organization while the external factors are those opportunities and threats that reside outside the organization. External factors are items such as new legislation, environmental factors, and the health of suppliers. Furthermore, if there are gaps in the analysis that competitors have not filled with their own resources, then an organization can deploy their strengths to take advantage of the opportunity.

Prior to attempting the development of any resource, an organization must ensure there is no competition for that same resource from the

beginning. These resource barriers hold the potential for high returns to the one competitor who holds the advantage [3]. If competition exists, the resource will yield no significant benefit to the organization. If the organization is late to acquire a resource, or has subsequent competition in acquiring a resource, it may not be feasible to undertake, unless that resource is a means to survival.

The establishment of a resource barrier is dependent upon the heterogeneity of the resource; if the heterogeneity is high then a resource barrier is feasible. Heterogeneity is a vital part of the components that determine whether an advantage can be obtained and has two distinct parts. If the resource can be easily acquired by another organization then there is essentially no advantage to the primary implementing organization. Furthermore, the ease of implementation of that resource is another obstacle to overcome. Any acquiring organization must be able to implement the resource to be able to realize the advantages. Therefore, a primary acquiring organization can either make it more difficult for another organization to acquire and/or implement the resource.

The mobility of a resource refers to the ease of transfer from one organization to another. If a resource can easily be transferred, that is, the resource is perfectly mobile between organizations; the early adopters will forfeit the advantages of the increased efficiency and effectiveness of their operations. If efficiency and/or effectiveness are not present, the resource is no longer valuable [14]. The consequence will be increased competition resulting in even distribution of financial stability within the market.

An imperfect mobile resource requires greater effort to implement and thus provides more value to the firm [17]. Imperfect mobility allows the foremost implementing organization to capitalize on competitor's limitations, creating a less attractive industry, resulting in a withdrawal of competitors and a more difficult entry into the market. Furthermore, if a resource can be employed that is imperfect, the result is fewer attempts from competitors to initiate that resource themselves.

After obtaining an advantage through the implementation of a resource, there must be little competition from other firms in attempting the same resource advantage. If the resource can be acquired, the rents realized through the resource may not offset the cost of acquiring the resource in the first place [17]. The return on investment must be considered to make a rational decision when considering the implementation of a particular resource. An analysis of the competitor's position and a determination of their strengths and weaknesses will aid the implementing organization in determining just how sustainable their advantage will become.

As the resource-based view is a dependent theory, the conditions are reliant upon one another [3]. The first condition is satisfied when heterogeneity exists. Further strengthening heterogeneity is the lack of mobility. Furthermore, an acquired resource must fit within the business model of each organization. If the resource is determined to be of benefit, available to acquire, but does not fit within the model, then a company can choose to change the model or acquire a different resource that compliments their strengths. Given that altering a business model is not only costly but also immensely risky, being the first to acquire the correct immobile resource acts as a deterrent to other organizations. Finally, subsequent limits to competition must exist for the advantage to be sustainable. If the advantage is not sustained, the yield of the resource declines in proportion to the number of competitors that employ the resource themselves and the length of time the resource is employed by competitors. The minimum timeframe a resource advantage must be sustained is the breakeven point. It is essential to choose a strategy that ties these four components together.

4. Total quality management

The practice of TQM set forth by W. Edwards Deming, Joseph Juran, and Kaoru Ishikawa has been accepted by United States management as a way to enhance the perception

of quality and value by focusing on the customer. Research indicates that TQM generates improved products and services, reduce costs, creates more satisfied customers, and improves profitability [16] [24].

TQM incorporates every aspect of an organization to achieve atypical outcomes based on three basic principles: open culture, employee empowerment, and executive commitment [21]. These principles contribute to a competitive advantage as they have proven to be imperfectly imitable features [21]. TQM serves as an appropriate tool for organizations because:

- all businesses are composed of an inter-dependent chain of suppliers and customers,
- the market place demands that all businesses are responsive to the needs of their customers,
- all industries have been affected by lower productivity, increased litigation, more competition, and adherence to traditional practices,
- provides a clear understanding of company standards and expectations
- focused on how an enterprise can better serve the customer through a search for excellence in every aspect of the operation [6].

As the last component of this theory, TQM serves as the implementing feature that is required to obtain a competitive advantage. In other words, TQM serves as the action steps that are necessary to put the other analysis into action.

5. The link of dependents

This theory links Porter's five forces model, RBV, and TQM. Each component acts as a foundation for the others to build upon to work towards a sustainable competitive advantage (see figure one). The first component of this theory is to analyze the state of the industry. What is limiting profitability? Are there weaknesses within the industry that could be capitalized? Do threats exist that could limit our survival period?

The company-wide strategies within Porter's model have the ability to create a competitive advantage [6] [22] [28]. Through the analysis of the industry and the determination of a company-wide strategy, the groundwork has been laid from which to build upon to transform the strategy into actionable results. If the idea of a company-wide strategy is used alone it will not be sufficient to gain ground on leaders and others within the industry. Further analysis is required to determine whether the organization possesses any resources that can be utilized to capitalize on the industry's opportunities identified through Porter's model.

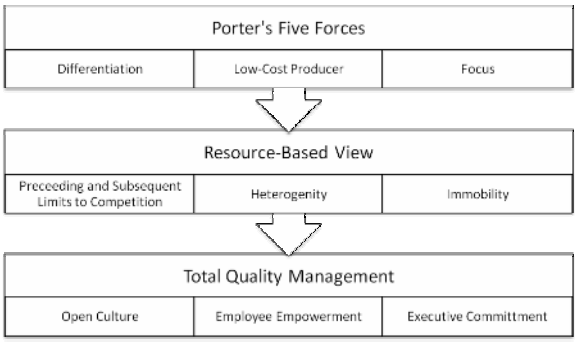


Figure 1. Decision matrix

Based on the company-wide strategy (i.e. differentiation, low-cost producer, or focus) the resources within an organization must be developed to capitalize on the opportunities within the industry. The four criteria of RBV are used at this point to determine exactly which resource does not have competition either before or after implementation and is immobile and inimitable. Once resources are identified, which meet the requirements of RBV and strengthens the company-wide strategy identified by Porter's model then the action steps can be developed to make the plan a reality.

"The failure of many firms' strategies stems from an inability to translate a broad competitive strategy into the specific action steps required to gain a competitive advantage" [20]. TQM procedures will allow the resources that have been identified to be implemented resulting in the precise steps needed to transform a generic strategic outlook into a resource driven set of processes.

TQM implementation has been popular in recent decades [1] [24]. However, even if several firms in an industry have attempted and perhaps succeeded in implementing TQM advantages still exist. As long as the number of firms that acquired a resource is less than the number of firms required to create perfect competition, a competitive advantage can still be attained [12]. One of the largest motivations for TQM implementation is when leaders of other organizations see what can happen when TQM is a success. The implementation phase is perhaps the largest single factor that determines if TQM will be a success in an organization [15] [25] [27].

6. Conclusion

Porter [19] continuously warns of being stuck in the middle. By being stuck in the middle, an organization has no clear strategy nor is focusing on any portion of the market to realize substantial gains. By utilizing Porter's model as a foundation to understand which necessary actions require implementation, other processes, and techniques can contribute that create a competitive advantage for the organization. By deploying this process, an organization has a guide to take the steps necessary to achieve a strategy that is comprehensive enough to create a competitive advantage.

Outlining this model of securing a competitive advantage through the combination of Porter's five forces, the RBV, and TQM allows an organization to not only formulate a broad strategy, but more importantly provides the action steps necessary to carry out that strategy, based on strengths through resources. By having more focused action steps, an organization will have a much more successful chance in their industry to gain market share, increase financial stability, provide jobs, and ultimately become the leader in their industry.

This work provides a basis from which further research can be derived. Suggestions include a more specific analysis of TQM and the action steps that can lead to competitive advantage. Additionally, although lengthy in nature, an empirical examination can be undertaken through a competitive analysis of a particular industry and the effectiveness of TQM programs.

7. References

[1] A. Durity, "Everyone is a Winner in the Baldrige Stakes", Management Review, 1991, 80, pp. 36–39.

[2] A. McGahan and M.E. Porter, "How Much Does Industry Matter, Really?", Strategic Management Journal, 18, 1997, pp. 15-30.

[3] B. Wernerfelt, "A Resource-Based View of the Firm", Strategic Management Journal, 5, 1984, pp. 171-180.

[4] Bank, J., The Essence of Total Quality Management (2nd ed.), Prentice-Hall, New Jersey, 2000.

[5] Burke, G., Microeconomics: Principles, Problems and Policies, McGraw-Hill/Irwin, New York, 2005.

[6] D. Miller, "Relating Porter's Business Strategies to Environment and Structure: Analysis and Performance Implications", The Academy of Management Journal, 31, 1988, pp. 280-308.

[7] D. Miller and P.H. Friesen, "Porter's Generic Strategies and Performance", Organization Studies, 7, 1986, pp. 37-55.

[8] Evans, J.R. and W.M. Lindsay, The Management and Control of Quality (3rd ed.), West Publishing Company, St. Paul, MN, 1996.

[9] G. Haubl and V. Trifts, "Consumer Decision Making in Online Shopping Environments: The Effects of Interactive Decision Aids", Marketing Science, 19, 2000, pp. 4-21.

[10] Hardjono, T.W., S. Have, and W.D. Have, "The European Way to Excellence", European Quality Publications, London, 1997.

[11] Ishikawa, K., What is Total Quality Control? The Japanese Way, Prentice-Hall, New Jersey, 1985.

[12] J. Hirshleifer, Price Theory and Its Applications (2nd ed.), Prentice-Hall, New Jersey, 1980.

[13] J. Sutton, "Vertical Product Differentiation: Some Basic Themes", The American Economic Review, 76, 1986, pp. 393-398.

[14] J.B. Barney, "Firm Resources and Sustained Competitive Advantage", Journal of Management, 17, 1991, pp. 99–120.

[15] J.V. Saraph, P.G. Benson, and R.G. Schroeder, "An Instrument for Measuring the Critical Factors of Quality Management", Decision Sciences, 20, 1989, pp. 810-829.

[16] M. Terziovski and D. Samson, "The Link between Total Quality Management Practice and Organisational Performance", International Journal of Quality & Reliability Management, 16, 1999, pp. 226-237.

[17] M.A. Peteraf, "The Cornerstones of Competitive Advantage: A Resource-based View", Strategic Management Journal, 14, 1993, pp. 179-191.

[18] O. Aktouf, "The False Expectations of Michael Porter's Strategic Management Framework", Revista Gestato e Planejamento, 2005, pp. 75-93.

[19] Porter, M.E., Competitive Advantage, Free Press, New York, 1985.

[20] Porter, M.E., Competitive Strategy, The Free Press, New York, 1980.

[21] Powell, T. "Total Quality Management as Competitive Advantage: A Review and Empirical Study", Strategic Management Journal, 16, 1995, pp. 15-37.

[22] P.R. Dickson and J.L. Ginter, "Market Segmentation, Product Differentiation, and Marketing Strategy," The Journal of Marketing, 51, 1987, pp. 1-10.

[23] Reid, G., Small Business Enterprises: An Economic Analysis, Routledge, London, 1993.

[24] S.L. Ahire, D.Y. Goihar, and M.A. Waller, "Development and Validation of TQM Implementation Constructs", Decision Sciences, 27, 1996, pp. 23–56.

[25] S.M. Shortell, J.L. O'Brien, J.M. Carman, R.W. Foster, E.F. Hughes, H. Boerstler, and E.J. O'Connor, "Assessing the Impact of Continuous Quality Improvement/Total Quality Management: Concept Versus Implementation", Health Services Research, 30, 1995, pp. 377-401.

[26] T. Grundy, "Rethinking and Reinventing Michael Porter's Five Forces Model," Strategic Change, 15, 2006, pp. 213-229.

[27] T.K. Douglas and W.Q. Judge Jr., "Total Quality Management Implementation and Competitive Advantage: The Role of Structural Control and Exploration", Academy of Management Journal, 44, 2001, pp. 158-169.

[28] W.R. Smith, "Product Differentiation and Market Segmentation as Alternative Marketing Strategies," The Journal of Marketing, 21, 1956, pp. 3-8.

Cost Consideration of a Supply Chain Using Ontologies

Nabeel Yousef, Nabin Sapkota, Jose Sepulveda, and Luis Rabelo
University of Central Florida
nyousef@mail.ucf.edu

Abstract

The purpose of this research is to create a general framework that can express the cost data for the partners of the supply chain in similar terms. A framework that can layout the common activities identified within the firm and the relationship of these activities between the partners of the supply chain. The framework should also be able to layout the effect of changes in activities on other partners within the supply chain. Cost information will help in making decisions about pricing, outsourcing, capital expenditures, and operational efficiency. It should enable the organizations to compare directly the costs of certain activities.

The framework will be able to track cost through the chain, which will improve the flexibility of the supply chain to respond to rapidly changing technology. The framework will help in developing product strategy paradigms that encompass the dynamics of the market, in particular with respect to the technology adoption lifecycle. It will be used for forecasting and decision making in product related problems. The proposed framework will have the ability to detect successful products and the driving performance measures within the supply chain that lead to this success.

1. Introduction

In today's world, since competition is not localized, most of the companies have to compete globally as the newer technologies are constantly emerging with the market behavior becoming even more dynamic. The flexible supply chain has become the core of competition in the new local or global market. To maintain a competitive edge over the similar businesses one has to have a cost effective and flexible supply chain structure that not only ensures that the company to be able to compete successfully in the market but also has enough flexibility in the chain structure to accommodate required changes promptly without losing any market share or time. In recent years, focus has been on how to manage the supply chain that is not only flexible enough but also provides a competitive edge over the competitors and ensures successful survival in the dynamic market conditions.

The supply chain is the network of customers, distributors, transporters, storage facilities, manufacturers, and suppliers that participate in the sale, delivery, and production of a particular product. There are definitions of the supply chain by many companies, practitioners, and organizations. Ganeshan and Harrison (1995) defined the supply chain as "a network of facilities and distribution options that performs the functions of procurement of materials, transformation of these materials into intermediate and finished products, and distribution of these finished products to customers". William C. Copacino (1997) defines the supply chain as the flow of materials and products from source to user. Late 80's and the early 90's saw the immense rise of the term Supply Chain

Management (SCM). Cooper et al. (1997) define supply chain management as the management of the flow of production from raw material to the finish product including the disposal process. This process includes order generation, order taking, information feedback, and the efficient and timely delivery of goods and services. The basic purpose of SCM is to reduce total costs through the chain that includes raw material and acquisition costs, logistics cost, facility and manufacturing costs, and distribution costs (Shapiro, 2001).

In § 2, a brief explanation of several relevant works are mentioned followed by our cost framework using ontologies for SCM in § 3 and we conclude in § 4.

2. Literature Review

Handfield and Nochols (1997) highlighted various important issues related to the supply chain management (SCM), significance of SCM, and why the integrated approach to SCP is necessary for the future competitiveness in the market. Cooper, Ellram, et al, (1997) define SCM as the practice of coordinating the flow of goods, services, information, and finances as they move from supplier to manufacturer to wholesaler to retailer to consumer. This process includes order generation, order taking, information feedback, and the efficient and timely delivery of goods and services. As Cokins(19990 stated, the four essential ingredients for successful SCM are driven continuous replenishment, electronic commerce (bar coding), category management, and the true cost information activity based costing (ABC). Even though there has been a lot of research in the true cost information activity, there are still many opportunities to improve and issues to be to be tackled. In a good SCM, helping the partners in the chain to understand their internal costs by tracking cost information through the supply chain is very important as it assists in the coordination and optimization of activities across firms in a value chain. Accurate cost information helps the managers to understand the profitable products and services and to devise newer strategies to increase future profits. Nevertheless, these strategies should be extended to all the members in the chain acting as a team for the common goal of making profit. This kind of collaboration and the information leverages to perform as a value chain. Understanding the processes along the supply chain will determine any low performing partners. In this event the chain may opt to replace the weakest of the partners. Each partner in the chain creates a cost for the others that can be reduced to improve the overall total cost. Lots of opportunities are available to improvise saving in the cost within the chain. That is the reason more and more firms, these days, are interested in establishing inter-relationships between each other. The relationships allow for the exchange of information and cooperation but also need a commitment and acceptance of the risk and rewards of partnering (Dekker and Goor, 2000). Consequently, partnering helps to create more efficient activities within the supply chain and provides better outcomes, which increase the competition among chains and individuals within the chain (Cooper et al, 1997).

Primary research interest in value chain analysis is to identify the different opportunities for cost savings. Analyzing cost has become harder and more complex. Varieties of products processes, technologies, labor efficiencies, taxation (national and international), local laws, geographic distribution of members in the

chain, environmental constraints, etc. necessitate many scenarios to be considered for cost analysis. It is very difficult to predict the influence of all these cost factors associated with each product or service which adds complexity in the analysis. Other complexities arise because of the fact that most of the cost models heavily depend on the historical data which make it is very hard to predict the cost associated with the new product processes with this scheme. Another complexity in cost analysis comes when comparison of alternative designs and the evaluation of strategic choices which have to be carried out.

Additionally, much of the literature has been narrow in scope and has addressed very specific problems in cost consideration in the supply chain. No literature has addressed the need for a general framework for a cost model that can be used as a standard template in supply chain cost management and optimization. The desired framework should express the cost data in similar terms that can be understood across firms, firm units, and the activities within.

This research proposes a framework that can be used as a template that can integrate existing supply chain models to monitor and analyze cost within the chain. The framework will drive future profits by creating newer strategies. The framework ontologies will help in defining and distributing cost centers. The ontologies will aid in information sharing through organization units while minimizing redundancy.

Ontologies will coordinate conflicting goals and objectives and improve performance. Using ontologies will aid in coordinating cost, quality, and efficiency while controlling excessive inventory and backlogs. Understanding the concepts and the activities will minimize uncertainties in production planning as well. (Ung Min, Bjomsson, 2002).

Ontology is a philosophical word that describes the nature and organization of being or reality. The notion of ontology as a formally specified conceptualization shared by a community of practice is now well established and is used and applied in several areas, including knowledge management, knowledge acquisition, information retrieval and extraction, knowledge engineering, and knowledge modeling. Ontologies will be used to define and specify the main domain concepts for the framework developed in this research. The essential role of ontology is to support reuse, which can take place in different scenarios. For instance, ontologies have been used to support the specification of reusable libraries of problem-solving components, to drive model-based knowledge acquisition, to allow semantic information retrieval, and to structure collaborative decision-making processes. Based on Gruber (1993), ontology is a formal, explicit specification of a shared conceptualization. Also Fishwick and Miller (2004) defined ontology as a knowledge representation used to capture information and knowledge about a subject. Gomez-Perez (1998) identified four main types of ontologies: domain ontologies, task ontologies, meta-ontologies, and knowledge ontologies. Domain ontologies are defined as the ontologies that provide a vocabulary for describing a particular domain which is a specific area of knowledge that is typically the focus of a certain group or community. Task ontologies provide a vocabulary for the terms involved in a problem solving process. Meta-ontologies provide basic terms to arrange domain and task ontologies. Knowledge ontologies provide tools and

methods for supporting knowledge management, relying on sharable and reusable knowledge. The supply chain cost model is considered domain ontologies, which describe the terms, standards, definitions, and their relationships through the supply chain. Concepts and terms include partners, activities, resources, cost drivers, and their relationships to produce the total cost of a product.

The purpose of the cost model ontologies is to provide common understanding of the product cost through the supply chain from raw material to the finished product. The ontologies will document the concepts with modeling primitives and semantic relationships. The model ontologies should be expressive enough to represent the formal semantics, which will help in sharing, using and reusing information (Lacy, 2005).

Fayez (2005) developed an approach to construct the Supply Chain Ontology using a standardized Ontology language. He used the Supply Chain Operation Reference model (SCOR) as guide in building Supply Chain Ontology. Due to some of the limitation and deficiencies of the SCOR Model, he developed and added a new layer in the ontology, which addressed all new supply chain concepts while enhancing the scope of the SCOR model. Fayez (2005) stated that the improved supply chain representation would provide the ability to integrate a supply chain cost model such as activity-based costing which could be used to improve cost representation across the supply chain.

Activity Based Costing (ABC) is an accounting methodology that assigns costs to activities rather than products or services. In order to correctly associate costs with products and services, ABC assigns costs to activities based on their use of resources, and then assigns costs to cost objects. The process of identifying the activities and their relevant costs is called Value Chain Analysis. Additionally,the revenues and assets would be assigned to these activities.

The economic behavior of the supply chain in the Value Chain Analysis can be understood by location the cost drivers for each activity. The activity is defined as any process or task that happens over time and has a role in the chain.

Activities are considered the common factor between business process improvement and information improvement. This is why it is important to understand the activities within the firm or group of firms in order to improve the flow of material between different units.

3. Development of the framework

3.1. General approach of development of framework

In order to develop a standard framework for a supply chain, the first step is to collect all domain information using historical data and data currently being used and to understand the different concepts in the supply chain. Experts are asked to verify the exhaustiveness of the domain for that particular supply chain in consideration. Secondly, each concept in the domain is identified and defined in terms of its functionality, and then its associated characteristics are developed as class diagrams showing the interaction among concepts. Thirdly, each concept is defined and all the attributes are listed. Sub concepts are the concepts that inherit some of the attributes from the related main concept and have other specific attributes of their own. All concepts, their attributes, and their definitions are then entered into the software later. The final step before the

ontology is developed is to define a relationship between these concepts. Then the ontology for the supply chain is developed, and the cost model based on ABC analysis is added into the framework. All the information is saved in a base knowledge for everyone to share, which constitutes the general framework. Figure 1 shows the approach that has been used to develop the framework. The ontology and the cost model have been explained under separate subsequent sub-headings in this section.

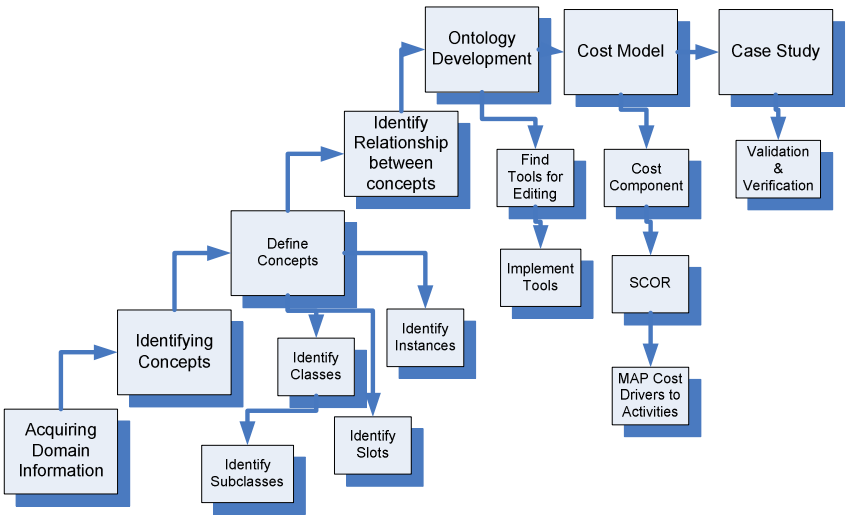


Figure 1: Approach to develop standard framework

3.2. Ontology development

Ontology for the supply chain is considered domain ontology. Defining the concepts will minimize redundancy and improve relations among different units within the supply chain. Different tools can be used to build and edit ontologies and ontology-based annotation using pre–defined concepts in ontology to mark–up a document, and ontology based reasoning. Some of the common software packages that can be used for ontology editing are: Protégé 2000, OntoEdit, and OILEd (Fensel et al, 2001). Protégé-2000 is open source software with an efficient graphical user interface (GUI); hence, we chose to use Protégé-2000 as a tool to develop the cost model ontology in this research.

Protégé-2000 is an integrated software tool developed by Stanford University. It has been used by system developers and domain experts to develop knowledge-based systems. Applications developed with Protégé are used in problem-solving and decision-making in a particular domain (Protégé-2000 user guide). The software has a uniform GUI whose top level consists of overlapping tabs for compact presentation of parts and for convenient co-editing between them.

In this research the SCOR model is used to identify the main concepts keeping the ontologies within a frame of standards. Customers, as one of the partners in the supply chain, trigger demand for a certain product. The chain consists of several processes (plan, source, make, deliver and return) and associated activities that will assist in the product flow. Moreover, each of these activities has one or more associated cost drivers and requires several resources to

accomplish these activities. The framework for cost ontologies is shown in the following Figure 2.

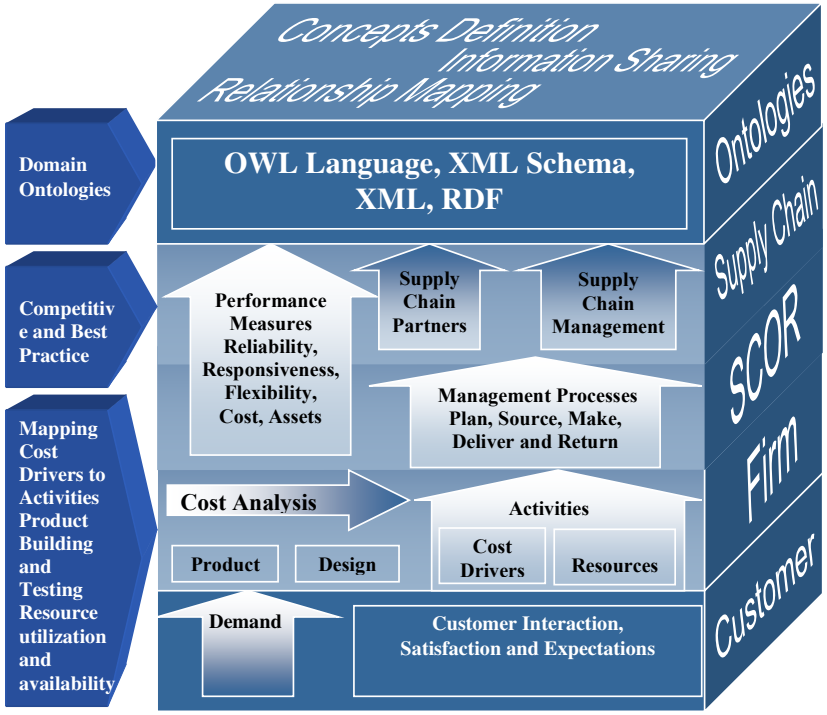


Figure 2: Framework for Supply Chain Cost Ontologies

Partners are firms, businesses, companies, manufacturer, and end user that participate in the flow of a certain product through a chain of activities. Suppliers supply the chain with the raw material, manufacturers transform raw materials into finished or semi-finished goods, and retailers sell finished goods to end users, and carriers move raw-materials or finished/semi finished goods within the chain. Within these partners there are lots of activities that support the product building process until the finished goods are delivered to the end users. For a nontraditional supply chain that may have more or different partners, a partner called "other" has been added to the ontology concepts to provide flexibility in modeling.

SCOR describes five main management processes for each of the partners as plan, source, make, deliver, and return. Within these processes are several but common activities that drive the product building through the chain, such as purchasing, sales, marketing, scheduling, etc. Some of these activities are considered main activities, while others are sub activities. For example, purchasing raw material is considered a main activity, while preparing a purchase order is a sub activity that supports the main one. The activities are operated and achieved by using resources such as people, equipment, facilities, land, money, and raw material. In building the ontology an additional term called other resources has been used to govern the flexibility of the adding other resources not yet encountered. The cost for each activity is driven by cost drivers.

The cost drivers are defined as events or actions that trigger the activity and can be the basis for the calculation of unit costs. They can also be defined as the factors responsible for variation in the cost of an activity. The cost drivers can be frequency measures, duration measures, physical measures or other measures that might not be included under any of the previous measures, such as time of year. The time of year or the calendar measure can affect the sales of a product, which might affect the primary cost. In the model ontologies, different types of products have been added, such as physical products, services, or other products. The "other" product term adds flexibility to incorporate future products not yet classified into the model. Cost concept will include the types of cost encountered in the supply chain. Figure 3 summarizes the cost model ontology hierarchy as developed in Protégé 2000.

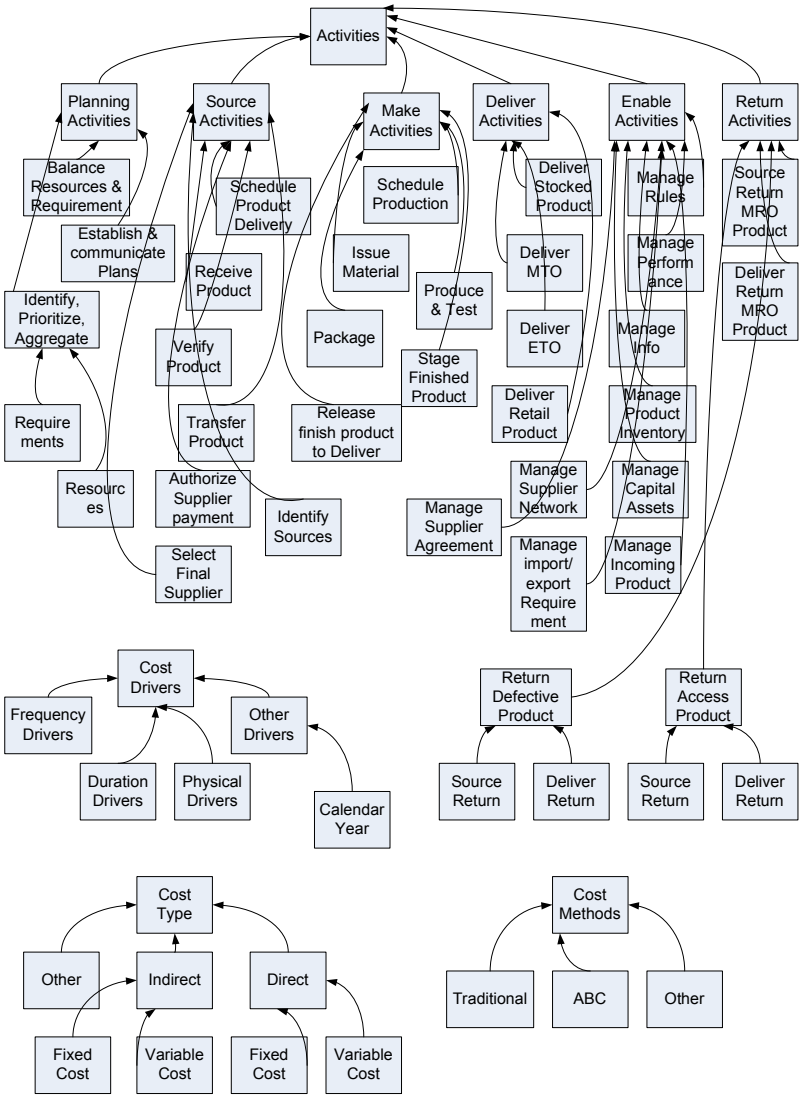


Figure 3: Cost model ontology hierarchy using ABC technique

In this step the activities could be classified into added value activities and non-added value activities. A decision has to be made about the non-added value activities either to eliminate or to minimize the effect of such activity on the total price of the product.

3.3. ABC cost model

Once the activities are identified, resources will be assigned to each activity. ABC will help in recognizing the actual dollar costs against individual activities. Activities could be classified into value-added activities or non value-added activities. A decision has to be made about the non-added value activities. It opens up opportunities for cost reduction as these non value-added activities with proper planning or modification in the process that has such activities can be either eliminated or can be minimized. Identification of non value-added activities and taking corrective actions against them bring a positive impact on the total price of the product. Additionally, the built ontologies help in analyzing activities within the supply chain. Each activity is defined as a class or sub-class, and resources are assigned to these activities. The ontology of the model also provides the definition of cost drivers and there measures. The next step will be to trace cost to activity to make sure that no cost was lost during the gathering. Then establish output measures and analyze costs for each product based on the activities that a certain product or service goes through.

4. Integrated framework

An integrated framework consists of extensive base knowledge for various activities that are required to be executed in order to accomplish various processes as explained in the SCOR model for the supply chain. Base knowledge also consists of relevant information about all the members in the supply chain. For example, information in a base knowledge for one of the suppliers in a supply chain has different parts or raw materials which it can supply, their rates, lead-time, etc. For each and every kind of activity, relevant cost drivers (frequency of task, duration of task, etc) and resources required are recorded so that costing can be done at supply chain level.

Supply chain ontologies with ABC costing is developed in Protégé that has source code written in Ontology Web Language which in turn is a customized version of xml for ontology development application. Therefore, any local application such as Visual Basic for Application (VBA) code can read base knowledge in xml and fetch knowledge (required data/information) from the base knowledge and transfer it to the spreadsheet application similar to Microsoft Excel. End users can select various configurations for any product from the drop-down menu in the GUI developed in VBA. Local application like VBA reads relevant data from the base knowledge as per the input of the end user in GUI and passes it on to the excel sheet where costs are computed based on ABC technique. Upon completion of the calculation of costing, cost is displayed to end users, cost by various categories or various costs by activity in a graphical form. End users need write any code or be

knowledgeable of spread-sheet application. The general working of the integrated framework with the end user application is shown in Figure 4. This way end users can try various scenarios to simulate different configurations for the existing or new product and get information such as whether the product is profitable by addressing such concerns as; which product among others is the least profitable, which activity can be reduced or eliminated, which partners are to be dropped from the chain, what is best configuration for the product to be competitive in the market, etc. Once the integrated model is in hand, the end user can have insights about the costing of the product by various categories or by activity which otherwise is impossible without integration of cost model in the supply chain ontology.

Since the base ontology is in OWL, it can be made to read relevant information about the partners or other relevant information directly from the internet, WAN, or LAN and update the base knowledge periodically.

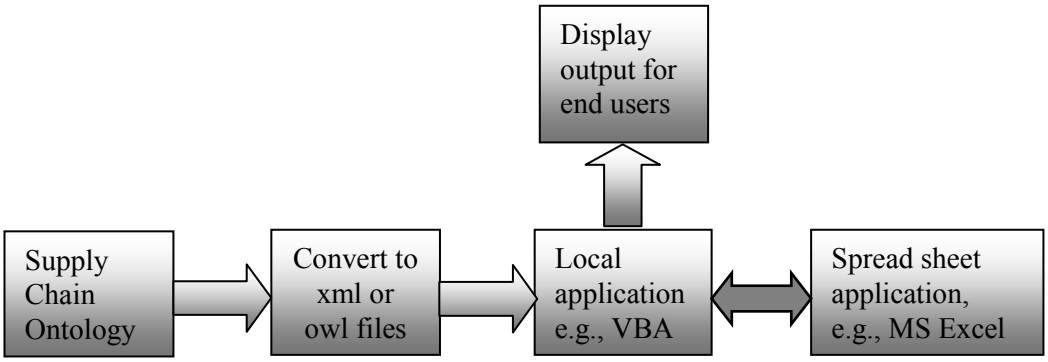


Figure 4: The general working of the integrated framework with end user application

5. Conclusion

Cost information is very important within the chain of firms that participate in producing a certain product. The information will help in making decisions about pricing, outsourcing, capital expenditures, and operational efficiency. The captured information should improve competences between supply chains. It should enable the organizations to compare directly the costs of certain activities. The framework being proposed in this research will be a template that can integrate existing supply chain models to monitor and analyze cost within the chain. The framework will drive future profits by creating newer strategies. The framework ontologies will help in defining and distributing cost centers. The ontologies will aid in information sharing through organization units, which will minimize redundancy. The standardization will help in:

- Improving comparability of cost data between firms
- Creating a more flexible chain that will react to the dynamic daily changes
- Creating communication between different management information systems
- Extending the cost reduction to the supply chain level

- Restructuring the activities to be integrated with each other across the firms of the supply chain
- Driving future profits by creating newer strategies
- Recognizing the profitable products and services and eliminating the non profitable ones.

References:

Armstrong, Peter. "The costs of activity-based management." Accounting, Organizations and Society 27.1-2 (2002): 99-120

Cho, Richard K. and Yigal Gerchak. "Supply chain coordination with downstream operating costs: Coordination and investment to improve downstream operating efficiency." European Journal of Operational Research In Press, Corrected Proof.

Claribel Castillo. "Cost Modeling System for the IC Packing Industry." Dissertation, Texas A&M University, 1998.

Cooper, Martha C., Lisa M. Ellram, John T. Gardner, and Albert M. Hanks (1997), "Meshing Multiple Alliances," Journal of Business Logistics, Vol. 18, No. l,pp. 67-89.

Cooper, Martha C., Douglas M. Lambert, and Janus D. Pagh (1997), "Supply Chain Management: More Than a New Name for Logistics," The International Journal of Logistics Management, Vol. 8, No. 1, pp. 1-14.

Dekker and Van Goor. "Supply Chain Management And Management Accounting: A case Study of Activity-Based Costing." International Journal of Logistics 3.1 (2000): 41-52.

Dekker, Henri C. "Value chain analysis in interfirm relationships: a field study." Management Accounting Research 14.1 (2003): 1-23.

Dekker, Henri C. "Control of inter-organizational relationships: evidence on appropriation concerns and coordination requirements." Accounting, Organizations and Society 29.1 (2004): 27-49

Duncan Williamson. "Introduction to Management Accounting.". Mar. 3, 2003. http://www.duncanwil.co.uk/manacc/trial/trial3.htm.

Escudero, L. F., et al. "Schumann, a modeling framework for supply chain management under uncertainty." European Journal of Operational Research 119.1 (1999): 14-34.

Fayez, Mohamed S. (2005) "An Automated Methodology for Comprehensive Definition of the Supply Chain Using Generic Ontological Components". Ph.D. Dissertation, University of Central Florida, Spring 2005.

Fishwick, Paul and Miller, John. " Ontologies for Modeling and Simulation: Issues and Approaches". Proceedings of the 2004 Winter Simulation Conference.

Gary Cokins. "Understanding Activity Based Costing and Supply Chain Management.". 1999. www.ascet.com.

Gunasekaran, A. "Supply chain management: Theory and applications." European Journal of Operational Research 159.2 (2004): 265-68.

Gunasekaran, A. and E. W. T. Ngai. "Information systems in supply chain integration and management." European Journal of Operational Research 159.2 (2004): 269-95.

Handfield, Robert B Nichols, and Jr Ernest L. Introduction to Supply Chain Management. Upper Saddle River, N.J. : Prentice Hal, 1999.

Jung Ung Min, Hans Bjornsson; e AEC research group, CIFE, Stanford (2002)

LaLonde and Pohlen. "Issues in Supply Chain Costing." International Journal of Logistics Management 7.1 (1996): 1-12.

Richard Cooper, Sandra Richtermayer, and Mark Frigo. 2003 Survey of Managment Accounting. 2003. Ernst and Young LLP and Institute of Managament Accountants. Report

Seal, Willie, et al. "Enacting a European supply chain: a case study on the role of management accounting." Management Accounting Research 10.3 (1999): 303-22.

Shapiro, J., “On the connections among activity-based costing, mathematical programming models for analyzing strategic decisions, and the resource-based view of the firm”, European Journal of Operational Research, 118, 1999, pp.295-314.

Shapiro, J., “Modeling the supply chain”, Duxbury, 2001.