ISSN: 2472-9256 (Print)
ISSN: 2472-9264 (Online)

Journal of Management Science and Business Intelligence

Volume 2, Number 2, August 2017

Published by Institution of Business Intelligence Innovation

Journal of Management Science and Business Intelligence

Vol. 2, No. 2; August, 2017

DOI https://doi.org/ 10.5281/zenodo.851581

Table of Contents

Editorial Board

IBII

Journal of Management Science and Business Intelligence, 2017, 2–2
Aug. 2017, pages 1-8
doi: 10.5281/zenodo.808452
http://www.ibii-us.org/Journals/JMSBI/
ISBN 2472-9264 (Online), 2472-9256 (Print)

CLUSTER ANALYSIS OF LIQUIDITY MEASURES IN A STOCK MARKET USING HIGH FREQUENCY DATA

Amin Salighehdar[1, *], Yang Liu[1] , Dragos Bozdog[1] and Ionut Florescu[1]

[1]Hanlon Financial System Laboratory, Financial Engineering Division, School of System and Enterprises, Stevens Institute of Technology, Hoboken, New Jersey, 07030

*Email: asalighe@stevens.edu

Received on February 28, 2017; revised on June 08, 2017; published on June 14, 2017

Abstract

Liquidity is one of the crucial factors in economy which reflects smooth operation of the markets. In a liquid market, traders are able to transact large quantities of security quickly with minimal trading cost and price impact. Many researchers have investigated the relationship between market liquidity and trading activity of a financial market. According to the existing literature, liquidity can measure different market characteristics such as trading time, tightness, depth, and resiliency. There is significant number of liquidity measures published in the literature. The main goal of this study is to use a hierarchical clustering algorithm to classify different liquidity measures. We examine the relationship between liquidity measures in order to detect commonality and idiosyncrasy among them. Then, we estimate the correlation among liquidity measures to quantify similarity between them and this quantity is used to develop a hierarchical clustering algorithm. At the end, we analyze the consistency in the structure of the clusters and we conclude that, clusters hold the same structure for almost 80% of the stocks in our sample. The data set that we are using for this study is NASDAQ High Frequency Trader (HFT) data. This data set contains trading and quoting activities of 26 HFT firms in 120 stocks on the Nasdaq exchange for various dates (in millisecond timestamp).

Keywords: Liquidity; High Frequency Trading; Correlation; Hierarchical Clustering.

1 Introduction

Liquidity is an essential feature of a financial market and it is used as an indicator for smooth operation of an economy. In a liquid market, traders are able to execute large quantities of security quickly with minimal trading cost and little price impact. A liquid market might be characterized as a continuous market that traders can buy or sell any amount of stock immediately (Black (1971), Kyle (1985)).

A significant amount of literature tries to shed light on the critical role of liquidity in finance. In the asset pricing literature, Acharya and Pedersen (2005) investigate how assets price is affected by liquidity risk. The authors show that the covariance of the security return and liquidity with the market return and market liquidity can affect a security's return.

Amihud and Mendelson (1986) investigate the relationship between stock bid and ask spread and stock return. They conclude that return increases with an increase in the spread. Amihud (2002) defines the daily ratio of absolute stock return to its dollar volume as an illiquidity measure, and it is shown that expected stock returns are an increasing function of expected illiquidity.

Chordia et al.(2001) study the relationship between liquidity and trading activity in order to analyze how liquidity and trading activity vary over different days in a week. The authors use daily data from 1988 to 1998, and they show that on Fridays we have lower liquidity and trading activity, while Tuesdays show opposite pattern. Bali et al. (2014) study how market reacts to liquidity shocks. The authors conclude that there is a positive and significant relation between liquidity shocks and future returns.

The existing literature shows that liquidity plays an essential role in the financial market worldwide. However, liquidity is not directly observable in the financial markets, and despite its importance, problems in quantifying liquidity still exist. Defining a globally accepted proxy for liquidity is an active area of research. O'Hara (2004) states that even though liquidity is a simple concept, it is hard to define it, and there exist a lot of different views of liquidity in literature. Furthermore, liquidity is a multidimensional variable that can capture different aspects of a market such as trading quantity, trading speed, trading cost and price impact (Liu (2006)). To the best of our knowledge, there are over 65 different measures of liquidity in literature developed using daily and monthly data. These measures may be classified into four categories based on the market feature they capture. The first category of the liquidity measures which includes liquidity measures such as number of trades, and turnover try to quantify *trading quantity* in the market. Amivest liquidity ratio, used by Cooper et al. (1985), Amihud illiquidity measure, Amihud (2002), Kyle lambda (λ), introduced by Kyle (1985), and Pastor reversal measure developed by Pastor and Stambaugh (2001) fall into a second category which captures *price impact*. The third category address the *trading cost* feature in the market.

For instance, difference between bid and ask price, known as spread, is one of the liquidity measures in this category. This group of liquidity measures has been studied in multiple papers (Roll (1984), Chordia et al. (2001), Acker et al. (2002), Dacorogna et al. (2001)). Finally, the last category of liquidity measures captures *trading speed*; e.g. Liu (2006) define a measure which falls into this category.

Although a lot of literature propose proxies to quantify liquidity in the market, some of the proposed measures suffer from serious shortcomings. For instance, among price impact measures, Amihud (2002) measures the lack of liquidity by dividing daily return over daily dollar volume. Although it is shown by Goyenko et al. (2009) that Amihud measure performs better than other measures, the measure cannot consider days without trading. Although there is no trading, these periods contain quotes which are important for a global liquidity measure. Amivest liquidity measure used by Cooper et al. (1985), and Amihud et al.(1997) suffer from the same issue. The quoted bid-ask spread is a noisy measure of illiquidity ((Lee (1993), Brennan and Subrahmanyam (1996)). Lee and Swaminathan (2000) show that trading volume, measured by the turnover ratio, is not a good indicator for liquidity. High trading volume does not necessarily indicate high liquidity in the financial market especially when we have significant volatility in the market, and this is proven during the *flash crash* on May 6, 2010 ((Van der Merwe (2015)).

Most of the measures existing in literature have been constructed using daily data (low-frequency). One of the reasons for using low-frequency based liquidity measures is saving in computational time compared to high-frequency based liquidity measures (Holden et al. (2014)). However, analyzing high-frequency based liquidity measures may provide more insight into time variation of liquidity. Therefore, more work is necessary to develop high frequency based liquidity measures.

Since limit orders play a crucial role in providing liquidity in the market, there are some works in the literature that propose liquidity measures based on information from a limit order book. For example, Cost of Round Trip (CRT), which is defined for any given transaction size, accumulates the status of the entire limit order book. This measure can capture the depth of a limit order book, and it is shown that CRT is correlated with other measures of liquidity such as quoted spread and effective spread (Irvine et al. (2000)). Kang and Zhang (2013) propose a liquidity measure which is capable of measuring the dispersion of a limit order book.

The main goal of this study is to classify the existing liquidity measures proposed in the literature into different groups. This classification have been done by considering the market feature that each liquidity measure captures. We are looking into the classification problem from different perspective. Correlation between individual stock liquidity and market liquidity, and co-movement between liquidity and other market factors such as return have been studied in literature (Von Wyss (2004), Sarr and Lybek (2002), Acharya and Pedersen (2005), Vu et al. (2015), and Amihud (2002)). In this study, we use the results obtained from the correlation analysis to develop a hierarchical clustering algorithm that is capable to detect a subset of liquidity measures that are similar to each other. We define the concept of the similarity of liquidity measures as a function based on Pearson correlation coefficient. Further, we analyze the consistency in the structure of the clusters. We are able to detect liquidity measures that introduce inconsistency. We call them *problematic liquidity measures*. We argue that our findings can provide more insight about the possible commonality and idiosyncratic feature between different liquidity measures. We perform all the analysis using a high frequency data set, so the first steps in our study is to investigate whether calculating the existing liquidity measures using a high frequency data is feasible.

This paper is organized as follows. In section 2, we provide a brief mathematical description for the liquidity measures used in this study. Also, we illustrate the correlation analysis and the hierarchical clustering algorithm used in this research. In section 3, we discuss the results including a discussion of the consistency in the structure of the clusters as well. We look into the consistency issue using an optimization concept. Finally, in section 4, we conclude the paper.

2 Methodology

In this section, we illustrate methodologies used in this study. First, we discuss the liquidity measures used in this study. Second, we explain the methodology used to analyze relation between different liquidity measures. Finally, we elaborate how liquidity measures can be classified into multiple groups based on the specific similarity function defined in this study.

2.1 Liquidity Measures

This section provides an overview of liquidity measures studied in literature. As we mentioned in section 1, there are more than 65 liquidity measures introduced in literature and most of them are developed using daily or monthly data. In this study, we are not proposing a new liquidity measure, but we are interested to analyze existing liquidity measures using high frequency data. We look into liquidity measures reviewed by Von Wyss (2004). Following same notation used by Von Wyss (2004), liquidity measures are divided into two groups: one-dimensional and multi-dimensional. The multi-dimensional liquidity measures are combination of multiple one-dimensional liquidity measures. In one-dimensional liquidity measures, only one market variable like price or size is used. Mathematical description of the liquidity measures used in this study is given in Table 6. We eliminate from the study all the liquidity measures that are not feasible to be calculated using our high frequency data set.

2.2 Correlation Analysis

In order to analyze the relation between different liquidity measures, we look into correlation among them. Correlation is a measure that quantifies the strength of linear relationship between two quantitative variables ((Moore and McCabe (1989)). Correlation between individual stock liquidity and market liquidity, and co-movement between liquidity and other market factors such as return have been studied in literature (Von Wyss (2004), Sarr and Lybek (2002), Acharya and Pedersen (2005), Vu et al. (2015), and Amihud (2002)). As we mentioned in the previous section, we are interested to use results from correlation analysis to classify liquidity measures into different groups. We believe that by this methodology, we can find a group of liquidity measures that provide same kind of information for the market. Therefore, we can reduce number of liquidity measures without loss of information. To achieve this goal, we build a correlation matrix for a set of liquidity measures.

Definition 1. *Let denote with* $X = \{x_i^j\}\ where\ (i, j) \in \{1, \dots, n\} \times \{1, \dots, m\}$ *a set of* $n \times m$ *liquidity measures.* m *and* n *denote respectively the number of days in our sample data set and the total number of liquidity measures used in this study. So, the correlation matrix* C^j *for any given day* j *is a* $n \times n$ *symmetric matrix whose* p, q *entry is the correlation between a pair of liquidity measures like* x_p *and* x_q *where* $p, q \in \{1, \dots, n\}$*, and it is defined as following:*

$$C^j{}_{pq} = \begin{cases} 1 & , if\ p = q \\ \frac{Cov(x_p,x_q)}{\sigma_{x_p}\sigma_{x_q}}, & if\ p \neq q \end{cases} \tag{1}$$

where $Cov(x_p, x_q)$ *is the covariance between two liquidity measures* (x_p, x_q) *and is defined as following:*

$$Cov(x_p, x_q) = E[(x_p - E[x_p])(x_q - E[x_q])] \tag{2}$$

and in Equation (1), σ_{x_p} *and* σ_{x_q} *represent the standard deviation for* x_p *and* x_q *respectively.*

By employing correlation analysis, we are able to detect a set of liquidity measures that have common movement. We use the term *commonality* to describe this pattern[1]. The reason for using superscript j in our notation is that we are interested to analyze the structure of the correlation matrix during our sample period. More details are presented in section 3.4.

2.3 Clustering Algorithm

In order to reduce sample space of the liquidity measures without losing any information, we use a clustering algorithm to partition liquidity measures into multiple groups. Clustering algorithms have been used in different research areas. The main purpose of these algorithms is to partition data points into some groups so that the points within each group are similar to each other, and the points from different groups are dissimilar (Tan et al. (2006)). Various clustering algorithms have been developed based on how a similarity between data points is measured (Ketchen and Shook (1996), Jain and Dubes (1988), Xu et al. (2005)). In addition to defining a similarity measure, most of these algorithms require a prior knowledge of the number of clusters that we want to divide the data points. For instance, in *k-means* algorithm, we need to define *k* which is the number of clusters. On the other hand, there exist some algorithms that do not require any information about the number of clusters in advance. For example, hierarchical clustering algorithms have this property. The result of a hierarchical clustering algorithm can be presented by a dendrogram (tree-like visual representation). Since, in this study we use a hierarchical clustering algorithm, we provide more details about this algorithm in the following section.

2.3.1 Hierarchical Clustering

First step in implementing a hierarchical clustering algorithm is to define a measure that can quantify similarity or dissimilarity among data points. In general, the scientific question that researchers try to address is a key factor in determining the dissimilarity measure (Gareth (2013)). In this study, we use Pearson correlation coefficient to define a dissimilarity measure.

Definition 2. *Let* ρ_{pq} *to be Pearson correlation coefficient between liquidity measures* x_p *and* x_q*. Mathematically*

$$\rho_{x_p x_q} = \frac{\sum_{k=1}^{l}(x_{pk}-E[x_p])(x_{qk}-E[x_q])}{\sqrt{\sum_{k=1}^{l}(x_{pk}-E[x_p])^2 \sum_{k=1}^{l}(x_{qk}-E[x_q])^2}} \tag{3}$$

Where l *is the total number of observation in one day. Then, we define a dissimilarity measure as following:*

$$D_{x_p x_q} = (1 - |\rho_{x_p x_q}|) \tag{4}$$

It is clear that the higher the correlation coefficient between two liquidity measures, the smaller the dissimilarity between them. Xu et al. (2005) review different similarity and dissimilarity measures used in clustering algorithms.

Agglomerative and divisive methods are two main methods to implement a hierarchical clustering algorithm. The agglomerative method is a bottom-up approach. It assigns each data point into a separate cluster. So, for N data points, we will have N clusters in the beginning. Then, it merges two similar clusters until a certain stopping criteria is achieved. The divisive approach is a top-down approach. It puts all the data points in a unique cluster. Then, it divides this cluster into two clusters. This process will continue until all clusters are singleton clusters (Xu et al. (2005)). Since, divisive method is computationally expensive (Everitt et al. (2001)), we use the agglomerative approach in this study. The dissimilarity measure defined in Equation 4, can quantify dissimilarity between two liquidity measures. In order to quantify dissimilarity between two sub clusters that each one has multiple data points, we use the complete linkage which considers the dissimilarity value between all pairs of observations in two clusters and keeps the largest value. More detail about different type of linkage methods can be found in Gareth (2013).

3 Results

In this section, we illustrate our results. First, we explain the data set used in this study. Second, we discuss the results obtained from the correlation analysis of the liquidity measures. Third, we present the results from the hierarchical clustering analysis. Finally, we provide a discussion about the consistency in the structure of the clusters.

3.1 Data Set

For this study, we use a high resolution data set provided by NASDAQ. This data set have been studied in different literature (Brogaard (2010), Kearns et al. (2010), and Khashanah et al. (2014)). These researchers use this data set to analyze the role of High Frequency Traders (HFTs) in United States equity market. This data set contains trading and quoting activity of 26 HFT firms in 120 stocks on the NASDAQ exchange. The sample is organized by market capitalization and is evenly split by NASDAQ to three different categories, large cap, medium cap and low cap. The sample period covers the week of Feb 22 - 26, 2010. Timestamp for trades is millisecond, and each trade has a flag to indicate whether it is initiated by a buyer or a seller. Also, trade reports contains a field with the following codes: HH, HN, NH, or NN. H represents a HFT and N denotes a non-HFT. The first term in the pair classifies the liquidity seeking side, and the second character classifies liquidity supplier. For example, a trade labeled HN would mean an HFT took liquidity from a non-HFT. In addition to trade information, for the week of Feb 22 - 26, 2010, the data set contains the collective best HFT quote along with the collective best non-HFT quote.

3.2 Correlation Analysis Result

In this section, we present the results from the correlation analysis. First, we calculate 22 different liquidity measures within one second time interval for all 120 stocks in our data set. We repeat this process for each

[1] The commonality and idiosyncratic patters are two terms used in literature to analyze cross-sectional variation in liquidity. See Chordia et al. (2000), Mancini et al. (2013),Karolyi et al. (2012)

day. Then, we calculate the correlation matrix for that specific day. We set the correlation coefficients to zero if they are not statistically significant (Casella and Berger (2002)). One of these matrices is shown in Figure 1. In this figure, the blue color circles display positive correlation, and the red color circles shows negative correlation. Color intensity is proportional to the correlation coefficient. As we were expecting and documented in literature (Von Wyss (2004)), the economically similar liquidity measures are highly correlated to each other. For instance, high correlation between *Sabs*, *SrelM*, *Srelp* is shown in Figure 1.

Further, by paying close attention to the structure of the correlation matrix, we can see that the correlation matrix can be decomposed into multiple blocks. This special structure of the correlation matrix is quiet interesting, since it brings up the idea that there are some liquidity measures that are highly correlated to each other. Therefore, we can reduce the number of liquidity measures without losing any useful information. So, this analysis reveals the presence of commonality pattern among liquidity measures.

We do further analysis to examine whether this structure can be verified for all the stocks in our sample period. Figure 2 shows the correlation matrix for three stocks during five days. These stocks, *AMZN*, *SFG*, and *ROG* are chosen from high cap, medium cap, and low cap categories respectively. We can clearly see that, the correlation matrices have almost same structure during sample period. This result shows that there is a consistency in the structure of the correlation matrix. Block structure of the correlation matrix proves that there is a high linear relationship between some of these liquidity measures. This finding is used to classify liquidity measures into different groups.

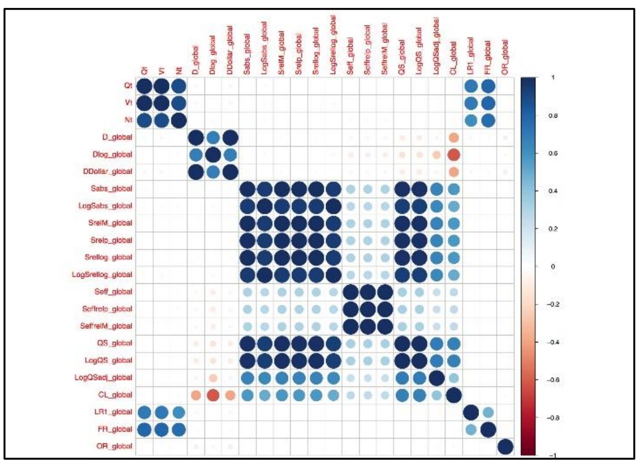

Fig 1: Correlation Matrix of Liquidity Measures for AAPL during Feb 22, 2010

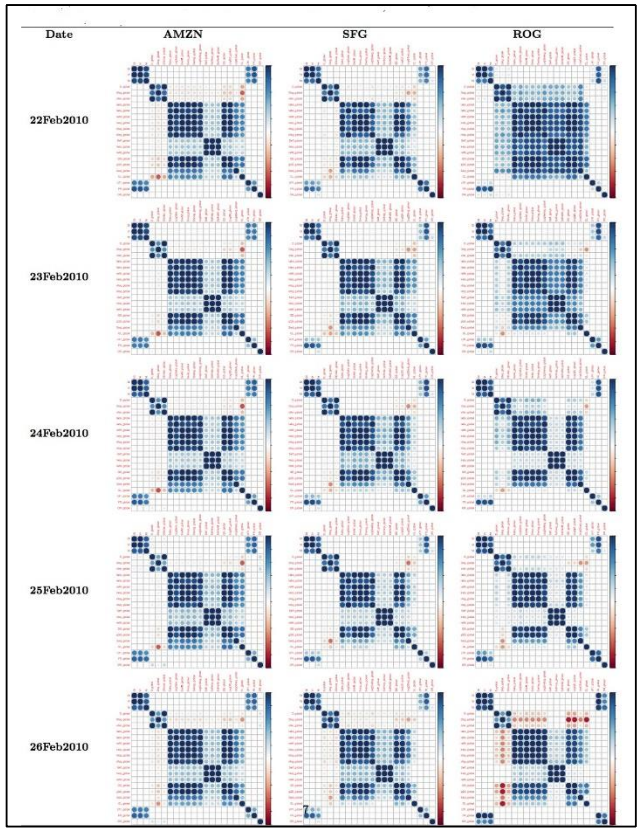

Figure 2: Comparing the structure of the correlation matrix for AMZ, SFG and ROG during a week from 22 Feb, 2010 - 26 Feb, 2010

3.3 Hierarchical Clustering Results

The main contribution of this study is to classify liquidity measures into different groups. This can help us to identify liquidity measures that are providing same kind of information to the market. It means that we can reduce the number of liquidity measures without losing any useful information. To achieve this goal, we employ a hierarchical clustering algorithm by using Pearson correlation coefficient to define our dissimilarity measure. The result of a hierarchical clustering can be presented by a dendrogram. For example, Figure 3 is obtained by employing hierarchical clustering on all 22 liquidity measures calculated in this study. By cutting the dendrogram at specific level (vertical axis), we will have different number of clusters. Table 1 summarizes the number of the clusters and the members of each cluster obtained from setting cutting level at 0.4. In this case, we are sure that the absolute value of the correlation coefficient among the liquidity measures is greater than or equal to 0.6. By changing the cutting level value, the structure of the clusters will be different. Figure 4 and Table 2 shows the dendrogram and the cluster structure obtained from choosing 0.2 as a cutting level. The presented results is just for one stock and it is considering only one day in our sample data set.

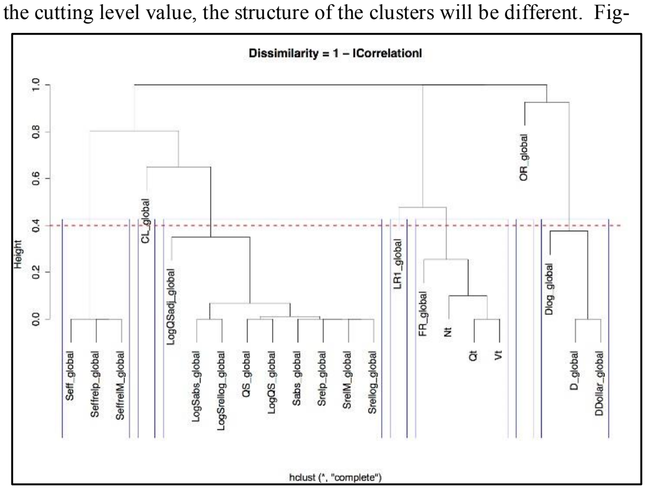

Fig 3: Hierarchical Cluster Analysis for AAPL during 26Feb, 2010, cutting level=0.4

Table 1: Cluster with their corresponding members . Ticker is AAPL, date is Feb 26, 2010. Cutting level of dendrogram is 0.4

Clusters	Liquidity Measures
1	Q_t, V_t, N_t, FR
2	$D, Dlog, D\$$
3	$Sabs, LogSabs, SrelM, Srelp, Srellog$ $LogSrellog, QS, LoggQs, LogQSadj$
4	$Seff, Seffrelp, SeffrelM$
5	CL
6	$LR1$
7	OR

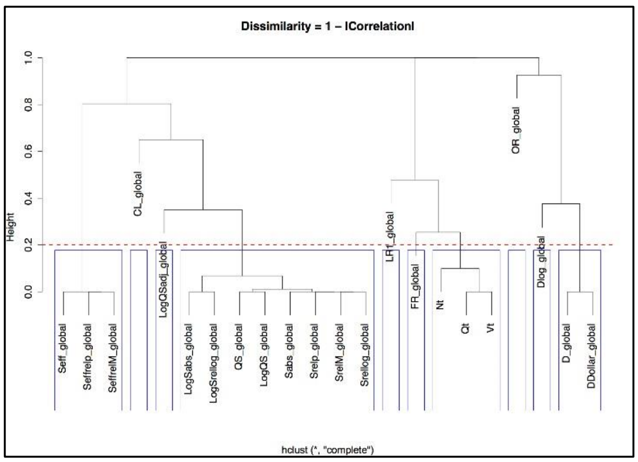

Table 2: Clusters with their corresponding members for AAPL during 26Feb, 2010, cutting level=0.2

Fig 4: Hierarchical Cluster Analysis for AAPL during 26Feb, 2010, cutting level=0.2

Clusters	Liquidity Measures
1	Q, V, N
2	$D, D\$$
3	$Dlog$
4	$Sabs, LogSabs, SrelM, Srelp$ $Srellog, LogSrellog, QS, LogQS$
5	$Seff, Seffrelp, SeffrelM$
6	$LogQSadj$
7	CL
8	$LR1$
9	FR
10	OR

As we mentioned in the above, the results we presented for the hierarchical clustering is only for one stock and one day. We need to investigate whether this structure is the same for all the stocks in our sample data set during all days. To do this, we repeat the process, and we count number of stocks that have same cluster structure for whole sample period. We find that this structure is not consistent. There are some liquidity measures that make inconsistency in the structure of the clusters. After doing analysis, we identified those liquidity measures to be *number of transactions per time unit* (N_t), *logDepth* ($DLog$), *composite liquidity* (CL_t), *liquidity ratio1* ($LR1$) and *adjusted log quote slope* ($LogQSadj$). After removing these 5 problematic liquidity measures , we get significant improvement in the results. Table 3 shows the cluster structure obtained after removing the problematic liquidity measures. By removing the problematic liquidity measures, the consistency in the structure of the clusters is significantly improved. Table 4 shows the results for two cases. It is clear that 79.2% of the stocks in our date set keep having same cluster structure for all days in our sample period.

Table 3: Cluster structure after removing problematic liquidity measures. For AAPL during 26 Feb, 2010, cutting level=0.4

Clusters	Liquidity Measures
1	Q_t, V_t, FR
2	$D, D\$$
3	$Sabs, LogSabs, SrelM, Srelp, Srellog$ $LogSrellog, QS, LogQS$
4	$Seff, Seffrelp, SeffrelM$
5	OR

Table 4: Comparing consistency in the structure of cluster before and after removing problematic liquidity measures.

Stock Category	Considering all the liquidity measures	Removing the problematic liquidity measures
All stocks	2.5%	79.2%
Large cap	4.8%	76.2%
Medium Cap	0%	77.5%
Small Cap	2.6%	84.2%

3.4 How to choose a cutting level for the dendrogram?

In all the analysis and the results we discussed so far, we assume a fixed cutting level. Specifically, we set this value to be 0.4. In this section we provide a discussion about how we get this value. Generally, this value depends on the nature of the problem and the goal we want to obtain through a hierarchical clustering. In this research, we are using the following procedure. Let α to be the value of dissimilarity function which in our case is defined as following:

$$\alpha = 1 - |\rho| \tag{5}$$

where ρ is the Pearson correlation coefficient. The vertical axis in a dendrogram obtained from a hierarchical clustering (Figure 3) shows the range of α. Since $|\rho| \leq 1 \Rightarrow 0 \leq \alpha \leq 1$.

- If $\alpha = 1$ ($|\rho| = 0$), then we have one big cluster.

- If $\alpha = 0$ ($|\rho|$ =1), then we have m clusters where m is the total number of liquidity measures.

So, let k to be the number of clusters we obtain by cutting the dendrogram at level α_i. \label{algo}The possible values of k are $1,2,\dots,m,$ where m is the total number of liquidity measures. For any given stock s_i (in our data set), we cut the dendrogram at level α_i. Then, we calculate k and record members of ecah cluster as well. We repeat this process for all days (in our sample period). We say stock $s_i$has a consistent cluster structure if the number of clusters and the members of each cluster are the same for all days. We repeat this procedure for all the stocks, and count how many stocks have these properties. We need to do this procedure for different value of α_i. It is clear that in two cases we can have the best consistency. First case happens when $\alpha = 0$ which means $|\rho|$ =1 that implies $k = m$. Second case happens when $\alpha = 1$ which means that $|\rho|$ =0 that implies

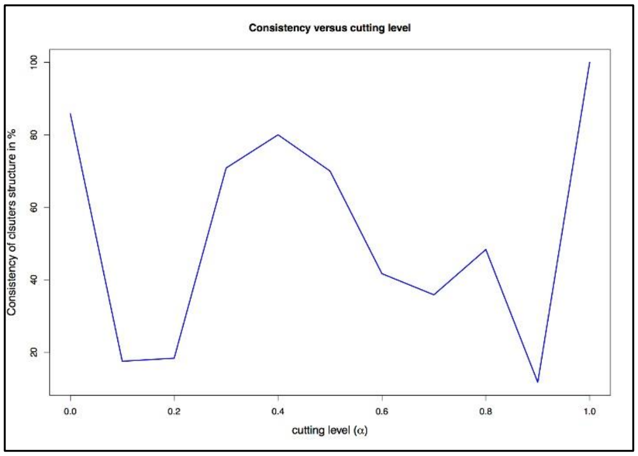

Fig 5: Consistency versus cutting level

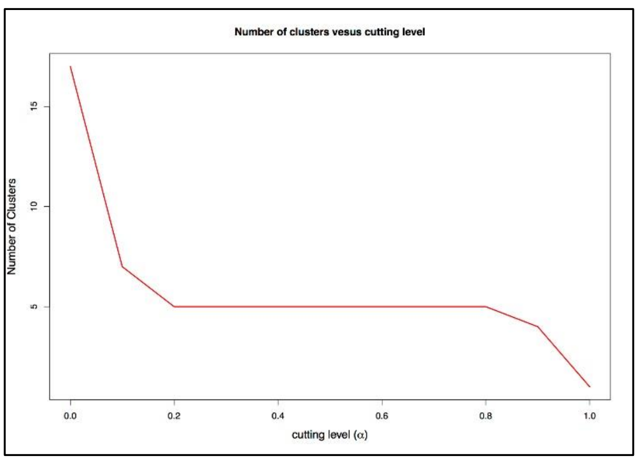

Fig 6: Number of clusters versus cutting level

$k = 1$. In other words, we can say that k is a decreasing function of α. Also, we need to take into account that the whole purpose of clustering is to partition similar liquidity measures into one groups. Therefore, we want to achieve the following while decreasing the value of dissimilarity function (α_i).

1. Increase consistency.
2. Decrease number of clusters.

We need to solve this optimization problem. We can look at the plot of the consistency versus α and the number of clusters versus α . Then, we can identify the region of interest (Figure 5, and Figure 6).

Numerically, by changing the value of α we record the consistency and the the total number of the clusters (Table 5). It is clear that number of clusters is a decreasing function of cutting level. And, For $\alpha = 0$ and $\alpha = 1$, we obtained highest consistency. We can see that by fixing the cutting level at 0.4, we can obtain the highest consistency in the structure of the clusters.

Table 5: Analyzing how cutting dendrogram obtained from hierarchical clustering at different levels will affect the consistency in the structure of the clusters

Cutting Level (α)	Consistency (Percentage)	Number of Clusters
0.00	85.83	17
0.10	17.50	7, 6
0.20	18.33	5,6
0.30	70.83	4,5
0.40	80.00	4,5
0.50	70.00	4,5
0.60	41.67	4,5
0.70	35.83	4,5
0.80	48.33	4,5
0.90	11.67	4
1.00	100.00	1

4 Conclusion

There exist a lot of liquidity measures in literature. These measures have been classified into different groups based on the market feature they capture. In this study we look into the classification of the liquidity measures from different perspective.

We perform a cluster analysis of the existing liquidity measures in the literature. Using a high frequency data assists us to investigate the intraday behavior of the liquidity measures more precisely. By analyzing the correlation matrix of the liquidity measures, we find that there is a block structure in the correlation matrix and further investigation reveals that this structure is consistent. We define a dissimilarity measure using a correlation concept. By using a hierarchical clustering algorithm, we are able to partition liquidity measures into different groups. We conclude that, there is a consistency in the the structure of the dendrograms obtained from the hierarchical clustering. This results obtained from the clustering can help us to reduce the number of the liquidity measures without losing any information. For future work, we are interested to use the results from the clustering procedure to develop a (ideally one) new liquidity index which is capable of capturing most of the market features.

Acknowledgements

The authors acknowledge the NASDAQ OMX Group for graciously providing us with a sample of 120 stocks with various levels of market capitalization listed on NYSE and NASDAQ where HFT traders are globally labeled under designation H.

Appendix

Mathematical description of the liquidity measures used in this study is presented in the following table. In this table, q_i and p_i denotes the number of shares of trade i and its price respectively. q_t^A and q_t^B refer to best ask volume and best bid volume respectively. Accordingly, p_t^A and p_t^B represent best ask price and best bid price respectively. p_t denotes the last paid price of the asset before time$t$.. r_t represents the return during the time interval for which we calculate liquidity proxies.

Table 6: Mathematical description of liquidity measures used in this study. The number of measures reviewed by Von Wyss (2014) is 31. We only present 22 of them which can be calculated using the high frequency data set that we analyzed for this study.

One Dimensional Liquidity Measures	
Liquidity Measures	Formula
Trading Volume	$Q_t = \sum_{i=1}^{N_t} q_i$
Turnover	$V_t = \sum_{i=1}^{N_t} p_i\, q_i$
Depth	$D_t = q_t^A + q_t^B$
LogDepth	$Dlog_t = \ln(q_t^A) + \ln(q_t^B)$
Dollar Depth	$D\$_t = \frac{q_t^A . p_t^A + q_t^B . p_t^B}{2}$
Number of Transaction per time unit	N_t
Absolute Spread	$Sabs_t = p_t^A - p_t^B$
Log Absolute Spread	$LogSabs_t = \ln(Sabs_t)$
Relative spread calculated with mid-price	$SrelM_t = \frac{p_t^A - p_t^B}{p_t^M} = \frac{2.\,(p_t^A - p_t^B)}{p_t^A + p_t^B}$
Relative spread calculated with last trade price	$Srelp_t = \frac{p_t^A - p_t^B}{p_t}$
Relative spread of log prices	$Srellog_t = \ln(p_t^A) - \ln(p_t^B)$
Log relative spread of log prices	$LogSrellog_t = \ln(Srellog_t)$
Effective Spread	$Seff_t = \lvert p_t - p_t^M \rvert$
Relative effective spread calculated with last trade price	$Seffrelp_t = \frac{\lvert p_t - p_t^M \rvert}{p_t}$
Relative effective spread calculated with mid-price	$SeffrelM_t = \frac{\lvert p_t - p_t^M \rvert}{p_t^M}$
Multi-Dimensional Liquidity Measures	
Quote Slope	$QS_t = \frac{Sabs_t}{Dlog_t}$
Log Quote Slope	$LogQS_t = \frac{Srellog_t}{Dlog_t}$
Adjusted Log Quote Slope	$LogQSadj_t = LogQS_t.\,(1 + \lvert \ln(\frac{q_t^A}{q_t^B}) \rvert)$
Composite Liquidity	$CL_t = \frac{SrelM_t}{D\$_t}$
Liquidity Ratio 1	$LR1_t = \frac{V_t}{\lvert r_t \rvert}$
Flow Ratio	$FR_t = N_t . V_t$
Order Ratio	$OR_t = \frac{\lvert q_t^B - q_t^A \rvert}{V_t}$

Conflict of Interest: none declared.

References

Acharya, V. V. and L. H. Pedersen (2005). Asset pricing with liquidity risk. *Journal of Financial Economics 77 (2)*, 375-410.

Acker, D., M. Stalker, and I. Tonks (2002). Daily closing inside spreads and trading volumes around earnings announcements. *Journal of Business Finance & Accounting 29 (9-10)*, 1149-1179.

Amihud, Y. (2002). Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Markets 5 (1)*, 31-56.

Amihud, Y. and H. Mendelson (1986). Asset pricing and the bid-ask spread. *Journal of Financial Economics 17 (2)*, 223-249.

Amihud, Y., H. Mendelson, and B. Lauterbach (1997). Market microstructure and securities values: Evidence from the tel aviv stock exchange. *Journal of Financial Economics 45 (3)*, 365-390.

Bali, T. G., L. Peng, Y. Shen, and Y. Tang (2014). Liquidity shocks and stock market reactions. *Review of Financial Studies 27 (5)*, 1434-1485.

Black, F. (1971). Toward a fully automated stock exchange, part i. *Financial Analysts Journal 27 (4)*, 28-35.

Brennan, M. J. and A. Subrahmanyam (1996). Market microstructure and asset pricing: On the compensation for illiquidity in stock returns. *Journal of financial economics 41 (3)*, 441-464.

Brogaard, J. (2010). High frequency trading and its impact on market quality. *Northwestern University Kellogg School of Management Working Paper 66*.

Casella, G. and R. L. Berger (2002). *Statistical inference*, Volume 2. Duxbury Pacific Grove, CA.

Chordia, T., R. Roll, and A. Subrahmanyam (2000). Commonality in liquidity. *Journal of Financial Economics 56 (1)*, 3-28.

Chordia, T., R. Roll, and A. Subrahmanyam (2001). Market liquidity and trading activity. *The Journal of Finance 56 (2)*, 501-530.

Cooper, S. K., J. C. Groth, and W. E. Avera (1985). Liquidity, exchange listing, and common stock performance. *Journal of Economics and Business 37 (1)*, 19-33.

Dacorogna, M., R. Gencay, U. Muller, R. Olsen, and O. Pictet (2001). An introduction to high frequency finance academic press. *San Diego*.

Everitt, B., S. Landau, and M. Leese (2001). Cluster analysis arnold. *A member of the Hodder Headline Group, London*.

Gareth, J. (2013). *An Introduction to Statistical Learning: with Applications in R*. Springer.

Goyenko, R. Y., C. W. Holden, and C. A. Trzcinka (2009). Do liquidity measures measure liquidity? *Journal of Financial Economics 92 (2)*, 153-181.

Holden, C. W., S. Jacobsen, A. Subrahmanyam, et al. (2014). The empirical analysis of liquidity. *Foundations and Trends in Finance* 8 (4), 263-365.

Irvine, P. J., G. J. Benston, and E. Kandel (2000). Liquidity beyond the inside spread: Measuring and using information in the limit order book. *Available at SSRN 229959*.

Jain, A. K. and R. C. Dubes (1988). *Algorithms for clustering data*. Prentice-Hall, Inc.

Kang, W. and H. Zhang (2013). Limit order book and commonality in liquidity. *Financial Review 48 (1)*, 97-122.

Karolyi, G. A., K.-H. Lee, and M. A. Van Dijk (2012). Understanding commonality in liquidity around the world. *Journal of Financial Economics 105 (1)*, 82-112.

Kearns, M., A. Kulesza, and Y. Nevmyvaka (2010). Empirical limitations on high frequency trading portability. *Available at SSRN 1678758*.

Ketchen, D. J. and C. L. Shook (1996). The application of cluster analysis in strategic management research: an analysis and critique. *Strategic management journal 17 (6)*, 441-458.

Khashanah, K., I. Florescu, and S. Yang (2014). On the impact and future of hft.

Kyle, A. S. (1985). Continuous auctions and insider trading. Econometrica: *Journal of the Econometric Society*, 1315-1335.

Lee, C. (1993). Market integration and price execution for nyse-listed securities. *The Journal of Finance 48 (3)*, 1009-1038.

Lee, C. and B. Swaminathan (2000). Price momentum and trading volume. *The Journal of Finance 55 (5)*, 2017-2069.

Liu, W. (2006). A liquidity-augmented capital asset pricing model. *Journal of financial Economics 82 (3)*, 631-671.

Mancini, L., A. Ranaldo, and J.Wrampelmeyer (2013). Liquidity in the foreign exchange market: Measurement, commonality, and risk premiums. *The Journal of Finance 68 (5)*, 1805-1841.

Moore, D. S. and G. P. McCabe (1989). *Introduction to the Practice of Statistics*. WH Freeman/Times Books/Henry Holt & Co.

O'Hara, M. (2004). Liquidity and financial market stability. *National Bank of Belgium Working Paper (55)*.

Pastor, L. and R. F. Stambaugh (2001). Liquidity risk and expected stock returns. Technical report, National Bureau of Economic Research.

Roll,R.(1984).A simple implicit measure of the effective bid-ask spread in an efficient market. *The Journal of Finance 39 (4)*, 1127-1139.

Sarr, A. and T. Lybek (2002). Measuring liquidity financial markets.

Tan, P.-N. et al. (2006*). Introduction to data mining*. Pearson Education India.

Van der Merwe, A. (2015). *Market Liquidity Risk: Implications for Asset Pricing, Risk Management, and Financial Regulation*. Palgrave Macmillan.

Von Wyss, R. (2004). *Measuring and predicting liquidity in the stock market*. PhD dissertation, University of St. Gallen.

Vu, V., D. Chai, and V. Do (2015). Empirical tests on the liquidity-adjusted capital asset pricing model. *Pacific-Basin Finance Journal 35*, 73-89.

Xu, R., D. Wunsch, et al. (2005). Survey of clustering algorithms. *Neural Networks, IEEE Transactions on 16 (3)*, 645-678.

Journal of Management Science and Business Intelligence, 2017, 2–2
Aug. 2017, pages 9-16
doi: 10.5281/zenodo.827384
http://www.ibii-us.org/Journals/JMSBI/
ISBN 2472-9264 (Online), 2472-9256 (Print)

Measuring a leader's ability to identify and avert crisis

Jamie Brownlee-Turgeon, Ph.D.

School of Business and Professional Studies, Brandman University, 16355 Laguna Canyon Road, Irvine, CA 92618

*Email: jbrownle@brandman.edu

Received on June 07, 2017; revised on July 10, 2017; published on July 14, 2017

Abstract

Leaders often have influence over the impact of pending crises by either preventing or minimizing the crisis (Pearson and Mitroff, 1993; Bonvillian, 2013). With crisis looming just around the corner, a leader's ability to identify, avert, and manage a crisis has become a fundamental element in organizational sustainability. Yet, most literature on crisis is focused in the field of communication or crisis management during the actual event. Wooten and James (2008) provide a conceptual model that describes leadership competencies in each of the five stages of crisis management. The development of the crisis identification and aversion instrument was to operationalize the Wooten and James (2008) conceptual model with a focus only on the pre-crisis stages of crisis management competencies. The crisis identification and aversion instrument has been validated through item reduction and content validation through the use of a Delphi panel of experts, item evaluation through the use of a large sample and factor analysis and assessment of construct validity. The validated instrument measures a leader's ability to identify and avert crisis by measuring three competencies: sensemaking, participatory management, and resourcefulness. Each scale has high internal consistency.

Keywords: Crisis leadership, crisis aversion, sense making, participatory management, resourcefulness, quantitative

1 Introduction

Crises continue to be an ever-present factor to our reality, in the past, the present, and the future (DuBrin, 2013). In fact, Fink (1986) posits that organizations should expect to always have a pending crisis right around the corner. Although there are numerous definitions of a crisis, this article utilizes Pearson and Clair's (1998) definition, "An organizational crisis is a low-probability, high-impact event that threatens the viability of the organization and is characterized by ambiguity of case, effect, and means of resolution, as well as by a belief that decisions must be made swiftly" (p. 60). In other words, any event that disrupts normal business practice describes a crisis. Fink (1986) suggests that a crisis has multiple stages with the event being only one stage. Many identify with a crisis event but have minimal knowledge to the other stages within crisis management.

Following the 1982 Tylenol poisoning and recall crisis management emerged as a field of study during. From that point, crisis management was studied by the field of communication with limited empirical research (Mitroff, 2004; Pearson & Mitroff, 1993). Historically, crisis management has focused solely on managing a crisis so business can continue as usual. A cost analysis approach is used to determine what action, if any, will be taken (Mitroff, 2004). Furthermore, the Institute of Crisis Management estimates that 50% of crises occur due to action or inaction by leadership. These data suggest a lack of focus on a leaders' ability to identify and avert crisis not only because of different leadership competencies but also due to traditional models of thinking such as risk management.

Crisis leadership is defined with two goals: (a) crisis aversion and (b) if aversion is not an option, mitigate the crisis in such a way that the organization becomes more resilient than before the crisis (Mitroff, 2004; Wooten & James, 2008). Bonvillian (2013) suggests that leaders with the ability to identify and avert crisis have a propensity to utilize crisis as a strategic catalyst to move the organization forward. Furthermore, these competencies lack appropriate attention in the academy and in leadership education and training; yet, Bonvillian (2013) postulates that the competencies needed for precrisis stages differ from what is needed during normal business practice. In a qualitative study based on archival data from the Institute of Crisis Management between the years of 2000-2006, Wooten and James (2008) developed a conceptual model of competencies needed by leadership in order to lead well through the five stages of crisis management. The five stages of crisis management are (a) signal detection; (b) preparation and prevention; (c) damage control and containment; (d) business recovery; and (e) learning and reflection (Mitroff, 2004). This article focuses on the first two stages of the Wooten and James (2008) model which are aimed at crisis aversion, more specifically with organizational generated or human-induced crises. These stages include signal detection and preparation and prevention.

2 Theoretical Foundation

Other constructs have a nomological interdependency with crisis leadership thus an understanding of them is relevant as they are distinct constructs. These constructs include crisis management, environmental scanning, risk management and emergency management.

Crisis management identifies probable crises and develops plans of action to prevent and mitigate the crisis event. However, it lacks the consideration of linking events together that may be predictions of a looming and low-probability crisis (Mitroff, 2004).

Environmental scanning includes the following areas: industry or market, regulatory, economic, social, and political (Albright, 2004). With a focus on identifying potential threats, it aligns closely with crisis leadership as crisis leaders continually scan the environment for a pending threat (James & Wooten, 2005). Yet again, the focus is more on the probable than the linkage of the improbable.

Risk management is the traditional approach that organizations utilize to assess potential liability. That said, risk management is more about the cost of the crisis occurring verses the cost of attempting to prevent the crisis (Williams, Bertsch, Dale, Smith & Visser, 2006). Furthermore, there are aspects of risk management that contribute to crisis leadership due to its focus on an organization's vulnerabilities and costs. If the cost to recover from the crisis is less than the cost of aversion, then risk management advocates to let the crisis occur.

Emergency management, differing from the other constructs, focuses on the low-probability events and develops a plan of action to prevent and mitigate (Waugh & Tierney, 2007). However, emergency management lacks the authority to decide to avert a crisis or even how to redesign after a crisis in order to create greater resiliency. Emergency management is merely tactical.

Crisis leaders do not just follow plans or limit themselves to probable events or a narrow perspective. Crisis leaders see the big picture, have an ability to link improbable events together in order to interpret a potential crisis, continuously engage in pre-crisis audits to identify warning signs and have an ability to redesign an organization toward greater resiliency following a crisis (Mitroff, 2004).

Currently, there is one other quantitative tool that measures crisis leadership. The Crisis Leader Efficacy in Assessing and Deciding (C-LEAD; Noonan Hadley, Pittinsky, Sommer, & Zhu, 2011) scale is an existing quantitative instrument that assesses a leader's ability in the third stage of crisis management, damage control and containment. However, there is no empirical data to support what effective leadership looks like in terms of competencies in the pre-crisis stages. The research focused on the pre-crisis stages and developed and validated a crisis identification and aversion tool to assess a leader's ability to avert crisis.

3 Methods

DeVellis (2012) suggests an eight steps process to develop and validate a new scale. These steps include (a) determine clearly what it is you want to measure based on a theoretical foundation, (b) generate an item pool, (c) determine the format for measurement, (d) have the initial item pool reviewed by experts, (e) consider inclusion of validation items, (f) administer items to a large sample, (g) evaluate the items, and (h) optimize the scale length.

First, a review of the literature was conducting utilizing the five competencies found in the first two stages of the Wooten and James (2008) model as well as the constructs with nomological interdependencies to crisis leadership. Four out of the five competencies had existing validated scales; as such, these scales were used as the foundation for the item pool. The validated instruments utilized include the following: (a) perspective taking, 7 items (M.H. Davis, 1980); (b) issue selling, 9 items (Bishop, Webber, & O'Neil, 2011); (c) organizational agility, 25 items (Charbonnier;Voirin, 2011); and (d) creativity, 30 items (Gough, 1979). The fifth competency, sensemaking, required a theoretically founded proposed set of 30 items. The original pool included 115 items. An initial evaluation of the items for duplicates or combining to strengthen an item reduced the list from 115 items to 97.

Next, 29 participants were identified for the Delphi panel in which 13 participated in the two iterations. Their expertise areas included higher education faculty in the field of organizational leadership, crisis management practitioners, and senior-level management who have encountered crisis in their tenure. The Delphi panel refined and reduced the item pool by identifying the level of importance of each item toward the construct on a 5-point Likert scale (1 = Not at all important, 2 = Minimally important, 3 = Somewhat important, 4 = Moderately important, and 5 = Very important). Items with a score of 4 or higher and had 80% of the experts' rating over 4 remained in the item pool. Additionally, there was opportunity for the experts to provide feedback on which items were duplicates or needed more clarity. The first iteration reduced the item pool from 97 to 54 items. The second iteration reduced the item pool from 54 to 41 items.

Following the Delphi panel two-stages of refinement and reduction, a large sample was used to evaluate the scale. This study utilized snowball sampling. The minimum sample size needed was 205 participants based on a rate of 5-10 respondents per item with 41 items remaining (DeVellis, 2012; Nunnely, 1978). Originally, there were 389 responses with 111 missing over 50% of the questionnaire; thus, the final sample size was 278. Of the 278, mean substitution was utilized as the imputation approach for any missing items; however, these included no more than two per respondent.

The first section of the survey included demographic questions about the respondent. These demographics included gender, industry, years of employment at current organization, and years of work employment experience. There were also questions to describe the relationship between participant and the leader identified for evaluation. The data collected included position of the leader identified, positon of the respondent in comparison to the leader identified, years worked with or for the identified leader, and lastly, currently working for the identified leader.

The following section included the 41 remaining items to describe the five competencies. These were rated on 7-point Likert scale and utilized "describes him/her very accurately" to "describes him/her very inaccurately" as anchors. Respondents were asked to identify a leader they had worked with or for over the last five years and to answer the questions based on how accurately or inaccurately the statement described that leader.

The final section included three scales for validation of the crisis aversion measurement: (a) Discriminant Validity: C-LEAD, 9 items (Noonan Hadley et al., 2011); (b) Predictive Validity: General Risk Propensity in Multifaceted Business Decisions, 5 items (Hung & Tangpong, 2010); and Predictive Validity: Leadership Effectiveness Scale, 6 items (Ehrhart & Klein, 2001).

Analysis of the large scale began with the Kaiser-Meyer-Olin measure of sampling adequacy and Bartlett's test of sphericity in order to determine if factor analysis was the appropriate method of evaluation. Once determined that it was the appropriate method, the second step utilized principle component factor analysis in SPSS. Direct oblimin was used for factor rotation and interpretation based on the strength of the correlation of items. The next step evaluated the eigenvalue, the scree plot, and the communalities. Lastly, factor analysis was used for further item reduction and factor loadings. Factor analysis was run two times. Any item that was cross-loaded or below a .35 significance after each iteration was removed. There were three remaining factors determined by loading on separate factors and all three had high internal consistency with a Cronbach alpha over .90. Once the three factors were determined, each factor was tested with the three validity scales. Predictive validity was measured based on the correlation between the factors and the Leadership Effectiveness Scale (Ehrhart & Klein, 2001) and the General Risk Propensity in Multifaceted Business Decisions (Hung & Tangpong, 2010). Discriminant validity utilized factor analysis with the C-LEAD (Noonan Hadley et al., 2011) and the three factors to determine if they loaded separately.

4 Results

The purpose of this section is to provide a summary of the data analysis conducted at each stage of the data collection process. There are three stages: item development, scale evaluation, and scale validation.

3.1 Item Evaluation

After the initial reduction of the item pool based on duplicates and combining of similar themes, a Delphi panel was the next step. Of the 29 invited participants, 13 responded to two iterations of the item pool. The experts were asked to rate the importance of the item to describe the variable. A 5-point Likert scale was used with "not at all important" and "very important" as anchors. Items retained in the instrument were based on items that averaged a 4.0 ranking or higher as well as at least 80% of the participants ranking the items over a 4.0. The first iteration reduced the 97-item pool to 54 items. The second iteration reduced the 54 items to 41 items.

3.2 Scale Evaluation

In the evaluation of the scale, demographics were collected about the respondents along with the leader they evaluated in the survey. Next, tests were conducted to determine whether or not factor analysis was the appropriate method; then, factor analysis was run two times.

3.2.1 Demographics

The demographic data casts a picture of the respondents. In terms of industry, 43.2% work in education followed by healthcare (9.4%), government (6.8%) and all other industries under 6%. The split between genders was male = 42.6% and female = 57.2%. 86.3% have over 10 years work experience.

Respondents were asked to consider a leader that they observed over the past 5 years and respond based on how the items described the leader. 42.1% evaluated their current supervisor with the second highest rating conducted on senior leadership (35.3%). The position of the respondent ranked highest with 56.1% being immediate subordinates.

Table 1. Data Corresponding to the Leader Identified and Respondent (N=278)

Variable	Percentage
Position of the Leader Identified	
Immediate Supervisor	42.1
Department Head (one level above supervisor	14.4
Senior Management	35.3
Peer Leader	7.9
Position of Respondent	
Immediate Subordinate	56.1
Member of Department	21.2
Member of Organization	12.6
Peer Leader	9.7
Years worked with or for the identified leader	
1.2 years	36.3
3-4 years	29.5
5+ years	33.5
Currently working for this leader	
Yes	50.7
No	49.3

3.2.2 Scale Identification and Validation

Factor analysis was utilized for scale identification. The data were evaluated on the number of missing questions per respondent. Initially, there were 389 respondents with 111 responses missing over 50% of the questionnaire. Hair, Black, Babin, and Anderson (2010) recommend the removal of respondents with over 50% missing data. With the removal of 111 respondents, N=278 remained statistically significant. The 278 respondents had no more than 2 missing questions totaling 34 missing items in the remaining 278. With 34 equating to .003% of all items, it was determined to utilize mean substitution as the imputation approach.

Factor analysis requires additional testing to determine the factorability of the data. The Kaiser-Meyer-Olkin (KMO), which measures the sampling adequacy, was .972. The Bartlett's test of sphericity had a significance of .000. Lastly, the communalities were assessed and all items had communalities greater than .50. The data were supportive of the use of factor analysis.

The correlation matrix supported that most of the items have a medium to large strength correlation, ranging from .30-1.0. Principle component factor analysis was used to extract factors. Due to the high correlation, an oblique approach was used with the use of direct oblimin for rotating the factors. Hair et al. (2010) suggests using a factor loading of .35 with a sample size N=278. The first iteration of factor analysis showed a four-factor solution with 72.07% variance. Both the Kaiser criterion or eigenvalue rule and the scree test were also utilized and supported the analysis.

Further refinement of the factors required the removal of cross-loading items. There were three cross-loaded items. The fourth factor only had one item load at a significant level and that item was also cross-loaded. Therefore, the removal of cross-loadings reduced the factors from four to three factors. Factor analysis was run again with the remaining 36 items and had one cross-loaded item to remove. Thus, the final analysis supported three distinct factors. The factors were interpreted as Participatory Management, Sense making, and Resourcefulness. Table 2 shows the final rotated patter matrix.

Table 2. Final Rotated Pattern Matrix for Reduced Set of 36 Items

Item	Factor 1	Factor 2	Factor 3
Is Sincere	**.959**	-.026	-.165
Encourages employees to suggest ideas and new solutions	**.899**	-.012	-.165
Is honest	**.834**	-.068	.015
Encourages cooperation between people with different skills and profiles	**.828**	.029	.034
Implements solutions to facilitate internal cooperation	**.826**	.121.	-.049
Tries to look at everybody's side of a disagreement before making a decision	**.816**	-.195	.160
Encourages employee participation in the crisis identification process	**.780**	-.085	.148
Believes there are two sides to every question and tries to look at both sides	**.779**	-.160	.152
Encourages employees to act with a view to continuously improve products, processes, and/or working methods	**.775**	.179	-.041
Encourages employees to take initiative to learn new things	**.748**	.074	.046
Organizes the management and sharing of knowledge and know-how among employees	**.739**	.206	.029
Develops employees skills with a view to the organization's future development	**.738**	.185	.027
Informs employees about upcoming changes and their implementation	**.738**	.008	.142
Communicates information about the organization and its action plans to all levels in terms easily understood by all	**.647**	.169	.106
Clearly distributes strategy to all hierarchical levels	**.547**	.226	.165
Is insightful	**.510**	.235	.210
Is capable	**.413**	.319	.273
Is confident	-.020	**.693**	.083
Able to make decisions quickly when circumstances change	.236	**.545**	.268
Handles pending crisis information in real time	.213	**.519**	.266
Deploys resources easily to respond to opportunities and threats encountered	.218	**.513**	.293
Able to identify and seize rapidly the best opportunities which come up in the environment	.325	**.485**	.222
Is inventive	**.383**	**.441**	.088
Is resourceful	.251	**.426**	.322
Does not dismiss things that do not seem normal but rather tries to interpret them	.021	-.208	**.953**
Able to see how events link together when others do not	-.064	.193	**.807**
Able to see patterns well	-.038	.133	**.797**
Tells someone when something is not normal routine	.066	-.074	**.781**
Spends time reflecting on events or behavior that does not fit the norm to determine if there is a link	.117	-.081	**.768**
Able to provide meaning to discrepancies in the normal routine	.099	.045	**.758**
Able to identify something that does not fit with normal routine	-.097	.182	**.728**
Recognizes when something seems off	.083	.124	**.715**
Brings potential failures in the system to direct supervisor	.111	.040	**.637**
Provides meaning for glitches in the system	.175	.152	**.559**
Scan and examines the environment to anticipate and prevent risks	.172	.312	**.450**

Note. Significant loadings are in bold.

3.2.3 Factors

Dimension 1: Participatory Management

There are seventeen items with this latent variable. Participatory Management is described as the inclusion of employees in terms of communication, training, information, solutions, and interactions. The scale has high internal consistency as evidenced by the Cronbach's alpha, a=.97.

- Tries to look at everybody's side of a disagreement before making a decision
- Clearly distributed strategy to all hierarchical levels
- Communicates information about the organization and its action plans to all levels in terms easily understood by all
- Informs employees to suggest ideas and new solutions
- Encourages employee participation in crisis identification processes
- Employee's skills are developed with a view to the organization future development
- Organizes the management and sharing of knowledge and know-how among employees
- Encourages employees to act with a view to continuous improvement of products, processes, and/or working methods
- Implements solutions to facilitate internal cooperation
- Encourages cooperation between people with different skills and profiles
- Encourages employees to take initiatives and to learn new things
- Believes there are two sides to every question and tries to look at both sides
- Is capable

- Is insightful
- Is honest
- Is sincere

Dimension 2: Resourcefulness

Seven items describe the latent variable of resourcefulness. These items describe a leader's ability to be agile in terms of resources. It includes agility in terms of decision making, identifying opportunities, actions, adapting to circumstances, handling information, deploying resources or assessment of the situation along with a confidence in one's ability to navigate a system fluidly. The scale has high internal consistency as evidenced by the Cronbach's alpha, a=.95

- Able to make decisions quickly when circumstances change
- Handles pending crisis information in real time
- Adapts very quickly to pending crisis developments
- Deploys resources easily to respond to opportunities and threats encountered
- Able to identify and seize rapidly the best opportunities which come up in the environment
- Is confident
- Is resourceful

Dimension 3: Sense making

There are eleven items within this latent variable. These items emphasize the ability to identify warning signs of a looming crisis and bring it to the attention of others. An important element of sense making is the ability to acknowledge what may seem implausible, interpret events as being linked, and observe what is out of a normal routine. The scale has high internal consistency as evidenced by the Cronbach's alpha, a=.95.

- Able to identify something that does not fit with normal routines
- Able to see patterns well
- Able to see how events link together even when others do not
- Spends time reflecting on events or behavior that does not seem to fit the norm to determine if there is a link
- Recognizes when something seems off
- Does not dismiss things that do not seem normal but rather tries to interpret it
- Tells someone when something is not normal or routine
- Able to provide meaning to discrepancies in the normal routine
- Provides meanings for glitches in the system
- Brings potential failures in the system to direct supervisor
- Scans and examines the environment to anticipate and prevent risks

3.3 Scale Validation

Four validation tests were conducted to further strengthen the crisis identification and aversion tool. These types include content, predictive, and discriminant.

Content validity was measured through the Delphi panel process. The individuals were deemed experts in the field of leadership, or more specifically, crisis leadership. The refinement and reduction of the items led to 41 items toward the construct and established content validity.

Predictive validity was tested with the three final factors with the Risk Propensity Scale and the Leadership Effectiveness Scale. Table 3 demonstrates that the bivariate correlations support that the three scales correlate to risk propensity. According to Pallant (2010), r=.30-.49 determines medium strength correlation and r=.50-1.0 determines large strength correlation. Thus, Resourcefulness has a large correlation with Risk Propensity. Participatory Management and Sense making have a medium correlation with Risk Propensity.

Table 3. Intercorrelations of the Three Factors for Crisis Identification and Aversion with Risk Propensity (N=278)

Variable	Participatory Management	Resourcefulness	Sense making	Risk propensity
Part.Man	-			
Resource	.838	-		
Sense Mak	.866	.852	-	
Risk Prop	.410	.522	.387	-

Because the literature posits that crisis leadership competencies differ from competencies in leadership during normal business operations, the goal was to determine if there was a correlation between the three scales and leadership effectiveness. Table 4 demonstrates that the bivariate correlations support that the three scales correlate to leadership effectiveness. Based on Pallant's (2010) guidelines for correlation, all three scales have large strength correlation with leadership effectiveness. Thus, one could postulate that developing effective leaders is correlated to crisis identification and aversion abilities within a leader.

Table 4. Intercorrelations of the Three Factors for Crisis Identification and Aversion with Leadership Effectiveness (N=278)

Variable	Participatory Management (P.M)	Resourcefulness (Res)	Sense making (S.M)	Leadership Effectiveness (L.E)
P.M	-			
Res	.838	-		
S.M	.866	.852	-	
L.E.	.861	.755	.776	-

Discriminant validity was tested with the C-LEAD scale. The C-LEAD scale assesses a leader's ability to lead during the third stage of crisis management, damage control and containment. The purpose of testing for discriminant validity is to identify if any of the factors required in the pre-crisis stages are also required in the third stage which is during the crisis event. Table 5 supports that C-LEAD and Participatory Management loaded separately with only four items cross-loaded.

Table 5. Discriminant Validity – C-LEAD and Participatory Management Scale

Item	Factor 1	Factor 2
Organizes the management and sharing of knowledge	.943	
Encourages employees to take initiatives and to learn new things	.912	
Encourages employees to suggest ideas and new solutions	.911	
Employee's skills are developed with a view to the organization's future development	.898	
Encourages cooperation between people with different skills and profiles	.879	
Implements solutions to facilitate internal cooperation	.870	
Informs employees about upcoming changes and their implementation	.857	
Encourages employee participation in crisis identification processes	.843	
Tries to look at everybody's side of a disagreement before making a decision	.792	
Encourages employees to act with a view to continuous improvement of products, processes and/or working methods	.781	
Communicates information about the organization and its action plans to all levels in terms easily understood by all	.775	
Clearly distributed strategy to all hierarchical levels	.731	
Is capable	**.508**	**.419**
Can summarize key issues involved in a situation to others regardless of how much data he/she has		.913
Can make decisions and recommendations even when he/she doesn't have as much information as he/she would like		.874
Can estimate the potential deaths and injuries that may occur as the result of his/her decisions or recommendations at work		.704
Can anticipate the political and interpersonal ramifications of his/her decisions		.697
Can modify his/her regular work activities instantly to respond to an urgent need		.619
Can make decisions and recommendations even under extreme time pressure		.595
Can determine which information is critical to relay to other units in advance of them requesting it	**.385**	**.573**
Can assess how the members of the general public are being impacted by his/her unit's actions or inactions during times of adversity	**.355**	**.569**
Can keep others abreast of his/her work activities without over-informing or under-informing them	**.453**	**.472**

Note. Cross-loading items are in bold

Table 6 supports that C-LEAD and Sense making loaded on separate factors with only two items cross-loaded.

Table 6. Discriminant Validity – C-LEAD and Sense making Scale

Item	Factor 1	Factor 2
Spends time reflecting on events or behavior that does not seem to fit the norm to determine if there is a link	.905	
Does not dismiss things that do not seem normal but rather tries to interpret it	.893	
Tells someone when something is not normal or routine	.866	
Able to see how events link together even when others do not	.813	
Able to provide meaning to discrepancies in the normal routine	.808	
Able to see patterns well	.800	
Recognizes when something seems off	.779	
Brings potential failures in the system to direct supervisor	.698	
Able to identify something that does not fit with normal routines	.689	
Provides meanings for glitches in the system	**.523**	**.364**
Can summarize key issues involved in a situation to others regardless of how much data he/she has		.882
Can make decisions and recommendations even when he/she doesn't have as much information as he/she would like		.879
Can anticipate the political and interpersonal ramifications of his/her decisions		.799
Can keep others abreast of his/her work activities without over-informing or under-informing them		.666
Can estimate the potential deaths and injuries that may occur as the result of his/her decisions or recommendations at work		.626
Can assess how the members of the general public are being impacted by his/her unit's actions or inactions during times of adversity		.624
Can make decisions and recommendations even under extreme time pressure		.617
Can determine which information is critical to relay to other units in advance of them requesting it	**.374**	**.587**
Can modify his/her regularly work activities instantly to respond to an urgent need		.567

Note. Cross-loading items are in bold

The items on the C-LEAD and Resourcefulness scale did not load on separate factors. Thus, Resourcefulness is a relevant dimension for crisis aversion as well as during the crisis event.

5 Discussion

The purpose of the study was to operationalize the two first stages in the conceptual model of crisis leadership developed by Wooten and James (2008) by developing and validating an instrument to measure competencies for crisis identification and aversion. Through the development and validation steps, the item pool describing Wooten and James (2008) five competencies loaded onto three separate factors: Participatory Management, Sense making, and Resourcefulness. The Leadership Effectiveness Scale and the General Risk Propensity Scaled established predictive validity for all three dimensions. The C-LEAD Scale established discriminant validity for Participatory Management and Sense making.

5.1.1 Practical Application

There are differing perspectives where practical application is relevant. These areas include the leadership scholar, the crisis manager practitioner, the human resource and development practitioner, and the educator.

The leadership scholar is now able to increase the quantitative research on crisis leadership, specifically on a leader's ability to avert crisis. Correlations, predictions, and differences can be studied with crisis leadership and other constructs such as transformational leadership, servant leadership, and organizational theories. This research will enhance the literature on crisis leadership.

The crisis manager practitioner can utilize the tool to assess the strengths and weaknesses of the current organizational structure as it relates to personnel and crisis plans. Training can be developed to increase cross-functional leaders in order to better equip an organization to be crisis averse.

Human resource and development practitioners often create leadership development programs. Now with a stronger understanding of the competencies needed to be crisis averse, HRD practitioners can build training specific to improve these competencies. An additive for HRD practitioners is that crisis leadership and effective leadership have a strong correlation.

Lastly, the educator would benefit due to the ability to include empirical data on crisis leadership in the leadership and business curriculum. By including crisis leadership in the education of future leaders, these leaders will be more prepared to avert organizationally generated crises.

5.1.2 Limitations

Due to the utilization of snowball sampling, there was a high percentage of respondents in the field of education, 43.2%. A post-hoc analysis was conducted to determine if this overrepresentation skewed the results. An independent samples t test was conducted with the three factors' mean scores. The first independent variable was higher education and the second independent variable was all other industries. The findings report that no significant differences exist in the mean leaves of the three scales.

5.1.3 Future Research

Due to the large number of items within the crisis identification and aversion instrument along with the items in the validity scales, there was hesitancy to include too many demographic questions because of test fatigue. This assumption proved itself true 111 responded stopping at or around question 29. That said, demographics on ethnicity, geographic regions, or age would add to the literature.

Secondly, the instrument was taken from a follower's perspective. Future research with the same tool taken as a self-report would be valuable. The challenge is whether or not a self-report would bring biases that skew the analysis.

5.1.4 Conclusion

Due to the increased crises occurring in organizations today, and the negative impact they have on organizations, leaders need to understand what it means to be a crisis leader. The Crisis Identification and Aversion Tool provides a means to understand how a leader can develop into a crisis averse leader. The first step of this understanding is quantitative assessment.

Acknowledgements

I must acknowledge those colleagues that encourage me and challenge me to continue my research – Dr. Loren O'Connor, Dr, Donald Scott, and Mr. Paul Turgeon (emerging doctor).

Conflict of Interest: none declared.

References

Albright, K. S. (2004). Environmental scanning: Radar for success. *Information Management Journal, 38*(3), 38-45.

Amabile, T. M. (1988). A model of creativity and innovation in organizations. In B. M. Straw & L. L. Cummings (Eds.), *Research in organization behavior* (Vol. 10, pp. 123-167). Greenwhich, CT: JAI Press.

Amabile, T. M., Hadley, C. N., & Kramer, S. J. (2002). Creativity under the gun. *Harvard Business Review, 80*(8), 52-61.

Bandura, A. (1997). *Self-efficacy: The exercise of control.* New York, NY: W. H. Freeman and Co.

Barron, F., & Harrington, D. M. (1981). Creativity, intelligence, and personality. *Annual Review of Psychology, 32*, 439-476

Bell, D. (1985). Disappointment in decision making under uncertainty. *Operations Research, 33*, 1-27.

Berkes, H. (2012). *Remembering Roger Boisjoly: He tried to stop shuttle Challenger launch*. Retrieved from http://www.npr.org/sections/thetwo-way/2012/02/06/146490064/remembering-roger-boisjoly-he-tried-to-stop-shuttle-challenger-launch%208/28/15

Bishop, K., Webber, S. S., & O'Neil, R. (2011). Preparation and prior experience in issue-selling success. *Journal of Managerial Issues, 23*(3), 323-340.

Blythe, B. (2002). Preparing for crisis. *Executive Excellence, 19*(12), 8.

Boin, A., t'Hart, P., Stern, E., & Sundelius, B. (2005). *The politics of crisis management public leadership under pressure.* Cambridge, MA: Cambridge University Press.

Bonvillian, G. (2013). Turnaround managers as crisis leaders. In A. Dubrin (Ed.), *Handbook of research on crisis leadership in organization* (pp. 92-109). Northampton, MA: Edward Elgar.

Charbonnier-Voirin, A. (2011). The development and partial testing of the psychometric properties of a measurement scale of organizational agility. *Management, 14*(2), 120-156.

Coman, A., & Bonciu, C. (2014). Leadership and creativity. *Manager, 19*, 27-37.

Coombs, W. T. (2006). The protective powers of crisis response strategies: Managing reputational assets during a crisis. *Public Relations Review, 38*, 150-152

Crocitto, M., & Youssef, M. (2003). The human side of organizational agility. *Industrial Management & Data Systems, 103*(5/6), 388-397.

Dacey, J. S. (1989). *Fundamentals of creative thinking.* Lexinton, MA: Lexington.

Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. *JSAS Catalog of Selected Documents in Psychology, 10*, 85.

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology, 44*(1), 113-126.

Davis, G. A., & Rim, S. B. (1985). *Education of the gifted and talented.* Englewood Cliffs, NJ: Prentice Hall.

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology, 44*(1), 113-126.

Delquie, P., & Cillo, A. (2006). Disappointment without prior expectations: A unifying perspective on decision under risk. *Journal of Risk and Uncertainty, 33*(3), 197-215.

DeVellis, R. F. (2012). *Scale development: Theory and application* (3rd ed.). Thousand Oaks, CA: Sage.

DuBrin, A. J. (2013). *Handbook of research on crisis leadership in organizations.* Northampton, MA: Edward Edgar.

Dutton, J. E., & Ashford, S. J. (1993). Selling issues to top management. *Academy of Management Review, 18*, 397-428.

Dutton, J. E., Ashford, S. J., O'Neil, R. M., & Lawrence, K. A. (2001). Moves that matter: Issue selling and organizational change. *Academy of Management Journal, 44*(4), 716-736.

Ehrhart, M. G., & Klein, K. J. (2001). Predicting followers' preferences for charismatic leadership: The influence of follower values and personality. *Leadership Quarterly, 12*(2), 153-179.

Evans, N. J., Forney, D. S., Guido, L. D., Patton, L. D., & Renn, K. A. (2010). *Student development in college: Theory, research, and practice* (2nd ed.). San Francisco, CA: John Wiley & Sons.

Feldman D. H. (1999). The development of creativity. In R. Sternberg (Ed.), *Handbook of creativity* (pp. 169-188). New York, NY: Cambridge University Press

Fink, S. (1986). *Crisis management: Planning for the inevitable.* New York, NY: AMACOM.

Galai, D., & Sade, O. (2003). *The "ostrich effect" and the relationship between the liquidity and the yields of financial assets*. SSRN working paper.

Gioia, D. A., & Thomas, J. B. (1996). Identity, image, and issue interpretation: Sensemaking during strategic change in academic. *Administrative Science Quarterly*, *41*(3), 370-403

Gough, H. G. (1979). A creative personality scale for the adjective check list. *Journal of Personality and Social Psychology, 37*(8), 1398-1405.

Hadley, C., Pittinsky, T. L., Sommer, S. A., & Zhu, W. (2009). *Measuring the efficacy of leaders to assess information and make decisions in a crisis: The C-LEAD Scale*. IDEAS Working Pater Series from REPEc.

Hadley, C., Pittinsky, T. L., Sommer, S. A., & Zhu, W. (2011). Measuring the efficacy of leaders to assess information and make decisions in a crisis: The C-LEAD scale. *The Leadership Quarterly, 22*(4), 633-648.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis* (7th ed.). Upper Saddle River, NJ: Prentice Hall.

Harraf, A., Wanasika, I., Tate, K., & Talbott, K. (2015). Organizational agility. *The Journal of Applied Business Research, 31*(2), 675-685.

Heller, J., & White, J. (2000, August 3). Ford Motor Co. is investigating reports about failures of firestone truck tires. *Wall Street Journal* (Eastern ed.), p. A6.

Heller, N. A. (2012). Leadership in crisis: An exploration of the British Petroleum case. *International Journal of Business and Social Science, 3*(18), 21-32.

Hollenbeck, G. P. McCall, M. W., & Silzer, R. F. (2006). Leadership competency models. *The Leadership Quarterly, 17*(4), 398-413.

Hung, K., & Tangpong, C. (2010). General risk propensity in multifaceted business decisions: Scale development. *Journal of Managerial Issues, 22*(1), 88-106.

Ibarra, H., & Andrews, S. B. (1993). Power, social influence, and sense making: Effects of network centrality and proximity on employee perceptions. *Administrative Science Quarterly*, *38*(2), 277-303.

Jacques, T. (2009). Issue and crisis management: Quicksand in the definitional landscape. *Public Relations Reiveiw, 35*(3), 280-286.

James, E. H., & Wooten, L. P. (2005). Leadership as (Un) usual: How to display competence in times of crisis. *Organizational Dynamics, 34*(2), 141-152.

James, E. H., & Wooten, L. P. (2010). *Leading under pressure: From surviving to thriving, before, during, and after a crisis.* New York, NY: Routledge.

Kapan, R. S., & Mikes, A. (2012). Managing risks: A new framework. *Harvard Business Review, 90*(6), 48-60.

Karlsson, N., Loewenstein, G., & Seppi, D. (2009). The ostrich effect: Selective attention to information. *Journal of Risk and Uncertainty, 38*(2), 95-115.

Kidd, P. T. (1994). *A 21st century paradigm in agile manufacturing: Forging new frontiers.* Wokingham, England: Addison-Wesley.

Klann, G. (2003). *Crisis leadership*. Greensboro, NC: Center for Creative Leadership.

Marcus, L. J., Dorn, B. C., & Henderson, J. M. (2006). Meta-leadership and national emergency preparedness: A model to build government connectivity. *Biosecurity and Bioterrorism: Biodefense Strategy, Practice, and Science, 4*(2), 128-134.

Marcus, A., & Goodman, R. (1991). Victims and shareholders. *Academy of Management Journal, 34*(2), 281-305.

Meyer, D. J., & Meyer, J. (2005). Relative risk aversion: What do we know? *Journal of Risk and Uncertainty, 31*(3), 243-262.

Miles, J. A. (2012). *Management and organization theory*. San Francisco, CA: John Wiley & Sons.

Mitroff, I. I. (1988). Crisis management: Cutting through the confusion. *Sloan Management, 29*(2), 15-20.

Mitroff, I. I. (2004). *Crisis leadership: Planning for the unthinkable*. Hoboken, JY: Wiley.

Mitroff, I. I. (2005). Crisis leadership: Seven strategies of strength. *Leadership Excellence, 22*, 11.

Noonan Hadley, C., Pittinsky, T. L., Sommer, S. A., & Zhu, W. (2011). Measuring the efficacy of leaders to assess information and make decisions in a crisis: The C-LEAD scale. *Leadership Quarterly, 22*(4), 633-648.

Nunnely, J. O. (1978). *Psychometric theory*. New York, NY: MrGraw-Hill.

Paget, M. A. (1988). *The unity of mistakes*. Philadelphia, PA: Temple University Press.

Pallant, J. (2010). *SPSS Survival manual: A step by step guide to data analysis using the SPS program* (4th ed.). New York, NY: McGraw-Hill.

Parker, S. K., & Axtell, C. M. (2001). Seeing another viewpoint: Antecedents and outcomes of employee perspective taking. *Academy of Management Journal, 44*(6), 1065-1100.

Pearson, C. M., & Clair, J. A. (1988). Reframing crisis management. *Academy of Management Review, 23*(1), 59-76.

Pearson, C. M., & Mitroff, I. I. (1993). From crisis prone to crisis prepared: A framework for crisis management. *Academy of Management Executive, 7*(1), 48-59.

Schein, E. H. (1992). *Organizational culture and leadership* (2nd ed.). San Francisco, CA: Jossey-Bass.

Schoenberg, A. (2005). Do crisis plans matter? A new perspective on leading during a crisis. *Public Relationships Quarterly, 50*, 2-7.

Starbuck, W. H. & Milliken, F. J. (1988). Executives' perceptual filters: What they notice and they make sense. In D.C. Hambrick (Ed.), *The executive effect: Concepts and methods for studying top managers* (pp. 35-65). Greenwich, CT: JAI.

Thach, L. (2012). Managerial perceptions of crisis leadership in public and private organizations: An interview study in the United States. *International Journal of Management, 29*(2), 712-725.

Vroom, V. H. (1994). *Work and motivation*. NewYork, NY: John Wiley.

Wang, J., & Hutchins, H. M. (2010). Crisis management in higher education: What have we learned from Virginia Tech? *Advances in Developing Human resources, 12*(5), 552-572.

Waugh, W. L., Jr., & Tierney, K. (2007). *Emergency management: Principles and practice for local government* (2nd ed.). Washington, DC: ICMA.

Weick, K. E. (1979). The social psychology of organizing (2nd ed.). Reading, MA: Addison-Wesley.

Weick, K. E. (1988). Enacted sense-making in crisis situations. *Journal of Management Studies, 25*(4), 306-317.

Weick, K. E. (1995). *Sensemaking in organizations*. Thousand Oaks, CA: Sage.

Weick, K. E. (2012). *Making sense of the organization: The impermenant organization*. West Sussex, UK: John Wiley and Sons.

Weick, K. E., & Sutcliffe, K. M. (2007). *Managing the unexpected: Resilient performance in an age of uncertaintly*. San Francisco, CA: John Wiley and Sons

Weick, K. E., Sutcliffe, K. M., & Obstfeld, D. (2005). Organizing and the process of sensemaking and organizing. *Organization Science, 16*(4), 409-421.

Williams, R., Bertsch, B., Dale, B., Smith, M., & Visser, R. (2006). Quality and risk management: What are the key issues? *The TQM Magazine, 18*(1), 67-86.

Woodman, R. W., Sawyer, J. E., & Griffin, R. W. (1993). Toward a theory of organizational creativity. *Academy of Management, 18*(2), 293-321.

Wooten, L. P., & James, E. H. (2008). Linking crisis management and leadership competencies: The role of human resource development. *Advances in Developing Human Resources, 10*(3), 352-379.

Yukl, G. (2006). *Leadership in organizations* (6th ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

Journal of Management Science and Business Intelligence, 2017, 2–2
Aug. 2017, pages 17-24
doi: 10.5281/zenodo. 10.5281/zenodo.835606
http://www.ibii-us.org/Journals/JMSBI/
ISBN 2472-9264 (Online), 2472-9256 (Print)

Health Recommender System using Big data analytics

J.Archenaa[1] and Dr E.A.Mary Anita[2]

[1]Research Scholar,Department of Computer Science & IT,AMET University,Chennai-India, [2]S.A.Engineering College,Chennai

E mail: archulect@gmail.com

Received on June 07, 2017; revised on July 18, 2017; published on July 22, 2017

Abstract

This paper gives an insight on how to use big data analytics for developing effective health recommendation engine by analyzing multi structured healthcare data. Evidence-based medicine is a powerful tool to help minimize treatment variation and unexpected costs. Large amount of healthcare data such as Physician notes, medical history, medical prescription, lab and scan reports generated is useless until there is a proper method to process this data interactively in real-time. In this world filled with the latest technology, healthcare professionals feel more comfortable to utilize the social network to treat their patients effectively. To achieve this we need an effective framework which is capable of handling large amount of structured, unstructured patient data and live streaming data about the patients from their social network activities.

Apache Spark plays an effective role in making meaningful analysis on the large amount of healthcare data generated with the help of machine learning components and in-memory computations supported by spark. Healthcare recommendation engine can be developed to predict about the health condition by analyzing patient's life style, physical health factors, mental health factors and their social network activities.

Machine learning algorithms plays an essential role in providing patient centric treatments. Bayesian methods is becoming popular in medical research due its effectiveness in making better predictions.For example on training the model with the age of women and diabetes condition helps to predict the chances of getting diabetes for new women patients without detailed diagnosis.

Keywords: Predictive analytics, Recommendation Systems, Bayesian rules, Big Data, Machine learning algorithms

1. Introduction

In today's digital world people are prone to many health issues due to the sedentary life-style. The cost of medical treatments also keeps on increasing. It's the responsibility of the government to provide an effective health care system with minimized cost. This can be achieved by providing patient centric treatments. More cost spent on healthcare systems can be avoided by adopting big data analytics into practice (1). It helps to prevent lot of money spent on ineffective drugs and medical procedures by making useful analysis on the large amount of complex data generated by the healthcare systems. There are also challenges imposed on the growing healthcare data. It's important to figure out how the big data analytics can be used in handling the large amount of multi structured healthcare data.

What is the need for predictive analytics in healthcare?

To improve the quality of healthcare, it's essential to use big data analytics in healthcare.

Data generated by the healthcare industry increases day by day. Big data analytics system with spark helps to perform predictive analytics on the patient data (3). This helps to alarm the patient about the health risks earlier. It also supports physicians to provide effective treatments to their patients by monitoring the patient's health condition in real-time. Diagnosis can be improved by utilizing the expert recommendations from medical forums. Customized treatment can be achieved with the help of big data analytics, which helps in improving the quality of healthcare services. It also helps to alarm government about the seasonal disease that may occur in particular locality due to the change in weather condition.

2. Big Data Use Cases for Healthcare

Many organizations are figuring out how to harness big data and develop actionable insights for predicting health risks before it can occur. Spark is extremely fast in processing large amount of multi-structured healthcare data sets, as it offers ability to perform in-memory computations. This helps to process data 100 times faster than traditional map-reduce. Spark's support for lambda architecture allows to perform both batch and real time processing.

2.1 Data Integration from Multiple sources

Spark supports fog computing which deals with Internet of Things (IOT). It helps to collect data from different healthcare data sources such as Electronic Health Record(EHR), Wearable health devices such as Fitbit, user's medical data search pattern in social networks and health data which is already stored in HDFS[5]. Data is collected from different sources and inadequate data can be removed by the filter transformation supported by spark.

2.2 High performance batch processing computation and Iterative processing

Spark is really fast in performing computations on large amount of healthcare data set. It is possible by the distributed in-memory computations performed as different clusters. Genomics researchers are now able to align chemical compounds to 300 million [2] DNA pairs within few hours using the Spark's Resilient Distributed Dataset (RDD) transformations [6]. It can be processed iteratively then.

2.3 Predictive Analytics Using Spark Streaming

Spark streaming components such as MLib helps to perform predictive analytics on healthcare data using machine learning algorithm [10]. It helps to perform real-time analytics on data generated by wearable health devices. It generates data such as weight, BP, respiratory rate ECG and blood glucose levels. Analysis can be performed on these data using k-clustering algorithms. It will intimate any critical health condition before it could happen.

The below figure represent proposed big data healthcare ecosystem using Apache Spark and Hadoop.

Apache Spark's RDD based computations is extremely fast in processing large amount of data. Real time streaming data from social networking sites can be processed effectively. Mlib –Spark's in-built library supports machine learning which is essential for designing health recommender systems. Prediction and Recommendation component are built using machine learning algorithm.

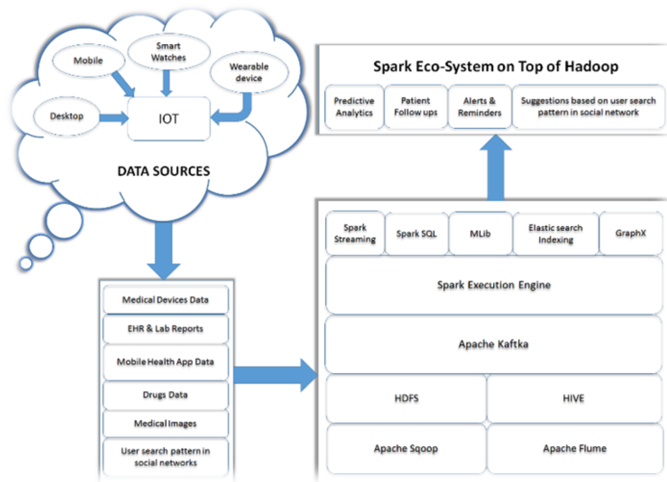

Fig. 1. Apache Spark Healthcare Ecosystem with Hadoop

3. Designing Healthcare Recommendation System

A health recommender system (HRS) suggests medical information which is meant to be highly relevant to the advancement in Medical treatment associated with the patient history.

HRS provide physicians, staff, patients and other individuals with knowledge and patient-centric information, intelligently filtered and presented at appropriate times, to enhance quality of healthcare services. Common features of HRS systems includes providing patient oriented guidance such as clinical information, integrating data from various sources (such as lab reports, medication history, imaging, wearable sensors, social media sites, health forums) into CDS (Clinical diagnosis system) application and provide relevant recommendations such as list of diagnosis, drug interaction alerts, preventative care alerts, suggesting patient centric health insurance plans, sending alerts about the hospital transportation required, sending alerts to patients about follow-ups, diet recommendations, Refilling medicines etc.

Based on user's medical expertise an HRS should suggest medical information, which is relevant to that user. Depending on the expertise of a HRS user, at least two separate use cases can be defined as follows:

1. Use case A = Health professional as end-user

In this scenario an HRS is used by a health professional to retrieve relevant information for a certain case. For example, existing clinical diagnosis, Clinical pathway or research articles from health forums can be computed automatically. This form of case-related information enrichment might support a physician with the process of clinical diagnostics as latest research results can be used for treatment decision support. In addition, naive-friendly documents can also be retrieved for the purpose supplying high quality information to patients in order to cope with a certain disease or adapt his or her lifestyle habits.

2. Use case B = Patient as end-user

In this scenario a patient interacts with a HRS-enabled PHR without direct support by a physician. HRS computes user-friendly content according to the person's case history. The relevant items are recommended to the user. By selecting the highest ranking content a patient is empowered in terms of health information.

3.1 PHR Enhanced with HRS

A PHR (Patient health record) is an electronic application through which patients can access and share their health information in a private, secure and confidential environment [6]. System is useful only if it gives valuable insights from the user health history. PHR enhanced with health recommendation component provides relevant information to the users based on their needs. It will be valuable add on to the existing healthcare system which suggests personalized or case based health recommendations.

Health Recommender Systems can assist its users in various stages in the care process, from preventive care through diagnosis and treatment to monitoring and follow-up. The most common use of HRS is for addressing clinical needs, such as ensuring accurate diagnoses, screening in a timely manner for preventable diseases, suggesting appropriate health insurance plans, alternative medicines, drug dosage recommendations or alerting adverse drug events.

4. Healthcare Recommender System Framework

Healthcare recommender system is represented by prediction and recommendation. It depends on a set of patients' case history, expert rules and social media data to train and build a model that is able to predict and recommend disease risk, diagnosis and alternative medicines. Predictions and recommendations are approved by physicians. HRS system requires input information to generate predictions and recommendations. In this work diabetes data is used as case study.

Training Data

Pile of historical medical records of diabetic patients (935 records) has been collected from hospitals. The collected data records are represented by a number of attributes, values and doctors diagnosis for each case. Diagnosis scale ranges from 1 to 10 based on the severity of the disease, 5-represents critical condition, 4-represents severe requires immediate treatment, and 3-represents moderate requires further investigation, 2-represents normal, 1-represents within control.

Demographic data of active patient

It refers to the user's data such as: name, age, location, education level, wearable device, lifestyle, food habits and type of connectivity.

Medical Case History of Diabetes Patient:

It comprises home test details such as blood sugar, blood pressure, weight. Diagnosis data comprises of physician notes, lab results, and medications. Diabetes data set is collected from KN specialty clinic and also downloaded from UCI repository.

The output of the system is

Prediction and recommendation: prediction is expressed as a numerical value that represents the disease risk diagnosis for future cases based on active patients. Recommendation is expressed as the suggestion required by the users. For example non healthcare professional might be requiring alternative remedies for treating diabetes. Healthcare professionals may be looking for disease diagnosis methods based on patient similarity.

4.1 Building the Predictive Model

Data Preprocessing

Feature selection methods can be used to identify and remove irrelevant and redundant attributes from data that do not contribute to the accuracy of a predictive model.

Data filtering is essential to avoid the creation of ambiguous or inappropriate models and improve the learning model performance. In our system, the diabetes dataset is filtered by determining the relevant features through InfoGainAttributeEval Attribute Selection method, furthermore, the data is also transformed to a form appropriate the classification.

4.2 Classification using Bayesian Network

Bayesian methods have become increasingly popular in medical research due its effectiveness in making better predictions. Diabetes is a chronic condition that occurs when the body cannot produce enough or cannot effectively use insulin [1].

Diabetes can mainly be of 3 types: Type-1 diabetes, Type-2 diabetes and Gestational diabetes. Type-1 diabetes results from non-production of insulin & Type-2 diabetes results from development of resistance of insulin, as a result of which the insulin produced is not able to metabolize the sugar levels properly. Bayesian classifier is used to predict diabetes accurately even with less amounts of training data.

Naïve Bayes is considered to be one of the most efficient and effective inductive learning algorithms for machine learning and data mining[5]. Bayesian statistics allow one to make an estimate about the likelihood of a claim and then update these estimates as new evidence becomes available.

In Bayes' probability of a hypothesis is obtained by multiplying the prior probability with the strength of the new data. The new, updated probability is called the posterior probability, or just 'the posterior'. This is the sum total of probabilities of all possible relevant hypotheses.

The posterior then becomes the new prior and the process may repeat. Let's consider how we can put Bayes' Theorem to practical use in everyday medical decision making.

1. For example 1 out of 1000 people die in diabetes it is known as the prior data.
   Prior Probability = 1/1000 = 0.001
2. Another test result indicates that there are 10% false positive result-indicates people who does not die due to diabetes
   10%*1000 = 100 false positives
3. On an average, people 101 test positive for death due to diabetes out of 1000 people(1 true positive- 1 die in accident and 100 false positive- who have diabetes but does not die)
4. Therefore 1 dies due to diabetes out of 101 people.

Without the Bayesian perspective, these 101 people will likely all become convinced that they will die due to diabetes. Bayesian statistics allows to get clearer perspective about test results by combining prior knowledge with new data and updating our position.

Baye's theorem is represented by the below formula:

$$P(H/D) = \frac{P(D/H)*P(H)}{[P(D/H)*P(H)] + [P(D/H_0)*P(H_0)]} \quad (1)$$

P(H/D) is the probability of the hypothesis (H) given the data (D),
P(D/H) is the probability of the data (D) given the hypothesis (H),

P(H) is the probability of the Hypothesis prior to the new data (also called the "prior probability" or just the "prior"), and $P(H_0)$ is the null hypothesis.
Combining background knowledge and evidence derived from data and missing data can be handled both in the construction process and in using a Bayesian network model. Expert systems based on Bayesian networks have the advantages of a formal mathematical foundation, relative computational tractability, and a graphical representation for presentation to an expert.

4.3 Parameters used in Estimation

Dataset of 1000 cases was prepared by collecting the data randomly from different groups of the society with an aim to have a variety in the dataset. To maintain accuracy and to avoid errors, data was preprocessed carefully.

Attributes	Description	Values used
Age	Age of the user	Discrete Integer Values
Sex	Male or Female	Male or Female
BMI	Body Mass Index (Height to weight ratio)	Discrete Integer Values
Family History	Any family member of the subject is suffering/ was suffering from diabetes.	Yes or No
Smoking	Smoking habits of the user	Yes or No
Drinking	Drinking habits of the user	Yes or No
Lifestyle	Lifestyle of the user	Active, Moderate, Sedentary
Eating Habits	Food habits of the user	Healthy Foods, Junk foods
Frequent Urination	Urination habits of the user	Frequent or Normal
Increased Thirst	Urge to drink more than usual	Yes or No
Fatigue	Does the user feel fatigue often?	Yes or No
Blurred Vision	Do you have blurred vision?	Yes or No
Waist Size	Waist size of the user in inches	Discrete Integer Values
Gestational Diabetes	Do you have gestational diabetes?	Yes or No
Polycystic ovaries	Do you have polycystic ovaries?	Yes or No
Fasting Plasma Glucose	Values of Fasting Plasma Glucose	Discrete Integer Values
Casual Glucose Tolerance	Values of Random Glucose tolerance test	Discrete Integer Values

Expert Rules

Fasting Plasma Glucose and Casual Glucose Tolerance Test

No Diabetes Range

If you had a fasting plasma glucose test, a level between 70 and 100 mg/dL (3.9 and 5.6 mmol/L) is considered normal.
If you had a casual glucose tolerance test, a normal result depends on when you last ate. Most of the time, the blood glucose level will be below 125 mg/dL
If your HBA1C test values is below 97 mg/dL then its normal.

Pre-diabetes Range

If your Fasting Plasma Glucose test ranges between 100 mg/dl to 125 mg/dl
If your Casual Glucose test ranges between 140 mg/dl to 199 mg/dl
If your HBA1C test values ranges between 97-154 mg/dL

Diabetes Range
If your Fasting Plasma Glucose test is 126 mg/dl or higher
If your Casual Glucose test ranges is 200 mg/dl or higher
If your HBA1C test values is greater than 180 mg/dL

Table 1. Life Style Based Analytics-Diabetes Profiling

Features	Employee A	Employee B	Diabetes Ratio A to B
Age	40	40	1 to 1
Vehicle Type	Cycle	Mini Van	1 to 10
Fast Food	Rarely	Frequent	1 to 40
Hobbies	Active Outdoor	Reading	1 to 80

This table describes about lifestyle based diabetes risk.

5. Implementation

Data from various sources combined with powerful learning algorithms and domain knowledge led to meaningful insights. Supervised pattern classification is the task of training a model based on labeled training data which then can be used to assign a pre-defined class label to new objects. Naive Bayes classifiers *are linear classifiers based on Bayes theorem*. Based on the conditional independence the presence of features are independent of each other [12]. Individual probability for all the features are calculated and classified into 3 classes: Diabetic, Pre-diabetic and No diabetic.
Individual probability is computed for all the features to classify the case as diabetic, pre-diabetic and no diabetes. **P(Diabetic='Yes') given** "Casual Glucose Tolerance Test" = 'Value from Test Data' and "Increased Blurred Vision"='Value from Test Data' .**P(Diabetic='No') given** "Casual Glucose Tolerance Test" = 'Value from Test Data' and "Increased Blurred Vision"='Value from Test Data'. Similarly the probabilities of all the features are calculated. To deal with the condition of zero probability values for unknown features Laplace smoothing is used. By calculating individual probability values, the test data gets classified into one of the three categories-Pre-Diabetic, Diabetic or Not Diabetic. The development of the system is done using Apache Spark and Python.

Table 2. Diabetes Classification using Naïve Bayes Algorithm

Class	**Entries**	**Percentage of Persons**
Pre-Diabetic	288	32
Diabetes	225	25
No Diabetes	387	43

This table describes about the results obtained after applying Bayesian network

5.1 Confusion Matrix

Confusion matrix gives a complete picture of how your classifier is performing [10]. It helps to get a picture of what your classification model is getting right and what types of errors it is making. Classification accuracy is the ratio of correct predictions to total predictions made.

Accuracy is calculated as = (TP+TN) / (P+N) (2)
P = TP + FN N = TN + FP

True positive (TP) - are the positive data set that were correctly labeled by the classifier. If the outcome from a prediction is p and the actual value is also p, then it is called a true positive (TP)[5].
True negative (TN) – are the data set predicted correctly for no diabetes. It is represented by TN.
False positive – are the data set predicted incorrectly for having diabetes. It is represented by FP.
False negative – are the data set predicted incorrectly for no diabetes. It is represented by FN.

Table 3. Number of Health Records

Number of Training Records	650
Number of Test Records	250

TP=100 FN=34 TN=70 FP=46
Accuracy= 170/134+116 = 0.68

5.2 Building the Recommendation Model

Hybrid Recommender System
Medical expert systems are a branch of artificial intelligence that applies reasoning methods and domain specific knowledge to suggest recommendations like human experts [6]. To enable reliable and fast decision making process, medical expert knowledge needs to be converted to a knowledge based system. Knowledge based system is not sufficient to suggest reliable recommendations due to the limitations in updating expert rules based on the population studies and limited personalization. Data driven approaches apply data mining and machine learning methods to extract insights from the heterogeneous data [10]. It provides individual recommendations based on the past learning experience and the patterns extracted from clinical data. Combination of information retrieval and machine learning can be used for medical database classification.

The below figure represents the source of Hybrid healthcare

recommender system.

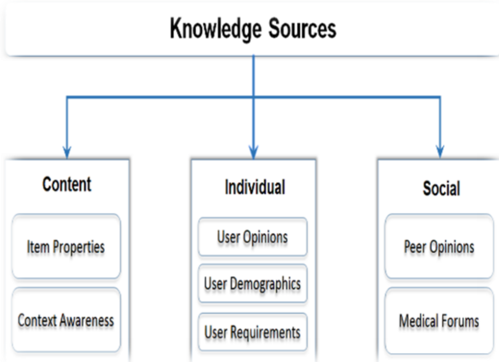

Fig. 2. Sources of Hybrid Recommender System

5.3 Types of Filtering

Collaborative filtering is the most common technique used by the recommender systems, in which the products are suggested to the user on the basis of users or items similarity. Correlations or similarities between users or items are calculated using K-Nearest Neighbor algorithm. Neighbor item ratings are combined to generate recommendations for the active user on unvisited or unrated items [7].

Content based filtering suggests the recommendations based on the user profile. For example type 2 diabetes diagnosis recommendations are suggested by keywords from patient case history [8]. The importance of words in the patient health profile can be evaluated using different weighted measure techniques, such as (a) Term Frequency/Inverse Document Frequency (TFIDF), (b) Bayesian classifiers, (c) clustering, and (d) Decision Trees (DT)[7]. For new users with few preferences, elicitation based recommendation method is used.

Domain ontologies are used to extract semantic information about items used in collaborative filtering algorithms and structured objects in medical websites as semantic entities. Ontologies are essential in healthcare domain as it helps to categorize disease, symptoms, medications, procedures, health insurance and so on.

To provide personalized healthcare recommendations, we proposed social profile enhanced recommendations based on the following criteria: (1) users with similar health concerns rate similar healthcare products, service, medication, home remedies and so on (2) users who liked similar healthcare-related items tend to like the same item in the future.

5.4 Social profile enhanced recommendation

In this approach profile similarity is computed based on the following:

1. Health Profile similarity
2. Patient behavior similarity.

Health profile information describes the patient age, location, gender and health-related concerns. Patient behavior similarity describes about the medical information accessed, healthcare social network actions, links accessed by user, user tagged by friends to health information, user's subscription to healthcare groups [5]. Combination of case based similarity and social health profile similarity is computed. If the value exceeds the threshold then the recommendation is given to the user.

5.5 Computing Case Based Similarities

Hybrid filtering approach is used to improve the accuracy of recommendations. Rule based filtering is used to filter profiles initially based on the user queries. Case based filtering is used to extract similar profiles based on patient health history. Case similarities are computed based on the KNN algorithm. Highest similarity score is suggested as recommendation.

New cases input to the HRS are compared with the existing case library. If there are no identical cases, HRS searches for the next similar cases. Case similarity is computed by KNN weighted average.

A weighted *K*-NN performs an evaluation on the attributes of the instances. Each attribute is evaluated to obtain a weight value based on how useful this attribute is for correctly identifying the classes of the dataset.

Similarity between cases is measured by the set of independent attributes. Attribute similarity is determined by the healthcare subject matter expert and stored as guiding rules. Similarity is measured by the numbers 0 or 1. Zero represents attributes are highly dissimilar and One represents attributes are likely similar. Importance of the attribute is measured as 1 –low importance or 5-high importance.

The below figure represents hybrid case based reasoning model:

Rule based methods are used for initial filtering which filters cases based on the expert rules.

Case based filtering is used to extract similar cases based on the similarities between their attributes.

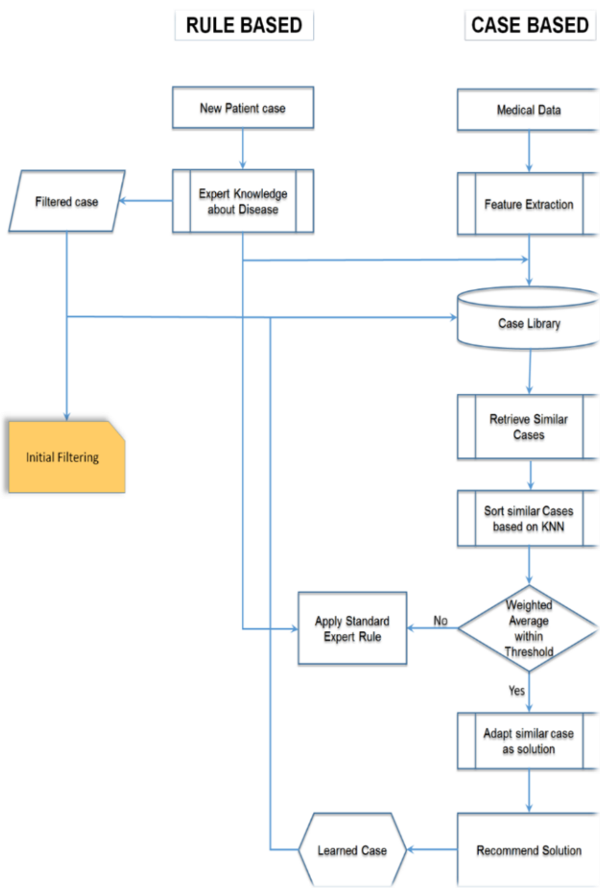

Fig 3. Hybrid Case Based Reasoning Model

The equation for computing similarity using KNN weighted average algorithm is represented as follows:

$$[1/\sum_F^N(I_F * A_F)] * \sum_X^N(I_X * A_X) \quad (3)$$

I_F and I_X represents the importance of attributes.

A_F and A_X represents the similarity scores between attributes

Table 4. Case Similarity between attributes

Features	**Case1**	**Case2**	**Case3**	**New Case**	**Similarity Score**
Age	36	25	35	27	Case1 vs new Case:0.89
Sex	Male	Female	Male	Male	
Case history	Pre-diabetes	Type 2 diabetes	Type 2 diabetes	Pre-diabetes	Case2 vs new case:0.65
Lifestyle	Moderate Exercising	Sedentary	No Exercise	Little Exercise	
Other Health concerns	Obese	Fatigue	Depression, Fatigue	Overweight	Case3 vs New case:0.37

This table shows the similarity between existing and new cases.

Similarity (New Case, Case 1)
= 1/17 * [(1*0.8) + (1*1.0) + (5*0.9) + (5*0.9) + (5*0.9)]
= 1/17 * (0.8 + 1.0 + 4.5 + 4.5 + 4.5)
= 0.89

Similarity (New Case, Case 2)
= 1/17 * [(1*1.0) + (1*0) + (5*0.7) + (5*0.6) + (5*0.7)]
= 1/17 * (1.0 + 0 + 3.5 + 3.0 + 3.5)
= 0.65

Similarity (New Case, Case 3)
= 1/17 * [(1*0.8) + (1*0) + (5*0.9) + (5*0.2) + (5*0)]
= 1/17 * (0.8 + 0 + 4.5 + 1.0 + 0)
= 0.37

Similarity computation determines which case can be suggested as recommendation, high similarity score is suggested as solution. Based on the above computation, the system will choose case 1 as the suggestion for the new case (0.89 > 0.65 > 0.37).

Can we trust recommendations?

Mass of unreliable, redundant information on the websites makes it hard to use the information for health decision making [6]. It is necessary to present a solution that user can "trust" to information and knowledge that retrieve from recommender systems.

Trust-aware recommender systems can be built by suggesting recommendations from trustable sources such as sites approved by Health on Net authority (HON)

6. Conclusion & Future Work:

In this work both the prediction and recommendation system in the context of diabetes was studied. Prediction is done using Bayesian classifier. Healthcare recommender systems are important as people use social network to know about their health condition. Accuracy of the prediction is measured using confusion matrix. Recommender systems outcomes are recommending diagnosis, health insurance, clinical pathway based treatment methods and alternative medicines to users based on their health profile similarity with others. Hybrid Recommender system filtering approach followed is –Content filtering, Rule based and Case based filtering algorithms was used. Further study on using collaborative filtering-social profile enhanced recommendation technique will be considered to improve the accuracy of recommendations. Reliability and security of social health information will be considered. Data from wearable device such as Fitbit will be considered for improving the prediction and recommendation performance.

References

1. Edwin Morely, Big Data Healthcare, IEEE explore discussion paper, 2013
2. Keith, Nitesh, Scaling Personalized Healthcare with Big Data, International big data analytics conference in Singapore, 2014
3. Cilcia Pinto, A Spark Based Workflow For Probabilistic Linkage Of Healthcare Data, Brazilian Research Council White Paper, 2013
4. Abid Sarwar, Vinod Sharma, Intelligent Naïve Bayes Approach to Diagnose Diabetes Type-2, ICNICT 2012
5. M. Kamran1, A. Javed, A Survey of Recommender Systems and Their Application in Healthcare, Technical Journal, University of Engineering and Technology (UET) Taxila, 2015
6. Punam Bedi, Trust based Recommender System for the Semantic Web, International Joint conferences on Artificial Intelligence,2014
7. Martin Wiesner and Daniel Pfeifer, Health Recommender Systems: Concepts, Requirements, Technical Basics and Challenges, International Journal of Environmental Research and Public Health, 2014
8. A. Felfernig, R.Bruke, Constraint-based Recommender Systems: Technologies and Research Issues, ICEC, 2008
9. Vladimir Hahanov, Volodymyr MizBig Data Driven Healthcare Services and Wearables, CADSM 2015
10. J.Archenaa,Dr E.A.Mary Anita, Interactive Big Data Management in Healthcare Using Spark,Springer Smart Innovation series,2016
11. https://medlineplus.gov/ency/article/003482.htm
12. http://sebastianraschka.com/Articles/2014_naive_bayes_1.html

Journal of Management Science and Business Intelligence, 2017, 2–2
Aug. 2017, pages 24-29
doi: 10.5281/zenodo.836241
http://www.ibii-us.org/Journals/JMSBI/
ISBN 2472-9264 (Online), 2472-9256 (Print)

A Manifold and Multi-Phase Framework for Bulk IT Procurement

Jim Tam[1, *], Junlian Xiang[2] and Tzu-Ming Lin[3]

[1]School of Information Technology Management, Ryerson University, Toronto, Ontario, Canada

[2]School of ICT, Seneca College, Toronto, Ontario, Canada

[3]Department of Information Management, National Central University, Taiwan

*Email: [1]jimtam@ryerson.ca, [2]junlian.xiang@senecacollege.ca, [3]jmlin@mgt.ncu.edu.tw

Received on April 11, 2017; revised on July 21, 2017; published on July 29, 2017

Abstract

This paper presents a multi-variate and multi-phase deterministic framework for IT acquisition on a very-large scale. The conventional approach of IT acquisition through open tenders or contractual negotiations invariably elicits different kinds of fallacies. The deployment of the deterministic framework minimizes the shortcomings associated with the conventional IT acquisition approach. The salient factors for ensuring procurement success for very-large-scale IT acquisition are discussed and the notion of “acquisition preparedness” is delineated. Finally, the tangible benefits of the framework are further discussed in and illustrated by an actual example of very-large scale IT acquisition undertaken by a federal government department in Canada.

Keywords: Bulk Procurement, Manifold, Multi-Phase, Framework, Acquisition Preparedness, Large-scale Acquisition, Fallacies.

1 Rationale underlying the need of very large-scale IT acquisition

In times of economic recession, business enterprises and different levels of governments strive to minimize their capital expenditures for IT hardware (HW) and software (SW) acquisition [5]. Traditionally, very-large scale or bulk purchases have offered a unified and standard approach for enterprises and governments to not only gain the best values for the capital expenditures but also minimize the administrative costs incurred in contracting [14]. The most impactful challenges that have plagued IT acquisitions over the last two decades have been examined recently [16, 17]. The importance of competitive negotiation is information technology procurement has also been highlighted in some research [18].

Typically, a very-large scale IT acquisition is characterized by the purchase of IT HW or SW on a multiple-million-dollar consumption level in US currency with the following objectives [9]:

- Satisfying customer's needs in terms of low cost, best quality, and expedient delivery;
- Minimizing administrative or operating costs;
- Conducting transactions with integrity, fairness, consistency and openness; and
- Fulfilling other enterprise or departmental objectives.

Over the past decades, enterprises have been utilizing a business strategy of acquiring other companies [2, 10] as a means of creating a competitive advantage in offering new products, reaching new markets and generating new revenues. Arising from such an “acquisition” undertaking, there is an ever-increasing need in merging and integrating dissimilar networks, systems and applications inherited from the acquired companies [1, 11]. The establishment of a common computing environment (architecture, processing platforms, networking technologies, support infrastructure, etc.), invariably facilitates data flows and further streamlines system interoperability [3]. As a result, large volume of HW and SW needs to be purchased to homogenize or consolidate various computing architectures.

Consequently, open tenders (RFPs) separately issued by individual governmental agencies or various departments in an enterprise can be consolidated into a very-large acquisition so as to leverage more buying power or negotiation advantage for better pricing schemes, merchandise delivery, support services or training offerings in IT acquisition.

However, as a result of the inflexible nature of procurement process in accommodating the “uncertainties of complex systems”, it is always a formidable challenge faced by governmental or corporate enterprises in IT acquisitions [15].

2 Conventional framework for very large-scale IT acquisition

Traditionally, in order to obtain "open and fair" competitive proposals from the vendor community, an enterprise or the governmental agency issues an RFP to the vendor community with a list of stipulated requirements and associated deadlines. Competition is normally carried out on the basis of:

(a) Pricing quotes; or

(b) Pricing quotes as well as details of proposed solution.

To ensure the fairness and confidentiality of the process, vendors are typically asked to submit their bids either through (1) sealed bidding; or (2) controlled, competitive negotiation.

Sealed bidding is characterized by a more rigid adherence to formal procedures. Those procedures aim to provide all bidders an opportunity to compete for the contract on an equal footing. In a sealed bidding acquisition, an enterprise normally awards the bidder with the lowest responsive bid (price). In contrast, competitive negotiation is a more flexible process that enables the enterprise to conduct discussions, evaluate offers, and award the contract using price and other pre-determined factors.

3 An Example for Very Large-Scale IT acquisition

A major federal government department [8] in Canada with responsibility for the provision and oversight of a safe and efficient transportation system employs more than 20,000 employees operating more than 200 facilities from coast to coast. The department operates over one hundred coast guard ships, a fleet of fixed wings and rotary wing aircrafts and many motor vehicles. It has responsibility for technical regulation, and for infrastructure, in the air, land and marine modes, providing regulation and oversight. In terms of computing architecture, the department has a heterogeneous and distributed computing environment with a myriad of computing machines ranging from main frames, minicomputers, supermicros, PCs and laptops. As a result of processing needs for ever increasing application transactions, the department undertakes to issue a very-large scale of IT acquisition with an award of a supply contract over a period of five years for meeting several hundred requirements in the following categories:

(1) Portability of applications among dissimilar systems;

(2) Interoperability of applications over various computing platforms currently in use or deployed in the future;

(3) Flexibility in replacing HW or SW components of a vendor by a different vendor;

(4) Continuity of support from vendors in operating and maintaining the computing environment;

(5) Commitment from vendors to continually develop and upgrade/update the proposed HW/SW technologies;

(6) The best pricing to reflect the ever-diminishing HW/SW pricings over the duration of the awarded contract;

(7) The expedient availability of parts/components in HW and SW (without substitutions) to be supplied throughout the duration of the contract;

(8) The commitment to jointly develop a standardized operating system environment with the governmental agency; and

(9) The training of system personnel in installing, operating, maintaining and upgrading the required system.

The winner of the contract award is required to supply, over a period of five years, the proposed system (HW, SW, installation, integration, maintenance, optimization, enhancement, and training services) to the federal government department.

4 Fallacies of the conventional Acquisition Framework

It is customary in conventional RFP award whereby winner primarily takes all. As a result, there is an imminent danger of technology lock-ins with the winning vendor. Despite the proliferation of open technologies over the past decades, different vendors still employ different kinds of proprietary "hooks" to induce involuntary lock-ins on the part of their customers. This proves to be "expensive" technologically and administratively for an enterprise.

Technological advances progress at great speed in both HW and SW. Consequently, in later stages of a multiple-year contract award, an enterprise will end up with the not-so-up-to-date HW/SW technologies or sometimes obsolete technologies of the vendor. This lagging behind in the state-of-the-art technology, nevertheless, needs to be circumvented.

HW/SW prices at the time of the contract award can be competitive. However, as time goes by, both the prices of HW and SW are dropping. The gradual drops in pricing are not normally factored into the acquisition of HW/SW in later years of the contract. As a matter of fact, the contract turns out to be an obstacle for an enterprise to get the best price for each HW/SW acquisition from the award contractor.

Delays in delivery or availability of HW/SW products are frequent in a very-large scale acquisition as there is a wide range of components to be provided by a bidding vendor which often forms partnerships with other auxiliary suppliers. The occurrence of mere minor delays in the delivery, installation, integration and testing, when all added up, can potentially be very costly for an enterprise and its operation.

Should a contractual dispute occur with the award contractor and its auxiliary partners, it is very costly for an enterprise to embark on any legal proceedings to iron out or ascertain the actual contractual responsibility of the award contractor and its auxiliary partners. The litigation will become very complex and time-consuming, particularly, when consequential damages are involved.

5 Multi-phase Framework for Very-Large Scale IT Acquisition

In view of the aforementioned limitations and fallacies of the conventional approach in IT acquisition, a deterministic and multi-phase framework, as shown in Figure 1, is deployed to issue a RFP to the vendor community by the Canadian government department.

This deterministic and multi-phase framework entails the following salient elements.

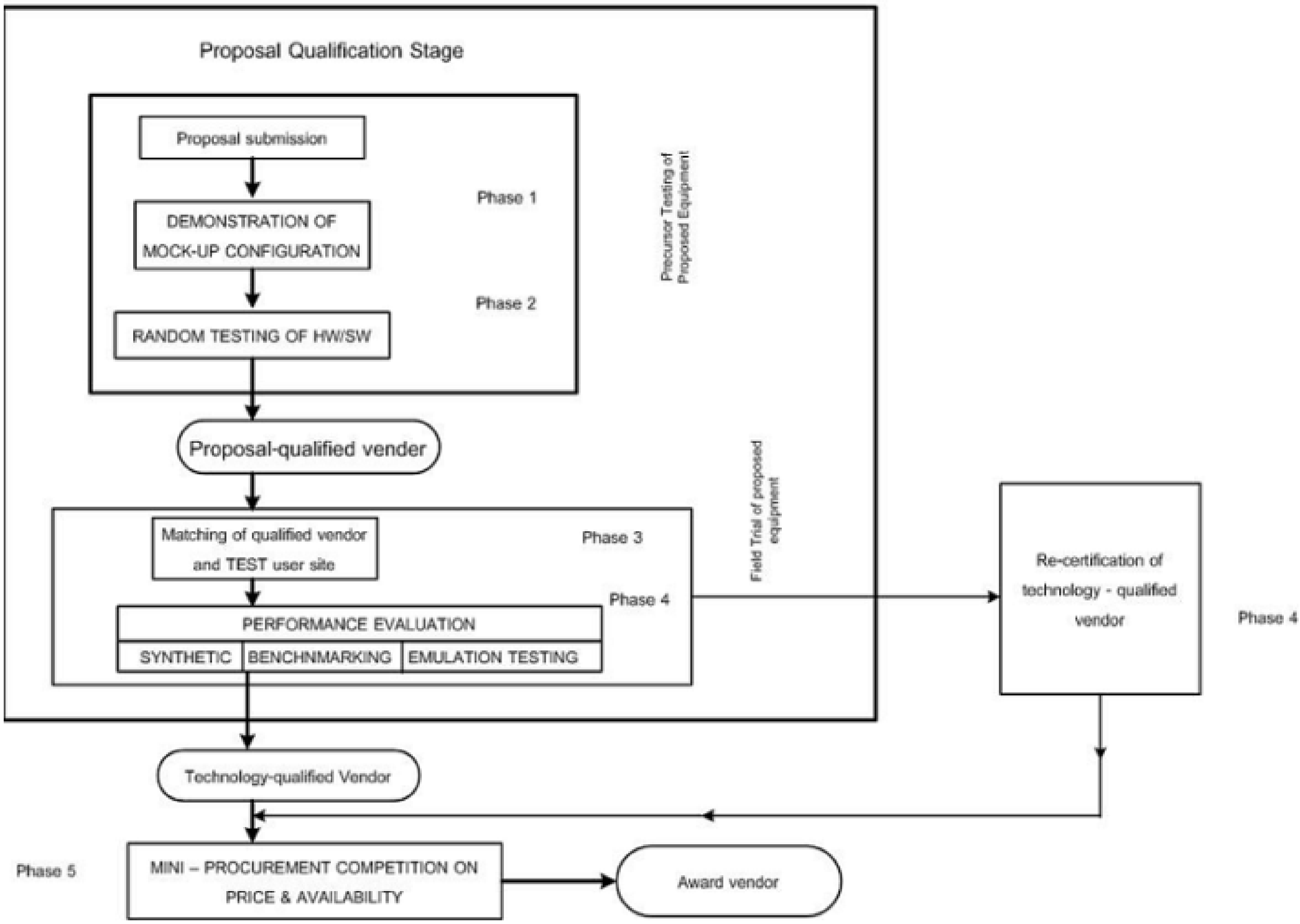

Figure 1: A Multi-Phase Large Scale IT Acquisition Cycle

5.1 Phase 1

Along with the submission of a proposal to fulfill the requirements of the RFP, each bidding vendor is required to further qualify and verify its technological capability by setting up a mock-up configuration at its company site(fully assembled at the expense of the bidding vendor) to prove the proper functioning of each proposed component, subsystem and integrated system.

5.2 Phase 2

Each bidding vendor is further required to be tested, on a random basis, over the supply of a specific HW/SW component from its partnering supplier, by the department so as to ascertain the efficiency and effectiveness with which the bidding vendor delivers the proposed component from its partner supplier. Test suites are designed and utilized by the government department to ascertain the proper functioning of the set-up systems in the specific context of simulated departmental operation.

5.3 Phase 3

The bidding vendor which has passed the previous two steps will be formally attained the "proposal-qualified" status which permits the bidder vendor to enter into the actual field trial evaluations. Each "proposal-qualified" vendor is randomly assigned to an actual user site within the department's operation to fully set up the proposed configuration (integrated systems and applications) under the normal system loadings of the user site for different kinds of performance evaluations (synthetic, benchmarking and user-transaction-emulated testing).

5.4 Phase 4

Those "proposal-qualified" vendors with attainment to pre-determined performance levels and proven capability to resolve assigned system problems within the pre-determined times will be upgraded to "technology-qualified" status.

Each "technology-qualified" vendor will annually be required to be evaluated by a technical team dispatched by the department particularly on the technological progress and new product developments. The purpose of this annual evaluation is to ensure that the "technology-qualified" vendors will continually provide the most-up-to-date technologies to the department.

At the end of the field trials in the second phase, each "technology-qualified" vendor will be upgraded to "award vendor" and be formally given a 5-year licence to supply the proposed technology (HW/SW) to the government department.

5.5 Phase 5

The actual purchase of HW/SW from each working unit of the government department will be subsequently administered through a mini-procurement, on pricing and availability only, among all the "award vendors" over the following 5-year contractual period. The award of such a multi-year license often saves significant administrative and contractual-award overheads in the procurements of IT systems. At the same time, this approach also centralizes the procurements of IT systems in accordance with the departmental guidelines and standards in terms of deployment of systems in governmental offices.

6 Salient factors ensuring success for enterprises in undertaking Very Large-Scale Acquisition

6.1 Transient Computer Markets

Computer market is constantly changing: regulatory changes (such as cyber law, telecommunication regulation, etc.); emergence of value-centered acquisition for users (best price, best availability and best technology) or an ever-changing technology (analog-digital, coaxial cable-optical fibre, baseband- broadband channel, etc.) [6]. This transience takes time to be transparent to the market and consequently creates an informational discrepancy in relation to targeted purchases to enterprises undertaking acquisitions. As a result, competing enterprises undertaking acquisitions must have the ability to comprehend the ramifications of this market transience and have the foresight (as opposed to their competitors) to preemptively acquire the best value targets. In order to develop an informational advantage surrounding potential target purchases over the competing enterprises, enterprises are required to possess a proficient knowledge of the entire value chain of the targeted market. This crucial knowledge is detrimental in deciding whether an enterprise will gain a competitive advantage in the target market and be successful in very-large scale acquisition.

6.2 Post-Contract Award Negotiation

One of the major fallacies of very-large-scale acquisitions is the overpaying of the values of acquired systems over the agreed period of contractual supply. Consistently overpaying for very-large scale acquisitions will doom any acquisition program. Enterprises undertaking very-large-scale acquisition must exercise the contractual right of negotiation with the supply vendors to iron out a differential pricing scheme that will benefit the enterprise over the whole period of the contractual award as both hardware and software depreciates in value every three to six months. In addition, a mechanism must be formulated to ensure that the supplied most up-to-date hardware and software will be delivered to the enterprises. This is in keeping with the best availability, best price and best-technology value proposition.

6.3 Preemptive Outlet and Field Testing

Given the nature and complexity of the very-large-scale acquisitions, many hardware, software and services suppliers with their affiliated subsidiaries are involved in an acquisition of this magnitude and size. As a result, there is a high probability of delays and mishaps in relation to the shipping, delivery, installation, integration and testing of the supplied hardware and software. It is therefore imperative to create a series of preemptive testing of all hardware and software at the manufacturing, supplying and shipping ends to ensure that the hardware and software are assembled in accordance with the stipulated technical specifications and standards. Corrective and remedial measures should be promptly formulated in the event of failure to adhere to the stipulated standards. This extra step will serve to eliminate any delays or mishaps in association with mal-functioning hardware or software after they are shipped to the enterprises. Further field trial testing can be done at the user sites to ensure the proper functioning of the acquired hardware and software in the actual user environments of realistic and simulated system loadings. The institution of these preemptive testing will not only minimize delays or mishaps in the shipping and delivery process but also eliminate the needs of incurring litigation against the suppliers and their subsidiaries for failed deliveries or mal-functioning of hardware or software. This will ultimately reduce unnecessary extra expenditures over the predetermined budgets and ensure success with the very-large-scale acquisition.

6.4 Streamlined System Integration

Most successful very-large-scale acquisitions need to address the following suite of questions with due diligence to ensure successful integration planning.

(1) What is the sequence by which various components or systems should be integrated?

(2) What are some of the foreseen or unforeseen obstacles in the process of system integration?

(3) What are the implications of the integration activities on users?

(4) What are our maximum points of leverage with minimum disruption in the acceptance testing of acquired HW/SW with the existing systems?

The migration and proliferation of the new integrated system into the production environment should be planned and executed incrementally in various phases to ensure a smooth and harmonized "roll-out" of the new system, interoperating in a streamlined manner, with existing systems within the enterprise.

6.5 Prior Procurement Experiences Converging to Established Procedure

Research on organizational learning suggests that procurement process stemming from experience guide organizational behavior in acquisition. Previous acquisition experience with a specific target system procurement provides opportunities to further improve the process and increases the probability of the process being utilized successfully in subsequent acquisitions [4, 12]. Empirical studies demonstrate that the more experience an enterprise's capability with a particular strategic action or direction over target system acquisition, the more likely they are to repeat acquisition with success [1]. In organizational learning theory, the resource-based view (RBV) advocates the notion that resources owned or controlled by the enterprise have the potential to provide a sustaining competitive advantage, particularly when they are not easily inimitable or substitutable [7]. The enterprise's acquisi¬tion experience can actually be viewed as a not-so-easily inimitable or construed as a non-substitutable resource. Some enterprises, based on their prior procurement experience, promulgate standardized guidelines and processes for procurement, proposal qualification, bid evaluation of hardware or software acquisition [4]. The established processes and guidelines enhance the integrity, consistency and fairness of these very-large-scale acquisitions.

7 Acquisition Preparedness

Acquisition preparedness means an ongoing attempt on the part of an enterprise to establish an infrastructure whereby it is ready to embark on very-large-scale acquisition. The preparedness involves the development of a set of seven core capabilities that facilitate acquisition on a very-large-scale. The key thing is that the organization is ready to translate them into the context of very-large-scale procurement when an acquisition comes up. The seven core capabilities for this acquisition preparedness are as follows.

(1) Strategic agility: insightful understanding of market dynamics and ability to formulate procurement action plans expediently.

(2) Market Insight: ability to grasp the changing nature of the system markets and uncover the new possibilities this has to offer.

(3)Technological Tracking: constantly building the enterprise knowledge base of technological advancement in pertinent areas.

(4) Enterprise Culture building: to create and perpetuate the responsible leadership and responsive culture throughout the enterprise.

(5) Resources management: the ability to deploy resources efficiently, effectively and productively throughout the enterprise

(6) Standardized Project and Process Management: to formulate and develop a fair, open and homogeneous procurement process with accountability, transparency and consistency. It has been found that the standardized solutions may be most effective for technology acquisitions [13].

(7) Experiential Management: the ability to learn from previous procurements and to innovate on the new possibilities that fulfill the requirements of future acquisitions.

Establishing acquisition preparedness sets the stage for generative value proposition thereby enhancing very-large-scale acquisition success.

8 Conclusion

By employing this multi-variate, multi-phase framework in very-large scale IT acquisition, the government department is able to:

(1) Select multiple "award vendors" (instead of one) which are all technologically qualified to supply the proposed system to the government;

(2) Keep the risks in malfunctioning or delays in delivery, installation, and integration of SW/HW to a bare minimum, as a result of pre-delivery trials;

(3) Eliminate substantively any system mis-claims, incapability or unavoidable failures that might occur in the system production environment, because the field trials provide the best verifications of technical claims in each bidding vendor's proposal;

(4) Get the best prices and most expedient delivery, for each working unit, from the most competitive "award vendor", as a result of each mini-procurement. This has totally eliminated the disadvantage of not getting the best price from cost deteriorations or obsolete technology in HW/SW;

(5) Obtain continually the most up-to-date HW/SW technology from the award vendor, given the annual requirement of technological re-evaluation;

(6) Eliminate the risk of potential lock-ins with a particular vendor by awarding of the contract to more than one single vendor.

Contrasting the conventional approach and the multi-phase approach in very-large-scale acquisitions, the two major strategies in acquisitions is a choice between an "expense" synergy (also known as "cost savings") strategy or a "growth" synergy strategy. The multi-variate and multi-phase approach described above embodies both strategies and has been adopted for very large-scale IT acquisition by government departments in Canada. With the identification of the salient factors for success in very-large-scale procurement and the establishment of acquisition preparedness for an enterprise, the probability of success for very large-scale IT acquisition will greatly be enhanced.

References

Amburgey,T.L., Kelly, D., Barnet, W.P., (1993), Resetting the clock: the dynamics of organizational change and failure, Administrative Science Quarterly, 38(2), 51-73.

Business Management Case Study: How Cisco Standardizes the IT Acquisition Process. (2006)

http://www.cisco.com/web/about/ciscoitatwork/business_of_it/it_acquisition_integration.html

Haleblian J, Finkelstein S. (1999), The influence of organizational acquisition experience on acqui¬sition performance: a behavioral learning perspective. Administrative Science Quarterly 44(1):29–56.

Haleblian J, Kim J. Y. (2006), The influence of acquisition experience and performance on acquisition behavior: evidence from the U.S. commercial banking industry. Academic Management Journal, 49(2):357–70.

Hayward, M. L. A., (2002), When do firms learn from their acquisition experience: evidence from 1990-1995, Strategic Management Journal 23, 21.

Kusewitt, J. B., (1985) An exploratory study of strategic acquisition factors relating to performance, Strategic Management Journal 6(2), 151-169.

Peteraf, M.J., (1993) The cornerstones of competitive advantage: A resource-based view, Strategic Management Journal, 14(2), 170-181.

Transport Canada, Unit Level System Establishment Program Detailed Report, (1993) (Publication No: TP 11519).

Vacdetta, C.L., (1999) Federal Government Contract Overview, Piper DLA, U.S.A. http://library.findlaw.com/1999/Jan/1/241470.html

Wang, C. H., Quan, X. I., Huang, S. Z. (2016). Technology acquisition through exploration alliance: network positions and technology diversity. International journal of technology intelligence and planning, 11(2), 93.

Ancarani, A., Di Mauro, C., Gitto, S., Mancuso, P., & Ayach, A. (2016). Technology acquisition and efficiency in dubai hospitals. Technological Forecasting and Social Change, 113, 475-485. doi:10.1016/j.techfore.2016.07.010

ASTAN, G. (2015). factors effecting technology acquisition decisions in national defense projects. Journal of Defense Resources Management, 6(1), 97-102.

Pierson, K., & Thompson, F. (2016). How you buy affects what you get: Technology acquisition by state governments. Government Information Quarterly, 33(3), 494-505. doi:10.1016/j.giq.2016.06.003

Sundarraj, R. P., & Talluri, S. (2003). A multi-period optimization model for the procurement of component-based enterprise information technologies. European Journal of Operational Research, 146(2), 339-351. doi:10.1016/S0377-2217(02)00553-2

Tarnoff, P. J. (2007). Principles of procurement for high-technology systems. Institute of Transportation Engineers.ITE Journal, 77(1), 38-43. Retrieved from http://ezproxy.lib.ryerson.ca/login?url=http://search.proquest.com/docview/224874655?accountid=13631

Rung, A., & Blum, M. (2015). A new vision for federal information technology procurements. Journal of Strategic Contracting and Negotiation, 1(3), 189-199. doi:10.1177/2055563615618757

Razak, A. R. A., Othman, A. A., & Sundram, V. P. K. (2015). The relationships of human success factor, information technology, and procurement process coordination on operational performance in building construction industry – A proposed conceptual framework. Procedia Economics and Finance, 31, 354-360. doi:10.1016/S2212-5671(15)01209-5

Metzger, R. S., & Kramer, L. B. (2013). The importance of competitive negotiations to state information technology procurement. The Procurement Lawyer, 48(3), 1

Journal of Management Science and Business Intelligence, 2017, 2–2
August. 2017, pages 30-35
doi: 10.5281/zenodo.851525
http://www.ibii-us.org/Journals/JMSBI/
ISBN 2472-9264 (Online), 2472-9256 (Print)

Diversity Assessment to Learn About Students' Attitudes and Awareness Concerning Diversity Prior to Enrolling In a Diversity Course and Students' Attitudes and Awareness About Diversity After the Completion of a Diversity Course

Pamela R. Rochester, Ph.D., LPC, NCC

Instructional Leadership and Support, The University of West Alabama, Station #33, Livingston, AL 35470

Email: prochester@uwa.edu

Received on August 25, 2017; revised on August 26, 2017; published on August 27, 2017

Abstract

This project assisted students in their understanding of what various beliefs are held in a cultural environment that relate to that culture's beliefs, practices, customs and behaviors that might be found to be common to people living within a particular population. Cultural environments impact the way people in the same culture develop, influencing their ideologies and personalities. Behaviors influence cultural environments which are emphasizged by the perspective of many aspects of culture which influences personal choices and behaviors.

Before the research began, the researcher obtained written approval from The University of West Alabama's Research Committee. Attached, as "ANNEX", is the approval letter from the University's Research Committee.

Keywords: Cultural Awareness, Diversity, Multicultural, Culture, Worldview, Counselling

1. Background Information

The diversity of the population in the United States continues to rise. This increase can be attributed to longer lifespans, births, as well as an increasing amount of immigrants to this country. Gandhi's familiar quote," Be the change you want to see" is a clarion-call for colleges', and universities' action.

The U.S. immigrant population stood at more than 42.4 million, or 13.3 percent, of the total U.S. population of 318.9 million in 2014, according to American Community Survey (ACS), data. Between 2013 and 2014, the foreign-born population increased by 1 million, or 2.5 percent. Immigrants in the United States and their U.S.-born children at that time numbered approximately 81 million people, or 26 percent of the overall U.S. population.

As a leading University, our facility has a responsibility to create a teaching and learning environment immersed in a variety of ideas, beliefs, and lifestyles. We live in a global society thus The University of West Alabama offers opportunities for students to experience and embrace diversity through classes, guest speaker programs, interactions with students and faculty from various backgrounds. A need for growth in multicultural competency throughout The University is evidenced by University's commitment to diversity through the Strategic Diversity Plan endorsed by the Board of Trustees in 2007 and through the increase of diversity among The University of West Alabama students with a ten-fold increase with international students from 2010 to 2013 (University Progress Report 2013).

Rochester (2016) noted this study will be conducted in sections of course CO547, Multicultural Counseling during the Summer Two session (two classes), the Fall One (two classes) 2016, and Spring One (two classes), 2017. The total number of students in these six classes was 110. Out of the 110 students in the six classes a total number of 102 participated in the research study. The researcher will is the same for all respective courses in the study. After recording the results of this study there may be more classes in the Multicultural Populations courses that the researcher will include through coordination of additional participation. Additional course sections could be added due to the recent Alabama State Department of Education determination that this course meets their requirement for diversity preparation for educators.

2. The Literature Review

"The National Center for Cultural Competence http://nccc.georgetown.edu.html promotes the value of self-assessment as a way to enhance delivery of services by individuals and organizations to cultural and linguistic populations that are increasing in diversity. As noted in this website, "Assessing attitudes, practices, policies and structures is a necessary, effective and systematic way to plan for and incorporate cultural competence" (Retrieved from http://nccc.georgetown.edu/orgselfassess.html on April 12, 2016). During the spring of 2016 the Department of Education at The University of West Alabama committed to require every graduate to enroll in a multicultural counseling course to expose students to an environment of learning more about divisive issues to develop multicultural competency.

Mariska, (2013) notes the future of multiculturalism lies with building effective multicultural skills training for counseling students and professionals. Mariska believes the greatest challenge is learning how to step outside of your own worldview. This is a challenge inherent in all aspects of counseling, as we consistently work to engage both our empathy for our clients' experiences and the knowledge that their perception of these experiences can be vastly different from our own. McCombs and Vakili emphasizes one way that diversity, e-Learning, and technology connect is through e-Learning experiences that are grounded in learner-centered principles and support the complex process of learning through collaboration and learning situated in inquiry and community. Features of such learning experiences and instructional practices are multiple ways of presenting course curriculum using a variety of technologies such as:

- Graphics
- Audio
- Video
- Animation
- Learner choice to match learner needs
- Class discussions
- Collaboration
- Problem-solving
- Flexible curriculum

Volckmann, (2012), writes, "Diversity encompasses different elements, such as "socio-economic, worldview, race, age, cultural, gender, sexual orientation, physical abilities, cognitive abilities, life experiences, and developmental stage." (D'Andrea & Daniels, 1992) Generally, multicultural counseling courses stress a combination of one or more of the three dimensions of multicultural competence-awareness, knowledge, and skills.

(Mitchell, 2015) In most four-year college strategic plans, there is a good-faith statement calling for increasing diversity as an institutional goal. There are good-even noble-reasons for doing so. The principal one is that American colleges and universities must look more like the rest of America if they are to remain relevant in the 21st century. Once federal and state governments adopted the principle of increasing access through programs like the GI Bill, direct state subsidies, the Pell Grant and various federal loan programs, there was no turning back. It's been good for America as the nation continues its chaotic march toward broader equality for its citizens. Further, the linking of access to diversity more directly reflects the changing demographics of American society, the need to retrain in a postindustrial economy with a strong manufacturing component and a growing service sector. Fundamentally, it affects America's ability to compete in a global economy.

Rosado (2015), who specializes in diversity and multiculturalism, described seven important actions involved in the definition of multiculturalism:

- **recognition** of the abundant diversity of cultures;
- **respect** for the differences;
- **acknowledging** the validity of different cultural expressions and contributions;
- **valuing** what other cultures offer;
- **encouraging** the contribution of diverse groups;
- **empowering** people to strengthen themselves and others to achieve their maximum potential by being critical of their own biases; and
- **celebrating** rather than just tolerating the differences in order to bring about unity through diversity.

"The goal of assessment is to contribute to the counselor's professional competence when dealing with diverse clientele" (Deardorff, 2009; Lonner & Hayes, 2004; Paniagua, 2010; Sternberg & Grigorenko, 2004; Sue, Arrendondo, Sternberg & Grigorenko, 2004; Sue, Arredondo & Davis 1992). "This process aims to bring people who are culturally or ethnically diverse (the "clients") together with psychologists and others who themselves differ from the clients culturally or ethnically. Counselors and therapists should be acutely aware of the responsibility they have in the assessment of persons as well as in the proper delivery of their professional skills." (American Psychiatric Association, 2000, 2013; American Psychological Association, 2001; Draguns, 1998).

Research indicates that education and the world of work have continued to increase in cultural diversity. As early as 2004, the US Department of Education Office of Civil Rights presented commitment to work with educators to strengthen multicultural competency in academia from Pre-K through Higher Education. Efforts from this commitment focused on increasing diverse inclusivfe academic communities to support student enrichment through exchange and exposure to others with talents, backgrounds, viewpoints, and experiences different from their own. As this presence of diversity in educational settings grew, need also grew for multicultural competency for educators and for support staff such as counselors. Warner (2002), notes six areas of competency within an individual's views and understanding about diversity and their perspective regarding awareness and commitment to a culturally diverse workplace. These are:
1. Awareness and Climate
2. Levels of Inclusion
3. Levels of Tolerance and Understanding
4. Degree of Empathy
5. Degree of Adaptation and Change
6. Persistence and Commitment

Research has supported inclusion of diversity training for both future educators and future counselors. Lonnquist, Banks, and Huber (2009) noted that inclusion of diversity training for future educators would serve to influence their increased cultural competency and better prepare them to provide a strong learning experience for learners with diverse needs. Carjuzza (2007) noted that while student bodies are becoming more culturally diverse, teacher bodies are becoming more culturally homogeneous. Carjuzza integrated an experiential component in a required multicultural foundations course for pre-service teachers. Currently graduate counseling and student affairs majors at UWA also participate in a similar experiential component which requires students to experience a facet of life in a culture other than their own. Shen (2007) conducted a study to assess school counseling students' self-perception of competence with Asian American students and found perceptions of reduced knowledge in comparison to awareness and skills. Shen noted that attainment of sufficient knowledge required resources outside the traditional classroom learning, thus indicating a need for identification of areas of knowledge need.

Hodges (2001) noted that, "a crucial task of college counseling centers in the 21st century was support of the growing multicultural landscape of higher education through design of service delivery to support the needs of a diverse student population. This support included examination of issues from a cultural framework and design of interventions that integrated the value of culture". Ethan and Siedel (2013) noted that professors felt they were brought into the lives of students for guidance and support whether trained for this or not. Indications were that faculty in addition to counseling center staff would benefit from enrichment of multicultural competency as they supported and guided increasing diverse student populations. A need for growth in multicultural competency throughout The University is evidenced by University commitment to diversity through the Strategic Diversity Plan endorsed by the Board of Trustees in 2007 and through the increase of diversity among UWA students with a ten-fold increase with international students from 2010 to 2013 (University Progress Report 2013).

The National Center for Cultural Competence http://nccc.georgetown.edu.html promotes the value of self-assessment as a way to enhance delivery of services by individuals and organizations to cultural and linguistic populations that are increasing in diversity. As noted on the website, " Assessing attitudes, practices, policies and structures is a necessary, effective and systematic way to plan for and incorporate cultural competence" (Retrieved from http://nccc.georgetown.edu/orgselfassess.html on April 12, 2016).

3. Project Description

The research study was designed to assess needs for growth in multicultural competency among university students as diversity grows among members of The University of West Alabama community and students prepare to serve in settings that are also growing in diversity of employees and students. The study consisted of the administration of pre and post diversity self–assessment. Students enrolled in the CO547 Summer Two, 2016 (which consisted of two six-week sessions), Fall One, 2016 (which consisted of two eight-week sessions), and Spring Two, 2017 Two, (which consisted of two eight-week sessions), and Spring One, 2017 (which consisted of two eight week sessions). Online students enrolled in these courses at The University of West Alabama were invited to participate via electronic completion of the pre and post survey. The same professor/researcher administered all of the pre-survey, and post survey instruments. The Informed Consent was completed with an electronic signature obtained. After submission of this consent, participants were invited to take an anonymous per-survey prior to learning about the Multicultural Immersion Project. The pre-survey was administered the first week of classes. The last week of classes the same students were invited to take an anonymous post survey. The study focused on the impact that the information learned in a multicultural course will increase student's skills, competence-awareness, empathy, and knowledge of a diverse population. Analyses of the pre and post diversity self-assessment tool indicated that students increased in the four dimensions which are: Multicultural skills, Multicultural competence-awareness, Multicultural empathy, and Multicultural knowledge. The diversity course which is CO547, Counseling Multicultural Populations, used for this study requires a Multicultural Immersion Experience and paper detailing the immersion. Following is the description of the Multicultural Immersion requirement. A requirement of CO547 is participating in a cross-cultural immersion experience designed by the student with supervision of the professor. The purpose of this field experience is to place the student into a cultural context where the student has little or no experiential familiarity. While such an experience is ideally suited to a study-abroad experience, meaningful cross-cultural experiences may be created at the local level in one's community. The immersion experience may focus on any of a number of cultural identity factors such as race, ethnicity, religion, sexual orientation, etc. Each student will propose a cultural immersion experience and obtain approval from the professor prior to beginning the experience. Please note that this activity requires placing the student into an identified cultural context, NOT bringing elements of an identified cultural context into the student's sphere of familiarity. This project was done during the Summer Two, Fall One, 2016 sessions, and the Spring One 2017-time frame. During the first week, the students were sent an example of a cultural immersion paper. The student does not have to visit another place to complete the cultural immersion experience. The student may do this immersion in their home city, town, or community. The students may do this study in another religion, a nursing home facility, a subculture, another school system, and with persons from a different background than the student's own background.

Time parameters for the field experience will vary, but the overall experience designed by the student should be sustained and ongoing, ideally involving an extended duration over several weeks. Limited contact duration and one-time activities or events will not fulfill the immersion requirement. The student is to do this study during the current course session; past experiences will not be approved for this project. Successful completion of the Cultural Immersion Experience will require submission of a summary which outlines dates, events, and reactions concerning activities and observations each date. A written paper summarizing the immersion experiences and overall reactions is to be submitted. The student's paper shall have a thesis statement, and background information pertaining to the culture studied. This paper is to be a minimum of 10 double spaced pages written in APA format. The cover page of this report does not count as one of the 10 pages of information. The paper is to have an introduction pertaining to the culture selected to study (the reason this culture was selected to study, and some historical background about the culture studied). The student must remember the primary focus of the paper is the immersion experience. Students are to present a list of dates worked on the project. Each date is to have the activity participated in and reactions to the activity. The paper is to have a conclusion with reactions to this experience. As with every assignment all information outside the realm of general knowledge must be referenced. Past activities undertaken by students in this course have included living with the family of a different race or ethnic group for several weeks, taking a trip to rural Mexico, staying at a Native American (Indian) reservation, participating in the religious and social events of another religious group over several weeks, and participating in the social/political activities of a gay or lesbian community. Creativity and innovation are encouraged with 25% of the grade for this activity being derived from the instructor's evaluation of the appropriateness, depth, and duration of the experience reported by the student. Students must place the Cultural Immersion papers in the Discussion Board if they wish to share with classmates. To measure the improvements in students' multicultural awareness a pre-diversity course self-assessment was administered to students enrolled in the CO547 Diversity Course before the course was taught, and before the Multicultural Immersion Experience was completed. After students had completed the CO547 Diversity course, and Multicultural Immersion Experience the post-diversity self-assessment was administered. Thus, it will be learned if students have gained multicultural awareness and positive attitudes after completing this course on diversity, and participating in a Multicultural Immersion Experience. Participation in this research project will be voluntary on the part of the students. It is anticipated that this research will show that students gain multicultural awareness and positive attitudes about diversity by taking this course. (Rochester, The University of West Alabama Blackboard CO547 courses 2016, and 2017).

4. **Conclusions and Findings**

The research study was designed to assess needs for growth in multicultural competency among University students as diversity grows among members of the University Community and students prepare to serve in settings that are also growing in diversity of employees and students. This information will support development of appropriate training and ongoing program support by offering more students an opportunity to enrich cultural competency and diversity equity in:

- Career preparation programs that require inclusion of preparation in cultural competency and diversity equity such as educator preparation and counselor preparation.
- Delivery of culturally competent counseling and other University services to members of the culturally diverse University community.

This information will also support efforts to attain seed grant funding for establishment of a sustainable program of support for the above.

Of the 102 students who voluntarily participated in this research study there was a positive outcome on the post-surveys as compared to the pre-surveys. The research reviled that the students gain knowledge as that their skills increased, as well as competence-awareness, empathy, and knowledge of a diverse population.

The intent of this proposal is to offer this training through an online module housed in a current Blackboard shell for the CO547 courses and to collaborate with University programs of study for integration as needed into their curriculums. Based on the current study, recommendations are to provide for additional CO547 Counseling Multicultural Populations Courses to the graduate student population. The current study provides stimulus for future research, and likewise, a longer term of study. A replication study will provide another safeguard to find out if this Multicultural Cultural Course offered is providing effective training. This study strengthens the University's commitment to offer diversity courses for every graduate student enrolled in a graduate program of study in the College of Education.

This information will also support efforts to attain seed grant funding for establishment of a sustainable program of support for the above. Suffice it to say that this research will assist in providing a comprehensive snapshot of the vision of a need for growth in multicultural competency throughout the University as illustrated by The University's commitment to diversity through the Strategic Diversity Plan endorsed by the Board of Trustees in 2007 and through the increase of diversity among The University of West Alabama students with a ten-fold increase with international students from 2010 to 2013 (University Progress Report 2013). It is anticipated that ongoing research will be conducted to assess the dynamic efficacy of a program to promote and sustain cultural competency among the University community.

References

ACS (2013, 2014). American Community Survey: *U.S. Department of Commerce*

American Psychiatric Association, 2000 & 2013

American Psychological Association, 2003

Carvin, S. & Wiggins, F. (1989). An Antiracism training model for white professionals. *Journal of Multicultural Counseling and Development.* 17, 105-114.

D'Andrea, M. & Daniels, J. (2012). The structure of racism: In T. Hidgers, M. Wunsch & V. Chattergy, Academic literacies in multicultural higher education. ISTE NETS Standards of *University Education of Hawaii, Honolulu, 92-97.*

Deardorff, D. K. (2009). *The Sage handbook of intercultural competence. Thousand Oaks, CA: Sage.*

Draguns, J. G, (1998). Transcultural psychology and delivery of clinical psychological services. In S. Cullari (ED.), *Foundations of clinical psychology*. Boston: Allyn & Bacon.

*International Society for Technology in Education (*2012). Retrieved from http://www.iste.og/standards

Lonner, W. J. & Hayes, S. A. (2004). Understanding the cognitive and social aspects of intercultural competence. In R. J. Sternberg & E. L. Grigorenko (Eds.), *Culture and competence Contexts of life success*. Washington, DC: American Psychological Association.

Markiska, M. (2013). Multicultural Counseling: A continual pursuit. Interviewer: Shallcross, L. *Counseling Today*, 56(3), 35-37.

Mitchell, B. (2015). Positing Multicultural Education across the Mirror of Globalization.

Multicultural Education, 22(4)*, 14*-18.

Paniagua, F. A. (2010). Assessment and diagnosis in a cultural context. In M. M. Leach & J. D. Aten (Eds.), *Culture and the therapeutic process: A guide for mental health professionals*. New York: Routledge.

Rochester, P. R. (2017). Counseling Multicultural Populations: CO 547 Course: *The University of West Alabama: Blackboard.* Livingston, Alabama: The University of West Alabama.

Rosado, C. (2015). What Makes a School Multicultural: *Critical Multicultural Pavilion; an Education Change Project*, 1-2

Sternberg, E. C., & Contexts of life success. Grigorenko, E. L. (2004). *Culture and competence: Contexts of life success*. Washington DC: American Psychological Association.

Sue, D. W., Arredondo, P., & McDavis. R. J. (1992). Multicultural counseling competencies and standards: A call to the profession. *Journal of Counseling & Development*, 70, 477-486.

UWA (2013). *The University of West Alabama: 5-year progress report.* Livingston, Alabama*:* The University of West Alabama.

Volckmann, R. (2012). Integral Leadership and Diversity—Definitions, Distinctions and Implications. *Integral Leadership Review*, 12(3), 1-21.

ANNEX to Abstract

THE UNIVERSITY of WEST ALABAMA

May 1, 2016

Pamela R. Rochester, Ph.D
Assistant Professor, Instructional Leadership and Support The
University of West Alabama
Station #33
Livingston, AL 35470

Dear Dr. Rochester

The University of West Alabama IRB has granted approval for your proposed collaboration study entitled *Diversity Assessment Learning About Student Attitudes.* The committee determined the risks are minimal/low in your project. This approval is valid from the date above through April 30, 2017.
I wish you well with your research endeavor. If you have any questions or concerns, please do not hesitate to contact me and use protocol reference # 16-033

Sincerely

Rodney J. Granec

Institutional Review Board Chair
Office of Sponsored Programs
The University of West Alabama

OFFICE OF SPONSORED PROGRAMS
The Choctaw Tavern • Station 47 • Livingston, AL 35470 • 205-652-3424 • Fax 205-652-383

Journal of Management Science and Business Intelligence, 2017, 2–2
August. 2017, pages 36-41
doi: 10.5281/zenodo.851527
http://www.ibii-us.org/Journals/JMSBI/
ISBN 2472-9264 (Online), 2472-9256 (Print)

The comparison of machine learning methods to achieve most cost-effective prediction for credit card default

Shantanu Neema[1, *], Benjamin Soibam[1]

[1]Department of Computer Science and Engineering Technology, University of Houston – Downtown

*Email: neemas1@gator.uhd.edu

Received on May 08, 2017; revised on August 23, 2017; published on August 27, 2017

Abstract

The purpose of this research is to compare seven machine learning methods to predict customer's credit card default payments in Taiwan from UCI Machine learning repository. By comparing different machine learning methods for classification; we aim to determine the best method and study the behavior of clients from each method based on a cost control perspective. Majority of customers do not default on their payments and hence a severe imbalance in classification accuracy pose a significant challenge. Objective of using various machine learning methods is to predict the best possible cost-effective outcome from the risk management perspective. Like any classification problem; the model is trained with different algorithms with re-sampling methods. A cost function is also implemented by implying a higher cost to defaulters classified not correctly. The cost function not only keeps a good balance in predictive accuracy but also a parameter well known as Mathew's Correlation Coefficient (MCC) to not compromise on losing potential customers. By varying the cost factor; we have also tried to see the behavior of each machine learning method (linear or non-linear) which will eventually help us to determine the best algorithm for the said problem. The outcome had different behavior of the results based on cost for original vs resampled data and between different methods. Depending on the trend of results (linear or non-linear) we preferred the method and type of data with non-linear trends. Non-linearity has more cofactors and hence more accuracy which was witnessed during the analysis. It was concluded that original data with Random Forest algorithm is the best in terms of a good balance on cost vs the accuracy.

Keywords: Cost factor, Predictive Accuracy, Machine Learning, Default payment

1 Introduction

In recent times, the credit card issuers regularly face credit debt crisis especially after 2008-2009 economic collapse. Many instances of over-issuing the credit cards to unqualified applicants have raised concerns. Our aim is to determine probable defaulters with reasonably good accuracy and to develop a cost-effective model where not all but the defaulters can be predicted with better accuracy while retaining good customers at the same time. It is a big challenge for any card issuing financial institution as well as for the shareholders and clients.

The use of machine learning methods has significantly increased post 2009. Butaru et al. (2016) used machine learning methods to predict delinquency across 6 major commercial banks using macroeconomic variables. Failure of commercial banks is very much related to bad credits. To evaluate the accuracy for the credit card default, many different approaches including linear discriminant analysis [Wiginton, 1980], k-nearest neighbor [Henley and Hand, 1996], classification trees [Bastos, 2007], artificial neural networks [Malhotra, 2003] etc. have been used in past. The performance of one method over the other usually depends on the problem. This paper is an attempt to address usage of the most frequent type of credit card data. This includes demographic information like age, gender, marital status etc. and the credit history showing billing and payment records to predict performance of an individual's risk when it comes to a potential defaulter. Present study tries to identify a standard method that lowers the cost along with maintaining the quality of results in terms of accuracy.

Many advanced machine learning methods can be used for classification of clients based on risky or non-risky with a predictive accuracy [Chen and Lien, 2009]. Chen and Lien used six different machine learning methods and concluded Artificial Neural Networks is the best when it comes to predictive accuracy using the same dataset. Chen and Lien, 2009 did not utilize Random forest [Leo, 2001] algorithm. These advanced methods can detect a client who might default on next payment with a high accuracy. But there is a high potential to lose many good customers as when the default detection is so specific it might categorize a lot of good customers as defaulters. But it might categorize a lot of good customers as defaulters when the default detection is so specific; it might list lot of potential good customers to fall in category of defaulters. Just to get a high predictability of defaulters, one cannot afford to lose such good customers as it might very well prove detrimental for financial institutions issuing credit cards. A good prediction will potentially have a mix of risky and non-risky clients with a better accuracy in predicting a defaulter in a cost-effective manner.

The models developed from these machine learning methods can be modified to implement a cost factor to have a risk control [Galindo and Tamayo, 2000] by penalizing false predictions of defaulters. with a good estimation on prediction of defaulters while maintaining a good number of consumers and reduce the overall cost. Present approach of implementing a risk control is an attempt to answer questions listed below by implementation of a cost control parameter:

(1) Is there any difference in cost effectiveness of different machine learning methods?
(2) Does cost is the only factor to be considered while assessing the risk?

The seven machine learning methods used in this project are as follows:

(1) Artificial Neural Networks
(2) K-Nearest Neighbor
(3) Linear Discriminant Analysis
(4) Logistic Regression
(5) Decision Tree
(6) Naïve Bayes Classifiers
(7) Random Forest.

2 Classification Accuracy

2.1 Data Description

This research is based on a multivariate classification dataset provided in UCI Machine Learning Repository. The data contains 30,000 clients with 23 attributes with no missing information (Table 1). Attributes X1 through X23 are independent variables; and class is the dependent variable with binary classes (0,1); 0 – Not defaulted, 1 – Defaulted on credit card payment.

A preliminary insight to data shows that there is a significant imbalance since approximately 78% of the clients never default (23,364 out of 30,000). All of the 23 variables from the dataset are described below and have been utilized in this research

- X1: Amount of the given credit (NT dollar): it includes both the individual consumer credit and his/her family (supplementary) credit.
- X2: Gender (1 = male; 2 = female).
- X3: Education (1 = graduate school; 2 = university; 3 = high school; 4 = others).
- X4: Marital status (1 = married; 2 = single; 3 = others).
- X5: Age (year).
- X6–X11: History of past payment. We tracked the past monthly payment records (from April to September 2005); as follows: X6= the repayment status in September 2005 X7= the repayment status in August 2005 X11 = the repayment status in April, 2005. The measurement scale for the repayment status is: -1 = pay duly; 1 = payment delay for one month; 2 = payment delay for two months; 8 = payment delay for eight months; 9 = payment delay for nine months and above.
- X12–X17: Amount of bill statement (NT dollar). X12 = amount of bill statement in September 2005; X13 = amount of bill statement in August 2005 X17 = amount of bill statement in April, 2005.
- X18–X23: Amount of previous payment (NT dollar). X18 = amount paid in September 2005; X19 = amount paid in August 2005. X23 = amount paid in April 2005.

A correlation heatmap of the data (Fig 1) was developed to check collinearity in the data. Very well-defined collinearity of the data is observed and therefore, one should consider use of penalized methods like Ridge or Lasso regression. A principal component analysis with LDA (Linear Discriminant Analysis) is also used to see if the results improve by reducing the dimensionality in the given dataset. For further analysis train dataset and test dataset are created with 2/3rd i.e. 20,000 clients for train data and remaining 1/3rd for the test data.

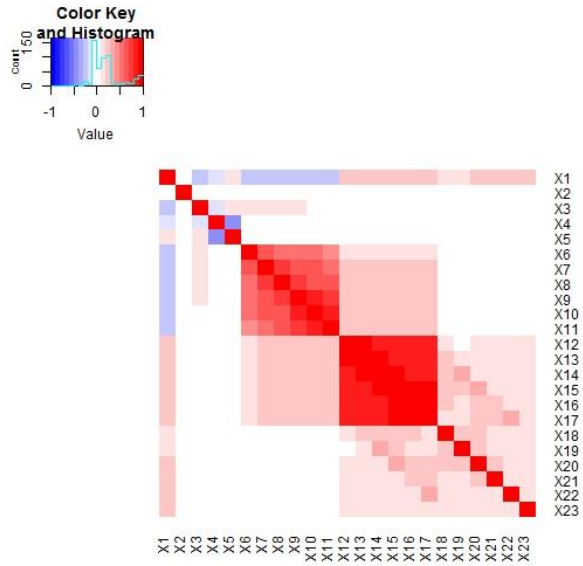

Fig. 1. Correlation Heatmap. Shows the Pearson's correlation coefficients between the different attributes.

2.2 Preliminary Accuracies

To evaluate the accuracy of default using the chosen methods in section 1, one can see 3 types of accuracies as follows

(1) Accuracy of default = No when the client is predicted as not defaulter
(2) Accuracy of default = Yes when the client is predicted as defaulter, and
(3) Overall accuracy for correct prediction of default = No and default = Yes

In imbalanced data, it is generally seen that overall accuracies might be very good but if one focuses on accuracy of default = Yes, it falls lower than 50% accurate. These preliminary accuracies are a direct result from imbalance in the dataset as explained in Section 2. A visual comparison of accuracies in all 7 methods are shown in Fig 2.

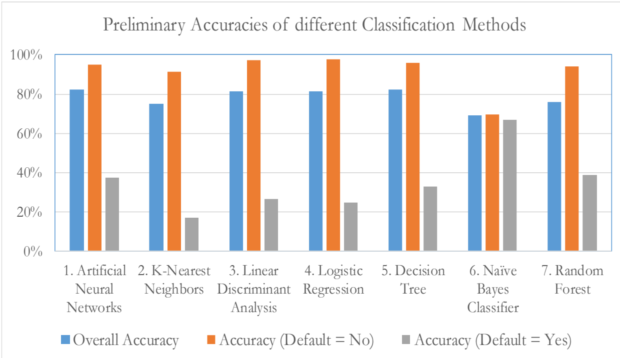

Fig. 2. Preliminary Accuracies. Shows the accuracies of default = No/Yes for all 7 machine learning methods along with overall accuracies for each method.

It can be clearly interpreted from Fig 2 that except Naïve Bayes classifiers with overall low accuracy is better than other methods in predicting both classes (i.e. default = Yes and default = No) with more than 60% accuracy. All other models provide better prediction in case when default = No but provide poor prediction (lower than 50%) in case of default = Yes. This makes them difficult to accurately predict who may potentially default on their credit card payment. One can clearly forecast a need to improve the balance in accuracies of default and no-default for all the models except Naïve Bayes which is good in predicting balanced accuracies.

3 Methodology

There are two ways to analyze data by using all seven methods to improve the balance in accuracies and keep cost effectiveness. For each method, client data is randomly divided into training data (about 2/3rd of all data) and remaining client data were used to validate the model. The dataset is used with two different approaches as shown below:

(a) Cost Function: A cost matrix shall be implemented using a cost factor (>1) for the more expensive clients presented by confusion matrix below to reclassify the classes based on cutoff probability which depends on the imbalance of the train data.

Table 1. Cost Matrix (Use of cost factor for bad clients)

		Observed	
		No	Yes
Predicted	No	0	> 1
	Yes	1	0

Cost Factor

Cost function also utilized another parameter widely known as Matthew's Correlation Coefficient (MCC) [Liu et al, 2015] defined below:

$$MCC = \sqrt{\frac{\chi^2}{n}} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$$

TP, TN, FP & FN are defined as follows:
TP – True Positive; client who did not default and predicted as not defaulter, TN – True Negative; client who default and predicted as defaulter, FP – False Positive; client who did not default but predicted as defaulter (less expensive) and FN – False Negative; client who defaulted but predicted as not defaulter (more expensive). These terms are clearly presented in Table 2.

Table 2. Confusion Matrix (Use of cost factor for bad clients)

		Observed	
		No	Yes
Predicted	No	TP	FN
	Yes	FP	TN

More Expensive

Less Expensive

Implementation of MCC will control the risk by minimizing the cost and have a better balance on the prediction as well. MCC shall be always above "zero" (means greater than 50% balanced accuracy; both sensitivity and specificity) and closer to +1. The selection of best model based on MCC and Cost Factor will be on getting "Less Cost & Similar/better MCC".

(b) Resample the train dataset such that the proportion of default is more balanced. This can be performed using following methods:

Subsampling methods:

- Under Sampling (choose less data with default = No)
- Over Sampling (choose more data with default = Yes), and

Synthetic data generation

- SMOTE (Minority oversampling)
- ROSE (Random oversampling)

From above 5 ways to analyze each machine learning method, following methodology is adopted to select the best model in each machine learning method (Fig 3). As discussed earlier cost is not the only parameter to choose the best model. In this paper, results with lower than 10% reduction in MCC were not selected as the best models to prevent significant reduction in customer base.

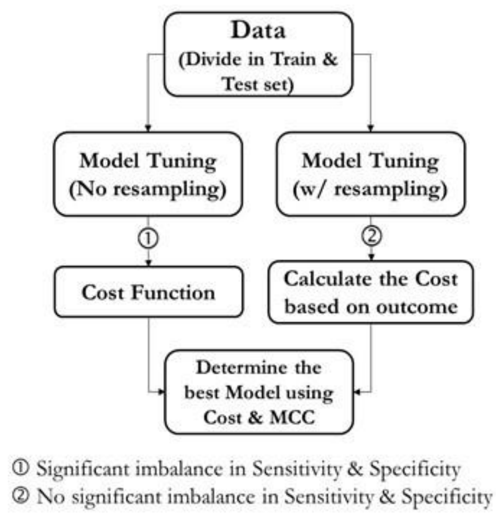

Fig. 3. Best Model Selection. Shows the flowchart on how to select the best model for each machine learning method.

4 Results

Any of the 5 ways mentioned above can be selected as the best cost-effective model for each method based on original vs resampled train data. Cost factor of 10 and 15 is chosen to understand effects of cost factor variation on performance of the model in terms of predictive accuracy and MCC. There are cases where the most cost-effective model is different for different cost factors. Furthermore, the results also justify usage of different models and, the reason behind choosing cost factor of 10 and 15.

4.1 Sample Results (ANN)

For each machine learning algorithm, the user defines range of few key parameters [Kuhn M. (2016)] and let the machine learning algorithm use the full range of these parameters. The model with the best accuracies can be called as model with best parameters. Initial part of this section summarizes results obtained from Artificial Neural Network (ANN) method (Table 3) to define the good model based on Cost and MCC (Table 2) for 30,000 clients. These sample results are presented to explain the following:

(1) Model with best parameters may not be the most cost-effective model.
(2) Models with better specificity after implementation of the cost function are the models best suited for our analysis.
(3) Models with higher drop in MCC (Old MCC vs New MCC) might be lower cost but not chosen as that is not good for our methodology.
(4) In models with resampled data; the cost function provides "Old Cost" which is the selected as cost of that model as the "New Cost" might be lower but the drop in MCC is significant drop in MCC (as much as 50% or more in many cases)
(5) A minor drop in MCC (as much as about 10%) may be acceptable if the cost reduces significantly.

Further, later part of this section summarizes (Table 4) results for all machine learning methods which includes cost of the model with (1) Original data, (2) down-sampled data, (3) up-sampled data, (4) Minority sampling (SMOTE) and (5) Random over-sampling (ROSE) All the models are chosen based on the methodology of low cost with significantly good MCC.

Table 3. Results from Artificial Neural Networks (ANN)

Cost Factor = **10** (lowest cost of **10657**)

Overall Accuracy	Old Sens	Old Spec	New Sens	New Spec	Old MCC	New MCC	Old Cost	New Cost
0.820	0.952	0.354	0.853	0.569	0.40	0.41	14613	10657
0.821	0.951	0.359	0.859	0.549	0.40	0.40	14512	11046
0.823	0.954	0.361	0.853	0.559	0.41	0.40	14451	10863
0.824	0.951	0.372	0.853	0.549	0.41	0.39	14221	11084
0.821	0.952	0.356	0.855	0.556	0.40	0.40	14566	10918
0.822	0.955	0.349	0.852	0.557	0.40	0.40	14709	10922
0.821	0.948	0.371	0.844	0.567	0.40	0.40	14285	10766
0.823	0.951	0.369	0.856	0.546	0.41	0.39	14293	11132
0.823	0.952	0.366	0.852	0.562	0.41	0.40	14351	10805

Cost Factor = **15** (lowest cost of **12204**)

Overall Accuracy	Old Sens	Old Spec	New Sens	New Spec	Old MCC	New MCC	Old Cost	New Cost
0.820	0.950	0.358	0.711	0.693	0.40	0.35	21613	12392
0.822	0.952	0.360	0.724	0.696	0.40	0.36	21551	12204
0.821	0.951	0.364	0.726	0.679	0.40	0.35	21415	12756
0.821	0.948	0.370	0.710	0.685	0.40	0.34	21238	12682
0.821	0.954	0.349	0.707	0.698	0.40	0.34	21880	12291
0.821	0.951	0.363	0.749	0.658	0.40	0.36	21443	13284
0.821	0.947	0.377	0.713	0.692	0.41	0.35	21023	12420
0.821	0.955	0.346	0.729	0.675	0.40	0.35	21965	12870
0.823	0.952	0.364	0.738	0.673	0.41	0.36	21401	12857

Model with highest accuracy
Model with lowest cost

4.2 Sample Results (ANN)

A summary of best model chosen from the cost perspective for each machine learning method are presented here. These results explain the following

(1) With different cost factors, best models may be different (i.e. with original data or with resampled data). No correlation with cost factor is observed from the results.
(2) Majority of the models with higher cost factor have shown significant reduction in MCC. This implies that the decision makers must carefully decide the cost factor to avoid risk of losing potential customers.
(3) Penalized methods have failed to show improvement in results in comparison to model with original data.
(4) SMOTE data generation method has also failed to provide good results.

4.3 Result Summary

Following conclusions can be summarized for different cost factors (10 and 15) presented in Table 4.

(1) With different cost factors; lowest cost model can be from different machine learning methods.
(2) With different cost factors; even though the best machine learning method is same, the model can be from original or resampled data.
(3) Higher cost factors have higher cost improvement but penalty on the MCC as in some models it reduces significantly.
(4) Regression and Discriminant Analysis are poorly performing models for this type of data.

At this stage; it seems difficult to establish which method is the best and based on which type of data (resampled or original). An initial look in the results are shown in table below:

Table 4. Summary of results from all Machine Learning Methods

For Cost Factor = 10

S. No	Machine Learning Method	SubSampling/ Synthetic Data	Cost	MCC
1	Random Forest	DownSampled	9478	0.37
2	Decision Tree	ROSE	9499	0.33
3	Naïve Bayes	-	9704	0.31
4	Artificial Neural Networks	DownSampled	9817	0.38
5	K-Nearest Neighbor	-	10333	0.28
6	Linear Discriminant Analysis	DownSampled	10429	0.31
7	Ridge Regression	UpSampled	10597	0.26
8	Logistic Regression	-	10676	0.26
9	Penalized Linear Discriminant	UpSampled	11235	0.37

For Cost Factor = 15

S. No	Machine Learning Method	SubSampling/ Synthetic Data	Cost	MCC
1	Random Forest	-	11435	0.30
2	Artificial Neural Networks	-	12204	0.36
3	Decision Tree	ROSE	13129	0.33
4	Linear Discriminant Analysis	-	13272	0.24
5	Naïve Bayes	-	13379	0.31
6	K-Nearest Neighbor	-	14272	0.28
7	Logistic Regression	DownSampled	14721	0.26
8	Ridge Regression	DownSampled	14721	0.26
9	Penalized Linear Discriminant	-	16205	0.37

5 Analysis and Discussion

It can be concluded from summarization of results (Section 4) that random forest is the best method for both down-sampled as well as the original data when the cost factor is 10.

Random forest models works well as for factor like cost a single model is not well suited by the fact that it has high variance due to multiple factors. On average, combined estimator using bagging based ensemble method like random forest works better as its variance is reduced.

A plot of the cost vs cost factor for random forest (Fig 4) was generated to present the difference between downs-sampled and original data.

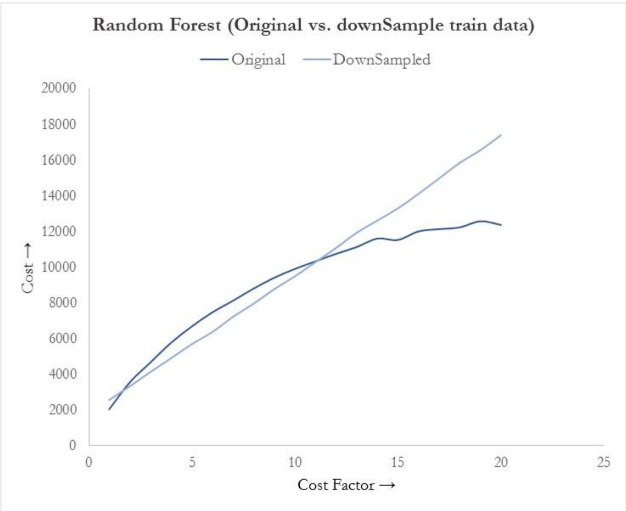

Fig. 4. Random Forest. Original vs. down-sampled train data

Similarly, in Fig 5 cost vs cost factor for ANN also represents difference between down-sampled and original data.

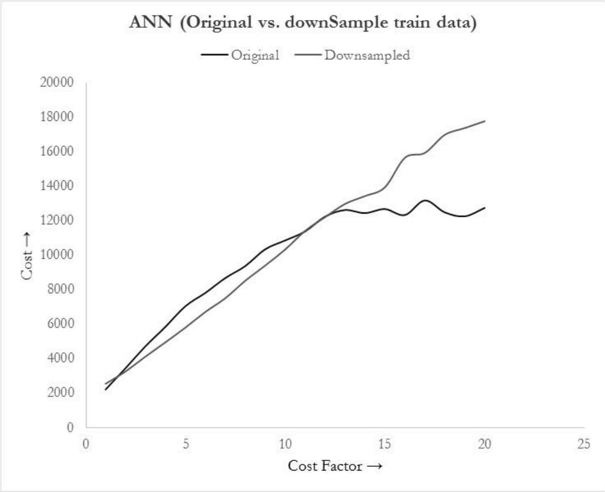

Fig. 5. Artificial Neural Networks. Original vs. down-sampled train data

From Fig. 4 and Fig. 5 it can be concluded that for low cost factor the down-sampled data presents better results than the original data. This is because; results from down-sampled data are linear in nature while the results from original data have non-linear characteristics. From risk management perspective, linear trends are in general not preferred. Non-linearity has more cofactors and hence more accurate which can be witnessed from the analysis. Therefore, it can be established that overall original data is better in terms of balance of cost vs accuracy. Also, no significant savings in cost is observed in terms of cost with down-sampled data, as compared to original data which provides greater number of savings for higher cost factors. Therefore, a non-linear model should be selected.

Similarly, random forest can be considered as best method only in the case of best cost outcome, because of the same reasons explained above. As seen in the figure below (Fig 6); random forest also has non-linear outcome with lower cost in comparison to any other method.

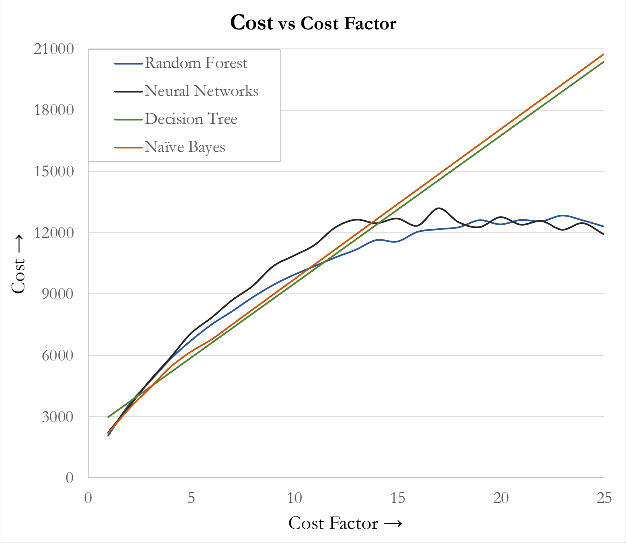

Fig. 6. Comparison of Machine Learning Methods.

For selected test data; Sample outcome of the confusion matrix is shown below:

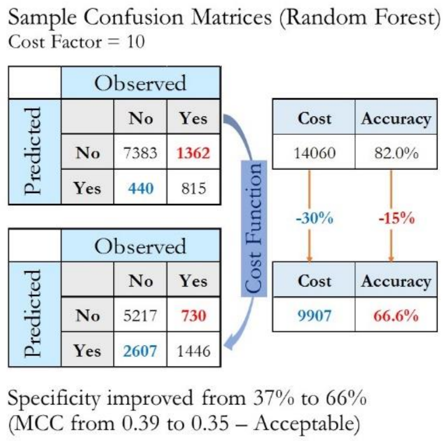

Fig. 7. Sample Results. (Confusion Matrix for 30,000 clients)

A significant reduction in the cost (30%) is observed by maintaining reasonably good MCC and Accuracy (Fig 7) is shown. Hence, random forest model should be utilized for any test set with original data.

6 Conclusion

This paper discusses 7 machine learning methods as defined in Section 1, and compares the performance of each method by considering cost-effectiveness. Each method is compared by using a cost function developed to penalized defaulters predicted as not defaulters. For a single cost factor, there are multiple results available from the confusion matrices and MCC.

However, over a range of cost factors among all 7 machine learning methods, only Random forest and artificial neural networks not only resulted in lower cost but also shows non-linearity in incurred cost per customer. Among these two methods, Random forest has the lowest cost over a larger range of cost factors.

In addition, random forest models have longer run-times and if one desire to have a better Matthew's Correlation Coefficient, Artificial Neural network is a better method. Choosing ANN model will also mean that the financial institution is more likely to take a little more risk which may be good as well.

In course of performing this analysis it was also noted that majority of machine learning methods used credit limit, billing & payment information with more importance. Random forest method was an exception and used Age as one of the top 5 variables. However, other discriminant variables like marital status, gender, education etc. are not as important as timely payments and credit limit across all 7 methods.

Overall our analysis indicates that the credit card default depends non-linearly on various factors. Therefore, ensemble method such as Random Forest and non-linear discriminators such as Neural Networks outperformed other models. We also used the Matthew's Correlation Coefficient, which has been shown to be a valid metric for evaluating model performance [Chen 2015].

Acknowledgements

I would like to thank Dr. Benjamin Soibam for letting me know about the opportunity to present our work at IIBI conference. I would also like to thank proof readers, participants, supportive family members, and Dr. Dvijesh Shastri.

Conflict of Interest: none declared.

References

Butaru, F., Chen Q., Clark B., Das S., Lo A. and Siddique A. (2016) Risk and risk management in the credit card industry, Journal of Banking and Finance, **72**, 218-239

Wiginton, J.C. (1980) A note on the comparison of logit and discriminant models of consumer credit behavior. Journal of financial quantitative analysis, **15**, 757-770

Henley, W.E. and Hand, D.J. (1996) A k-NN classifier for assessing consumer credit risk. Statistician, **45**, 77-95

Bastos, J. (2007) Credit scoring with boosted decision trees, Munich Personal RePEc Archive, 1

Makowski, P. (1985) Credit scoring branches out, Credit world , **75**, 30-37

Malhotra, R., and Malhotra, D.K. (2003) Evaluating Consumer Loans Using Neural Network, Omega, **31**, 83-96

Yeh, I. C., & Lien, C. H. (2009). The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients. Expert Systems with Applications, **36(2)**, 2473-2480

Galindo, J. and Tamayo, P. (2000) Credit risk assessment using statistical and machine learning: Basic methodology and risk modeling applications, Computational Economic, **15**, 107-143

Leo, B. (2001) Random Forests, Machine Learning, **45.1**, 5-32

Kuhn M. (2016) caret: Classification and Regression Training. R package version 6.0-73, https://CRAN.R-project.org/package=caret

Soibam, B. (2016) Performance Matrices [PowerPoint presentation]. University of Houston – Downtown

Liu, Y., Cheng, J., Yan, C., Wu X. and Chen, F (2015) Research on Matthews Correlation Coefficients metrics of personalized recommendation algorithm evaluation, International Journal of Hybrid Information Technology, **8.1**, 163-172

Liu, Y., Cheng, J., Yan, C., Wu, X., and Chen, F (2015) Research on the Matthews Correlation Coefficients Metrics of Personlized Recommendation Algorithm Evaluation, International Journal of Hybrid Information Technology, **8.1**, 163-172

Journal of Management Science and Business Intelligence, 2017, 2–2
August. 2017, pages 42-47
doi: 10.5281/zenodo.851530
http://www.ibii-us.org/Journals/JMSBI/
ISBN 2472-9264 (Online), 2472-9256 (Print)

An Empirical Analysis of Leader Personality and Servant Leadership

Yu Sun[1 *], Esther Gergen[1], Phyllis Duncan[1], Barbara Hinojosa[1] and Mark Green[1]

[1]Department of Leadership Studies, Our Lady of the Lake University, San Antonio, TX 78207

*Email: ysun@ollusa.edu

Received on May 20, 2017; revised on August 23, 2017; published on August 27, 2017

Abstract

Although servant leadership has been a popular philosophy for almost 50 years, it is only recently that peer-reviewed instruments with evidence of validity and reliability have emerged. A question related to the measurement of servant leadership is to what degree the instrument(s) are measuring constructs similar to leader personality. In this study 116 working adults completed a self-assessment of the big-five personality dimensions and their self-assessment of servant leadership using the *Servant Leadership Survey*. An exploratory factor analysis found that four of the nine scales of the *Servant Leadership Survey* appear to be measuring a construct similar to the personality dimension of agreeableness, but different from the other big-five dimensions of personality.

Keywords: Leader Personality, Servant Leadership, Factor Analysis

1 Introduction

Up until the mid-2000's, servant leadership was a popular philosophy but generally lacked a testable set of constructs. A step toward a more concrete definition occurred in 1995 with Spears' 10 aspects of servant leadership (1995). Barbuto and Wheeler (2006); Liden, Wayne, Zhao and Henderson (2008); Sendjaya, Sarros and Santora (2008); and van Dierendonck and Nuijten (2011) have each developed instruments to measure servant leadership. Each instrument has an associated peer-reviewed article that describes the development, reliability and validity of the instrument. The instrument used in this study was the *Servant Leadership Survey* by van Dierendonck and Nuijten (2011). Table 1 provides the dimensions measured by the instrument.

As theories progress and instruments are developed to measure those theories, two of the psychometric properties that gradually become clearer with increased reporting on an instrument are convergent and discriminant validity. Convergent and discriminant validity are considered subcategories of construct validity. To establish construct validity, we need to show that both convergent and discriminant validity are demonstrated. Convergent validity means two measures of constructs that theoretically should be related, are in fact related. Convergent validity can be established if two similar measures of a construct correspond with one another by calculating correlation coefficient. Correlations between theoretically similar measures should be high, while correlations between theoretically dissimilar measures should be low.

Table 1. Dimensions of the servant leadership survey

Empowerment	Is a motivational concept focused on enabling people and encouraging personal development
Accountability	Is holding people accountable for performance they can control
Standing Back	Is about the extent to which a leader gives priority to the interests of others first and gives them the necessary support and credits.
Humility	Is the ability to put one's own accomplishments and talents in a proper perspective
Authenticity	Is closely related to expressing the "true self," expressing oneself in ways that are consistent with inner thoughts and feelings
Courage	Is daring to take risks and trying out new approaches to old problems
Interpersonal Acceptance	Is the ability to understand and experience the feelings of others, understand where people come from and the ability to let go of perceived wrongdoings and not carry a grudge into other situations
Stewardship	Is the willingness to take responsibility for the larger institution and go for service instead of control and self-interest

Discriminant validity tests whether measures of constructs that theoretically should not be related to each other are, in fact not related. Discriminant validity applies to two dissimilar constructs that are easily differentiated. A successful evaluation of discriminant validity shows that a test of a concept is not highly correlated with other tests designed to measure theoretically different constructs.

Of the constructs for which there may be insufficient discriminant validity for the *Servant Leadership Survey*, personality is a logical possibility. While there are multiple models of personality, a frequently used model is the Big-Five model. This model conceptualizes personality as a combination of openness to new experiences, conscientiousness, extraversion, agreeableness, and neuroticism. As theories progress and instruments are developed to measure those theories, two of the psychometric properties that gradually become clearer with increased reporting on an instrument are convergent and discriminant validity. Convergent and discriminant validity are considered subcategories of construct validity. To establish construct validity, we need to show that both convergent and discriminant validity are demonstrated. Convergent validity means two measures of constructs that theoretically should be related, are in fact related. Convergent validity can be established if two similar measures of a construct correspond with one another by calculating correlation coefficient. Correlations between theoretically similar measures should be high, while correlations between theoretically dissimilar measures should be low.

Table 2. Dimensions of personality

Domain	Higher	Lower
Extraversion	Like People Prefer Large Groups and Gatherings Assertive, Active, and Talkative	Reserved but Not Necessarily Unfriendly Independent Rather Than a Follower
Openness	Curious about Both Inner and Outer Worlds	Prefer Familiar to the Novel
Agreeableness	Fundamentally Altruistic Sympathetic to Others and Eager to Help Them	Disagreeable or Antagonistic Toward People Skeptical of Others' Intentions
Conscientiousness	High Degree of Organization, Persistence, Control and Motivation in Goal Directed Behavior	More Lackadaisical in Working Toward Goals
Neuroticism	Identifies Individuals Who Are Prone to Psychological Distress	Emotionally Stable Usually Calm, Even-Tempered and Relaxed

The explanations for each of the big five personality types are based on: McCrae, R. R. and Costa, P. T., Psychological Assessment Resources, Inc. (2010). NEO inventories for the NEO Personality Inventory-3 (NEO-PI-3), NEO Five-Factor Inventory-3 (NEO-FFI-3), NEO Personality Inventory-Revised (NEO PI-R): Professional manual. Lutz, FL: PAR.

2 Purpose of the Study

Discriminant validity tests whether measures of constructs that theoretically should not be related to each other are, in fact not related. Discriminant validity applies to two dissimilar constructs that are easily differentiated. A successful evaluation of discriminant validity shows that a test of a concept is not highly correlated with other tests designed to measure theoretically different constructs.

Of the constructs for which there may be insufficient discriminant validity for the *Servant Leadership Survey*, personality is a logical possibility. While there are multiple models of personality, a frequently used model is the Big-Five model. This model conceptualizes personality as a combination of openness to new experiences, conscientiousness, extraversion, agreeableness, and neuroticism.

The purpose of this study was to analyze the degree to which scales of the *Servant Leadership Scale* (SLS) discriminate from the scales of the big-five personality model. Put simply, to what degree do the scales of the SLS measure something different than leader personality.

3 Previous Research

3.1 Leader Personality and Ratings of Servant Leadership

Two empirical studies were found in the peer-reviewed literature related to leader personality and ratings of servant leadership. Politis and Politis (2012) administered an abbreviated version of Page and Wong's (2000) *Servant Leadership Profile*. In this study, leaders self-assessed their personality and the degree to which they believed they were servant leaders. Leader openness, conscientiousness, extraversion and agreeableness were moderately to strongly positively related to all of subscales of the servant leadership instrument. Neuroticism was weakly negatively correlated with each servant leadership subscale.

3.2 Leader Personality and Other Aspects of Leadership Besides Servant Leadership

Table 4. Leader Conscientiousness and Ratings of Leadership

Leadership Dimension	k	N	Effect
Management by Exception Active*	6	1,469	-.04
Contingent Reward*	6	1,469	.03
Passive Leadership*	7	1,564	.04
Managerial Performance*	44	8,678	.10
Intellectual Stimulation*	8	1,828	.11
Individual Consideration*	8	1,828	.11
Transformational Leadership*	19	3,887	.15
Charisma*	9	1,706	.22
Leader Emergence**	20	NP	.24
Leader Effectiveness **	17	NP	.24

k is the number of effect sizes, N is the total sample size from those k studies, NP indicates the total N was not provided, effects provided are as follows: *Bono & Judge, (2004) reported the estimated population correlation; Judge, **Bono, Ilies & Gerhardt, (2002) reported the estimated corrected correlation; ***Barrick, Mount & Judge, (2001) reported the estimated true correlation.

Washington, Sutton and Feild (2006) administered Dennis and Winston's (2003) 23-item servant leadership scale to 283 followers of

155 supervisors. Supervisors reported their own personal agreeableness using 12 items from Costa and McCrae's (1998) NEO Five-Factor Inventory. A moderate positive correlation was found between the leaders' personality facet of agreeableness and follower ratings of those leaders' use of servant leadership ($r = .38$, $p < .01$).

The combination of leader personality and leadership style have been studied extensively. To date, that primary body of literature has been done of the full range mode of leadership, leader emergence and leader effectiveness (Green, 2015).

Table 3 provides a summary of three meta-analyses related to leader openness and follower ratings of leadership style.

Table 4 provides a summary of three meta-analyses related to leader conscientiousness and follower ratings of leadership style. Generally, while the effect sizes are also weak to moderate, the overall conclusion from the literature is that leader conscientiousness is related to several dimensions of leadership.

Table 4. Leader Conscientiousness and Ratings of Leadership

Leadership Dimension	k	N	Effect
Passive Leadership*	7	1,564	-.11
Management by Exception Active*	6	1,469	-.02
Contingent Reward*	6	1,469	.02
Intellectual Stimulation*	8	1,828	.03
Charisma*	8	1,605	.05
Individualized Consideration*	8	1,828	.11
Leader Effectiveness**	18	NP	.11
Transformational Leadership*	18	3,516	.13
Managerial Performance***	60	11,325	.25
Leader Emergence**	17	NP	.33

k is the number of effect sizes, N is the total sample size from those k studies, NP indicates the total N was not provided, effects provided are as follows: *Bono & Judge, (2004) reported the estimated population correlation; Judge, **Bono, Ilies & Gerhardt, (2002) reported the estimated corrected correlation; ***Barrick, Mount & Judge, (2001) reported the estimated true correlation.

Table 5 provides a summary of three meta-analyses related to leader extraversion and follower ratings of leadership style. As seen with the previous two personality dimensions, while effect sizes are weak, the overall conclusion from the literature is that leader extraversion is related to dimensions of leadership.

Table 5. Leader Extraversion and Ratings of Leadership

Leadership Dimension	k	N	Effect
Passive Leadership*	6	1,310	-.09
Management by Exception Active*	5	1,215	-.03
Contingent Reward*	5	1,215	.14
Intellectual Stimulation*	7	1,574	.18
Individual Consideration*	7	1,574	.18
Managerial Performance***	67	12,602	.21
Charisma*	9	1,706	.22
Transformational Leadership*	2	3,692	.24
Leader Effectiveness**	2	NP	.24
Leader Emergence**	3	NP	.30

k is the number of effect sizes, N is the total sample size from those k studies, NP indicates the total N was not provided, effects provided are as follows: *Bono & Judge, (2004) reported the estimated population correlation; Judge, **Bono, Ilies & Gerhardt, (2002) reported the estimated corrected correlation; ***Barrick, Mount & Judge, (2001) reported the estimated true correlation.

Table 6 provides a summary of three meta-analyses related to leader agreeableness and follower ratings of leadership style. Generally, while the effect sizes are weak, the overall conclusion from the literature is that leader openness is weakly related to several dimensions of leadership.

Table 6. Leader Agreeableness and Ratings of Leadership

Leadership Dimension	k	N	Effect
Passive Leadership*	7	1,564	-.12
Management by Exception Active*	6	1,469	-.11
Managerial Performance***	5	9,864	.10
Transformational Leadership*	2	3,916	.14
Intellectual Stimulation*	8	1,828	.14
Individual Consideration*	8	1,828	.17
Contingent Reward*	7	1,622	.17
Charisma*	9	1,706	.21
Leader Effectiveness**	1	NP	.21

k is the number of effect sizes, N is the total sample size from those k studies, NP indicates the total N was not provided, effects provided are as follows: *Bono & Judge, (2004) reported the estimated population correlation; Judge, **Bono, Ilies & Gerhardt, (2002) reported the estimated corrected correlation; ***Barrick, Mount & Judge, (2001) reported the estimated true correlation.

Table 7 provides a summary of three meta-analyses related to leader neuroticism and follower ratings of leadership style. Generally, while the effect sizes are weak, the overall conclusion from the literature is that leader neuroticism is weakly and negatively related to several dimensions of leadership, with the exception of weak positive relationships with Management by Exception Active and Passive Leadership.

Table 7. Leader Neuroticism and Ratings of Leadership

Leadership Dimension	k	N	Effect
Managerial Performance***	63	11,591	-.09
Individual Consideration*	9	1,772	-.10
Contingent Reward*	7	1,532	-.10
Intellectual Stimulation*	9	1,772	-.12
Transformational Leadership*	18	3,380	-.17
Charisma*	10	1,650	-.17
Leader Effectiveness**	18	NP	-.24
Leader Emergence**	30	NP	-.22
Management by Exception Active*	7	1,532	.02
Passive Leadership*	8	1,627	.05

k is the number of effect sizes, N is the total sample size from those k studies, NP indicates the total N was not provided, effects provided are as follows: Bono & Judge, (2004) reported the estimated population correlation; Judge, Bono, Ilies & Gerhardt, (2002) reported the estimated corrected correlation; Barrick, Mount & Judge, (2001) reported the estimated true correlation.

Table 8. Comparison of Leader Personality, Relationships Between Transformational and Servant Leadership

	O	C	E	A	N
Transformational Leadership k = 18 to 20	P = .15 FR	P = .13 FR	P = .24 FR	P = .14 FR	P = -.17 FR
Servant Leadership k = 2	Mod to Strong LR	Mod to Strong LR	Mod to Strong LR	r = .38 FR Mod to Strong LR	Mod to Strong LR

FR indicates Leaders Rated Themselves on Personality and Followers Rated the Leader on Leadership. LR indicates Leaders Rated Themselves on both Personality and Leadership. Columns are for Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism

4 Sample

The sample from this study consisted of 116 working adults from Houston, San Antonio, Laredo and smaller cities throughout south Texas. The working adults came from a wide mix of small, medium and large organizations, both for and non-profit. Participant ages ranged from 29 to 71 with concomitant years of work experience. Seventy-three percent of the participants were female, and 27% were male. The sample ethnicity composition consisted of 67% Hispanic, 21% White, 9% Black and 3% “other” ethnicities such as Asian and American Indian.

All participants held master’s degrees and had professional work experience. The areas of master’s preparation, however, was typically in a discipline other than leadership. Typical master’s degrees included business, education, sociology/social work, psychology and counseling, and various computer information systems areas. Each participant had been selected for admission as a part-time student in a doctoral program in leadership studies. As part of the orientation to the program, the students self-assessed their personality using the *Mini-International Personality Item Pool* and the *Servant Leadership Survey.* The self-assessments occurred prior to the beginning of training/education in theories of personality and servant leadership.

5 Instruments

5.1 Mini-International Personality Inventory Pool

Donnellan et al. (2006) created the *Mini-International Personality Item Pool* (Mini-IPIP) (2006) with consideration of the need to assess each personality domain, while retaining desirable psychometric properties of validity and reliability (Baldasaro et al., 2013). Donnellan et al. (2006) noted the Mini-IPIP (2006) was based on the *50-item International Personality Item Pool Five-Factor Model* (IPIP-BF) measure (Goldberg, 1999) and was “developed and validated across five studies.” The Mini-IPIP is a 20-item instrument which covers all five traits, with only four questions per trait, measured on a five-point Likert scale (Baldasaro et al. 2013).

The instrument was developed and validated through five studies. According to Donnellan et al. (2006), the first study administered the instrument to a large sample (N = 2,663) of students from multiple colleges and universities. The second study had a smaller sample (N = 329) and was used to examine how well the Mini-IPIP related to the Big Five, the IPIP FFM, and the TIPI. Study three examined relativity to another Big-Five measure and criterion measures. Studies four and five focused on the test-retest reliability for both short-term and long-term scales and criterion levels. The five studies revealed internal consistencies which were consistent, and were at, or above, (α = .60) (Donnellan et al. 2006).

Researchers have continued to assess reliability and validity of this instrument since its creation in 2006. Cooper, Millie, and Corer (2010) conducted a confirmatory factor analysis on the Mini-IPIP. Participants (N = 1,481) took the assessment online. The mean scores for each scale were broadly consistent with the previous data of Donnellan et al. (2006). Cronbach’s alpha scores were also found to be acceptable. Cooper et al. (2010) concluded that the Mini-IPIP is a good instrument to be used in situations with time constraints with Cronbach’s alpha scales for each at: Conscientiousness (α = .65), Extraversion (α = .71), Agreeableness, (α = .70), Openness (Intellect/Imagination) (α = .65), and Neuroticism (α = .62) (Baldasaro et al., 2013).

5.2 Servant Leadership Survey

The instrument underwent three stages in its development. In the first stage, 688 volunteers completed an early version of the survey that had 99 items. Based on those data, the authors conducted an exploratory factor analysis (EFA) that found fourteen factors with eigenvalues greater than 1. An iterative set of exploratory factor analyses using *Varimax* and *Oblimin* rotation eventually produced a six-factor model based on 28 items. At this stage of development, neither *Humility* nor *Stewardship* loaded on a unique single component. The authors added 11 additional questions designed to measure those hypothesized dimensions. This resulted in 39 possible questions.

The authors next asked an additional 263 individuals to complete the 39-question instrument. Based on those responses, the authors conducted a confirmatory factor analysis (CFA). Following the initial CFA, nine questions were removed. The reduced 30-question model produced a good fit for an 8-factor model (X^2 = 623, df = 377, CFI = .93, TLI = .92, $SRMR$ = .05, AIC = 19354, $RMSEA$ = .05).

The authors next asked an additional 236 individuals to complete the 30-question survey. The authors repeated a CFA with these data and again found support for an 8-factor model (X^2 = 600, df = 397, CFI = .94, TLI = .93, $SRMR$ = .06, AIC = 17148, $RMSEA$ = .05).

The combined sample of all three studies demonstrated Cronbach Alpha scores of .89 for empowerment (7 items), .81 for accountability (3 items), .76 for standing back (3 items), .91 for humility (5 items), .82 for authenticity (4 items), .69 for courage (2 items), .72 for forgiveness (3 items) and .74 for stewardship (3 items).

Seven of the eight scales from the *Servant Leadership Survey* were correlated in the range of .47 to .85 with the seven scales of the *Servant Leadership Scale (*Liden, Wayne and Henderson 2008*)*. The accountability scale of the *Servant Leadership Survey* was either uncorrelated or correlated at .20 or below for the seven scales of the *Servant Leadership Scale*.

Five of the eight scales were highly correlated with leader-member exchange *LMX-7* scores in the range of .38 to .85. Three of the *Servant Leadership Survey* scales were also highly correlated with the subscales of Rafferty and Griffin’s (2004) measure of transformational leadership. Six of the *Servant Leadership Survey* scales were also highly correlated with the *Brown Ethical Leadership Survey* (2005).

6 Results

Table 9 provides an initial analysis between participant personality scores and servant leadership scores. Scores in bold are significant. Generally, each of the nine dimensions of servant leadership were correlated with at least two dimensions of personality.

Table 9. Abridged Correlation Matrix

	O	C	E	A	N
Empower	**.476****	.122	**.578****	**.303****	**-.292****
Stand Back	.142	-.001	.094	**.425****	-.161
Account	.190*	**.590****	**.197***	.160	**-.375****
Forgive	.055	**.213***	.108	.117	**-.422****
Courage	**.534****	-.035	**.325****	.135	-.139
Humility	.143	**.195***	.047	**.217***	-.075
Authenticity	.087	.026	**.253****	**.246****	.068
Stewardship	**.365****	.112	.170	**.366****	**-.192***

Columns are for Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. **. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).

Table 10 provides the results of an exploratory factor analysis using the principal components method with varimax rotation. The factor analysis identified four factors with Eigenvalues greater than one. Using a factor loading standard of > .50, the first factor seems to be measuring aspects of charismatic leadership. In his 1976 theory of charismatic leadership, House posited that, among other things, charismatic leadership includes extremely high levels of self-confidence coupled with a strong conviction in the moral righteousness of his/her beliefs. This seems to capture much for the first factor called charismatic leadership.

Table 10. Factor Analysis of Personality and Servant Leadership

	CH	SL	C	N
Courage	**.78**	.08	.11	-.33
Empower	**.77**	.32	.29	-.11
Openness	**.73**	.16	-.04	.13
Extroversion	**.68**	.05	.03	.26
Stand Back	.16	**.71**	.14	-.08
Authenticity	.12	**.66**	-.12	-.10
Humility	-.12	**.62**	.24	.16
Stewardship	.38	**.58**	.23	-.02
Agreeableness	.34	**.57**	-.08	.37
Account	.26	.27	**.86**	-.12
Conscientiousness	-.05	.02	**.82**	.26
Forgive	-.07	.02	.06	**.87**
Neuroticism	-.33	.05	-.47	**-.59**

CH- Charisma, SL-Servant Leadership, C – Conscientiousness,
N - Neuroticism

Of the five factors of personality, agreeableness is the most likely to be associated with servant leadership. The second factor, called servant leadership contains agreeableness and four dimensions from the *Servant Leadership Survey (SLS)*.

The third factor was labeled conscientiousness. one of the facets of conscientious is called dutifulness, which is defined as am emphasis on the importance of fulfilling moral obligations. This partially explains why the accountability scale of the SLS loaded with the personality dimension of conscientiousness.

Finally, neuroticism loaded inversely with the SLS scale of forgiveness. One of the facets of neuroticism is called angry hostility and is defined as the tendency to experience anger and related states such as frustration and bitterness. This likely explains why neuroticism loaded inversely with the SLS scale of forgiveness.

7 Summary

After forty years as a philosophy without empirical support, in the last decade instruments have been developed to measure servant leadership. To date, however, none of those instruments seems to have emerged as a dominant choice for researchers. In 2016, Hoch, Bommer, Dulebohn and Wu published a meta-analysis related to servant leadership. While the studies analyzed for 11 different outcome variables ranged from 4 to 11, the fact that a meta-analysis has now been published supports the contention that interest in empirical research related to servant leadership is increasing.

The results of this analysis indicate that standing back, authenticity, humility and stewardship may be the most likely scales to be measuring something different that leader personality. Even, however, within these four scales, two are often measured by other scales. The HEXACO-PI-R is a humility scale that some researchers argue should be considered a sixth dimension of personality (Lee, 2013). Authentic leadership (Avolio, Gardner, & Walumbwa, 2007) has been researched in dozens of peer-reviewed articles (Hoch, Bommer, Dulebohn & Wu 2016).

Given these additional, possible overlaps, perhaps the scales from the SLS that are most unique to servant leadership are the scales that measure standing back and stewardship.

References

Avolio,B.J. *et al.* (2007) *The Authentic Leadership Questionnaire*. Mingarden.com.

Baldasaro,R.E. *et al.* (2013) Psychometric properties of the Mini-IPIP in a large, nationally representative sample of young adults. *Journal of Personality Assessment*, **95**(1), 74-84.

Barbuto,J.E.,Jr, and Wheeler,D.W. (2006) Scale development and construct clarification of servant leadership. *Group & Organization Management*, **31**(3), 300-326.

Barrick,M.R. *et al.* (2001) Personality and performance at the beginning of the new millennium: what do we know and where do we go next? *International Journal of Selection & Assessment* **9**, no. 1-2- 9.

Bono,J.E. and Judge,T.A. (2004) Personality and transformational and transactional leadership: a meta-analysis. *Journal of Applied Psychology*, **89**(5), 901-910.

Cooper,A.J. *et al.* (2010) A confirmatory factor analysis of the Mini-IPIP five-factor model personality scale. *Personality and Individual Differences*, **48**, 688-691.

Donnellan,M.B. *et al.* (2006) The mini-IPIP scales: tiny-yet-effective measures of the Big Five factors of personality. *Psychological Assessment*, **18**, 192-203.

Goldberg,L.R. *et al.* (2006) The International Personality Item Pool and the future of public-domain personality measures. *Journal of Research in Personality*, **40**, 84-96.

Green,M.T. (2015) *Graduate Leadership, 3rd Edition*, Vol. 2, North Charleston, SC, Leadership Press, ISBN-13: 978-0692419168.

Hoch,J.E. *et al.* (2016) Do ethical, authentic, and servant leadership explain variance above and beyond transformational leadership? a meta-analysis. *Journal of Management*, **20**(10), 1-29.

House,R.J. (1977) A 1976 theory of charismatic leadership. In J.G. Hunt and L.L. Larson (eds.), *Leadership: the cutting edge.* Southern Illinois University Press, Carbondale, IL

Judge,T.A. *et al.* (2002). Personality and leadership: a qualitative and quantitative review. *Journal of Applied Psychology*, **87**(4), 765-780.

Lee,K. (2013) *H factor of personality:why some people are manipulative, self-entitled, materialistic, and exploitive' and why it matters for everyone*. Wilfrid Laurier University Press.

Liden,R.C. *et al.* (2008). Servant leadership: development of a multidimensional measure and multi-level assessment. *Leadership Quarterly,* **19**(2), 161.

Page,D. and Wong,P.T.P. (2000) Conceptual framework for measuring servant-leadership, in S. Adjibolosoo, ed., *The Human Factor in Shaping the Course of History and Development*. Lanham, MD: University Press of America.

Politis,J. and Politis,N. (2012). The relationship between servant leadership and personality characteristics: The 'Big Five'. Proceedings of The European Conference On Management, Leadership & Governance, 332-339.

Sendjaya,S. *et al.* (2008). Defining and measuring servant leadership behavior in organizations. *Journal of Management Studies*, **45**(2), 402–424.

Spears,L.C. (1995) (ed.) *Reflections on leadership: How Robert K. Greenleaf's theory of servant-leadership influenced today's top management thinkers.* John Wiley & Sons, Inc., New York.

van Dierendonck,D. and Nuijten,I. (2011) The servant leadership survey: Development and validation of a multidimensional measure. *Journal of Business and Psychology*, **26**(3), 249-267.

Washington,R.R. *et al.* (2006) Individual differences in servant leadership: the roles of values and personality. *Leadership & Organization Development Journal*, **27**(8), 700-716. doi:10.1108/01437730610709309

Journal of Management Science
and Business Intelligence
Vol. 2, No. 2, August 2017
Price: $80

Published and printed by
Institution of Business Intelligence Innovation
P. O. Box 218241, Houston, TX 77218, USA
E-mail: contact@ibii-us.org
Website: http://www.ibii-us.org

ISSN 2472-9256
9 772472 925600

IBII