Effect of Leadership Styles, Social Capital, and Social Entrepreneurship on Organizational Effectiveness of Social Welfare Organization in Malaysia: Data Screening and Preliminary Analysis

Ng Chun Won*, Chong Yen Wan2, Mohmad Yazam Sharif3

1School of Business Management, Universiti Utara Malaysia, Malaysia, 2School of Business Management, Universiti Utara Malaysia, Malaysia, 3School of Business Management, Universiti Utara Malaysia, Malaysia. *Email: wcng.2t@gmail.com

ABSTRACT

The purpose of this paper is to discuss the process of screening, editing and preparation of initial data before any further multivariate analysis of the study concerning effect of leadership styles, social capital and social entrepreneurship on organizational effectiveness of social welfare organization in Malaysia. It is vital to conduct data screening to identify any potential violation of the basic assumptions related to the application of multivariate techniques. Moreover, initial data examination enables the researcher to gain a deeper understanding of the data collected. For this research, simple random sampling will be adopted as the sampling technique to draw conclusion about the entire population. Samples of 159 were selected from the total population of 271 welfare organization in Malaysia. Data screening and preliminary analysis were conducted to meet the assumptions of multivariate analysis. Thus, the study carried out non-response bias test, missing data detection and treatment, multivariate outliers detection and treatment, normality assessment, linearity assessment, common method variance assessment, multicollinearity assessment, homoscedasticity assessment and descriptive analysis. All the assessment was conducted using IBM SPSS statistical software version 21.0 (SPSS). In brief, the data found to fulfil the requirements for further multivariate analysis.

Keywords: Organizational Effectiveness, Social Welfare Organization, Data Screening, Preliminary Analysis

JEL Classifications: D23, L2

1. INTRODUCTION

Data screening implicates certain requirements in the quantitative research process. The first requirement is to meet the assumptions of psychometric property concerning the data, therefore making it safe to proceed to use the data for statistical analyses. Second, is the need to follow certain procedure by checking for errors and correcting the error, if any, in the data file. Failure to do this may result in distorting the following data analysis (Pallant, 2010). To meet these requirements, this study adopted the approach of detection and treatment of missing values, identification of outliers, normality assessment and linearity assessment (Hair et al., 2016). Sarstedt et al. (2012) and Hair et al. (2016) have highlighted on the important of normality test. The researchers claimed that the bootstrapping procedure used in partial least squares (PLS) is prone to standard error where the data is highly skewed or Kurtotic. This view reinforces the fact that multivariate data need rigorous examination in order to overcome the problems of outliers, violation of assumptions and the challenge of missing data which can significantly affect the results of the study (Hair et al., 2016).

The remainder of the paper is organized as follows, literature about leadership styles, social capital, social entrepreneurship and organizational effectiveness of social welfare organization. Next, highlight of the technique used in this research, result and discussion of the findings. Lastly, conclusion was drawn based on the research findings.

2. LITERATURE REVIEW

It is hard to define organizational effectiveness in the social sector (Mowbray, 2004). Organizational effectiveness can define as the
capability of an organization to achieve its objectives and goals efficiently (Selden and Sowa, 2004). Organizations always have more than one goal to achieve, single dimension is not sufficient to assess the effectiveness of the organization effectively (Herman and Renz, 2004; Yacinthe, 2004). Therefore, the effectiveness of social welfare organizations should be constructed in multiple factors (Light, 2008; Niven, 2003; Selden and Sowa, 2004). Mission achievement and financial efficiency are the dimensions to signify organizational effectiveness in this research. Leadership is defined as the action of someone who leads a group of individual to achieve a common goal by the influence and power (Hemphill and Coons, 1957). Fiedler (1969) defined leadership styles as the process by which leader influences a group of people to work together in achieving an established mission.

Past studies had shown that there is a need for leadership in addressing a known driving force to the social welfare organizations (Thompson et al., 2005). Nahapiet and Ghoshal (1998) defined social capital as the summation of the actual and potential resources in the network possessed by asocial unit, whereas Woolcock (1998) give emphasis to the information, trust, and norms inhering in one’s social networks. In this research, social capital will be referred to as a process of developing trusting relationships, mutual understanding and collective actions that connecting individuals, organizations and communities in social sector (Loffeffer et al., 2004). Social entrepreneurship can be explained by the word “entrepreneurship” with the modification of the term “social” (Martin and Osberg, 2007). On the other word, social entrepreneurship is just an expansion of the entrepreneurial approach that applied in business sector (Helm, 2007).

3. METHODOLOGY

Technique of data analysis is a method by which researchers analyse data, and consequently deliver better understanding of the phenomenon (Pallant, 2011). In this study, descriptive statistics was employed to analyse the data. Samples of one hundred and fifty nine were selected from the total population of 271 welfare organization in Malaysia through simple random sampling technique. The data collected was coded and inputted into the Statistical Package for the Social Science (SPSS version 21.0). Formerly the subsequent technique of data analysis was implemented to analyse the data. Firstly, non-response bias test was conducted. Subsequently, this study adopted the approach of detection and treatment of missing values, identification of outliers, normality assessment and linearity assessment for data screening (Hair et al., 2016). Lastly, common method variance (CMV), multicollinearity assessment, homoscedasticity assessment and descriptive analysis were conducted to meet the preliminary assumption for further multivariate analysis.

4. RESULT AND DISCUSSION

4.1. Response Rate
Response rate of survey is significant concern in a study because it ensures the questionnaires collected are valid for data analysis (Hair et al., 2010). Response rate defined by Hamilton (2009) as the percentage of respondents who participated in the survey from the sample size determined for the research. Bartlett et al. (2001, p. 46) recommend to follow Salkind (1997) cautionary step to increase the estimated minimal sample size by 40-50% to account for “lost mail and uncooperative subject” in survey studies. Therefore, by 50% increment, about 239 questionnaires set were distributed to respondents. However, 137 questionnaires were retrieved. Therefore, this makes the response rate of 57.32%, though, out of the 137 collected questionnaires only 134 were found to be useful for further analysis, because 3 questionnaires were excluded from the analysis due to outlier problem. This accounted for 56.07% valid response rate. According to Sekaran and Bougie (2010), response rate of 30% is acceptable for surveys. Hence forward, response rate of this study is adequate for further analysis (Table 1).

4.2. Response Bias Test
The issue of non-response bias arises when there is difference in the answers between non-respondents and respondents (Lambert and Harrington, 1990). Non-response bias can affect the findings of the research and the generalization of the result to the population. Henceforth, there is a need to conduct the non-response bias test to detect this type of error before moving to the main analysis. In regard to the possibility of non-response bias issue, this research followed a time-trend extrapolation method (Armstrong and Overton, 1977) by comparing the early and late respondents. The researchers claimed that late respondents share similar characteristics with non-respondents. Furthermore, to minimize the issue of non-response bias, Lindner and Wingenbach (2002) suggested that a minimum response rate of 50% should be achieved.

Subsequently, an independent samples t-test was conducted for all the study variables to inspect if there is any discrepancy between the two groups. As depicted in Table 2, the results of independent-samples t-test showed that the equal variance significance values for all the variables and the dimensions were >0.05 significance level of Levene’s test for equality of variances (Field, 2009; Pallant, 2011). Henceforth, it can be concluded that the assumption of equal the variances between early and late respondents has not been violated. Additionally, concerning Lindner and Wingenbach’s (2002) recommendation, since the research achieved 56.07% response rate, it can be considered that non-response bias was not a major concern.

4.3. Missing Data Detection and Treatment
The indication of a missing data is when a respondent failed to deliver answer concerning one or more questions thus making
Table 2: Independent samples test

<table>
<thead>
<tr>
<th>Dimensions</th>
<th>Levene’s test for equality of variances</th>
<th>t-test for equality of means</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F</td>
<td>df</td>
</tr>
<tr>
<td>Miss_mean</td>
<td>Equal variances assumed</td>
<td>0.650</td>
</tr>
<tr>
<td></td>
<td>Equal variances not assumed</td>
<td>0.296</td>
</tr>
<tr>
<td>Finan_mean</td>
<td>Equal variances assumed</td>
<td>0.219</td>
</tr>
<tr>
<td></td>
<td>Equal variances not assumed</td>
<td>−0.176</td>
</tr>
<tr>
<td>TransF_mean</td>
<td>Equal variances assumed</td>
<td>2.293</td>
</tr>
<tr>
<td></td>
<td>Equal variances not assumed</td>
<td>−1.257</td>
</tr>
<tr>
<td>Transac_mean</td>
<td>Equal variances assumed</td>
<td>1.243</td>
</tr>
<tr>
<td></td>
<td>Equal variances not assumed</td>
<td>1.161</td>
</tr>
<tr>
<td>STrust_mean</td>
<td>Equal variances assumed</td>
<td>0.182</td>
</tr>
<tr>
<td></td>
<td>Equal variances not assumed</td>
<td>0.107</td>
</tr>
<tr>
<td>PubSE_mean</td>
<td>Equal variances assumed</td>
<td>1.559</td>
</tr>
<tr>
<td></td>
<td>Equal variances not assumed</td>
<td>−0.222</td>
</tr>
<tr>
<td>Innov_mean</td>
<td>Equal variances assumed</td>
<td>0.018</td>
</tr>
<tr>
<td></td>
<td>Equal variances not assumed</td>
<td>1.108</td>
</tr>
<tr>
<td>Procac_mean</td>
<td>Equal variances assumed</td>
<td>2.038</td>
</tr>
<tr>
<td></td>
<td>Equal variances not assumed</td>
<td>−0.082</td>
</tr>
<tr>
<td>Risk_mean</td>
<td>Equal variances assumed</td>
<td>0.182</td>
</tr>
<tr>
<td></td>
<td>Equal variances not assumed</td>
<td>−1.143</td>
</tr>
<tr>
<td>Miss_mean</td>
<td>Equal variances assumed</td>
<td>2.156</td>
</tr>
<tr>
<td></td>
<td>Equal variances not assumed</td>
<td>−0.146</td>
</tr>
</tbody>
</table>

4.4. Multivariate Outliers Detection and Treatment

Outliers are extreme scores or values of data sets that may significant affects on the analysis and the result of the study (Hair et al., 2010). The presence of outlier may due to discrepancy in the measurement and can possibly show an experimental error (Churchill and Iacobucci, 2004). Presence of outliers in the data set can utterly distort the following data analysis and lead to erroneous results (Verardi and Croux, 2008). Mahalanobis distance (d2) was employed to detect the outliers. Mahalanobis distance (d2) defined by Tabachnick and Fidell (2007) as “the distance of a case from the centroid of the remaining cases where the centroid is the point created at the intersection of the means of all the variables” (p. 74).

With degree of freedom equating the number of items (61 items), the Chi-square value is 100.8878 (P = 0.001). Mahalanobis distance values that exceeded the Chi-square value were deleted (Tabachnick and Fidell, 2007). Following this criterion, three multivariate outliers (respondent 28 = 105.0353, respondent 88 = 110.2931, respondent 96 = 111.1706) were identified and deleted from the dataset because they could affect distort the result of the data analysis. Henceforth, after removing three multivariate outliers, the final dataset in this study was 134.

4.5. Normality Assessment

Though the preliminary assumption common in research analyses that are using the structural equation modelling (PLS-SEM) is to emphasize that since PLS grants accurate model estimation even that are using the structural equation modelling (PLS-SEM) is to emphasize that since PLS grants accurate model estimation even with the presence of extreme non-normal data (Henseler, 2012). However, Hair et al. (2012) and Hair et al. (2014) have highlighted on the importance of normality test. The researchers claimed that the bootstrapping procedure used in PLS is prone to standard error where the data is highly skewed or Kurtotic. Normality denotes to
the shape of the distribution of data for individual metric variable and its correspondence to the normal distribution of the benchmark for statistical methods (Hair et al., 2010).

In this study, the assumption of normality was inspected using two methods. The first method was examined by looking at the shape of data distribution graphically (Tabachnick and Fidell, 2007) and second method by evaluating the skewness and kurtosis value (Garson, 2012). The data collected in the present study have followed the normal pattern since all the bars on the histogram were close to a normal curve. Therefore, normality assumptions were not violated in the present study.

According to Garson (2012), the accepted range of absolute value of skewness and kurtosis is ±2. The values of both skewness and kurtosis in this study all fall within the range. Skewness is within the range of −0.773–0.049 while kurtosis is within the range of −1.137–1.081 as shown in Table 3. It means that the normality assumptions in this study were not violated.

### 4.6. Linearity Assessment

Linearity of relationship as an assumption in multiple regressions was used to denote the degree to which the change in the dependent variable was related with the independent variable (Hair et al., 2010; Tabachnick and Fidell, 2014). As multiple regression models were based on the linearity of multivariate relationships, the linearity assumption was necessary (Hair et al., 2010; Tabachnick and Fidell, 2014). The linearity test was conducted through the graph-legacy diagrams-scatter/dot-simple scatter procedures in SPSS 22. Linearity of data could be tested by examination of scatter plots or linearity residual plot (Hair et al., 2010; Pallant, 2013).

Visual examination of the plots in this study showed a roughly straight line and not a curve. This meant that the residuals had a straight-line relationship with the predicted values of the dependent variable (organizational effectiveness). Hence, there were a linearity of the relationship between the dependent variable of organizational effectiveness and the independent variables of leadership style, social capital and social entrepreneurship from each of the scatter plots. The data by this means satisfied the linearity assumption of multiple regressions.

#### 4.7. CMV Assessment

CMV can be defined as variance that is perpetually attributable to the measurement procedure rather than to the actual constructs the measures represent (Podsakoff et al., 2003). Common variance method is basically that of a measurement issue rather than constructs involved in the study. It is of interest due to its potential of bias when estimating the relationship among the theoretical constructs of the research (Podsakoff et al., 2003). Such errors may cause by social desirability or having a common rater; items ambiguity or item characteristics effects; the effects of grouping items or items context effects and measurement effects which happen through simultaneous measurement of predictor and criterion variables (Meade et al., 2007).

Some procedural and statistical controls were adopted to deal with the issue of CMV (Podsakoff et al., 2003; Podsakoff et al., 2012; Podsakoff and Organ, 1986; Viswanathan and Kayande, 2012).

The first step is the procedure in which the questionnaire design was subject to expect evaluation. These expect were selected through objective basis to avoid social effects. Secondly, the respondents were given an assurance that the research is meant for academic purposes; and that their responses are not about being right or wrong; and their responses are confidential. Efforts were also made to improve the scale items. This was achieved by avoiding vague concepts in the questionnaire and survey were written in a simple, specific and concise language.

Besides the procedural and statistical controls described above, the present study also adopted one of the most widely statistical approaches, Harman’s single factor test to inspect CMV (Podsakoff and Organ, 1986). The main assumption of Harman’s single factor test is that if a substantial amount of CMV is present, either a single factor may emerge, or one general factor would account for most of the covariance in the predictor and criterion variables (Podsakoff et al., 2003). This data does not have the problem of CMV serious enough to inflate relationships between the variables as the first (largest) factor accounting for 21.501% of the variance which is <50% (Kumar, 2011).

#### 4.8. Multicollinearity Assessment

Multicollinearity refers to a situation in which one independent variable is actually a combination of the other variables or when the independent variables are highly correlated (Hair et al., 2010; Pallant, 2010; Tabachnick and Fidell, 2013). The occurrence of multicollinearity among the exogenous latent constructs can potentially affect the estimates of regression coefficients and the statistical significance tests (Chatterjee and Yilmaz, 1992; Hair et al., 2006). Specifically, multicollinearity upturns the standard errors of the coefficients, which leads to decrease in the predictive power of the independent variables on the dependent variables (Tabachnick and Fidell, 2007). This is due to the reason that the variables cancel out each other (Hayes, 2005).

Two approaches were employed to examine multicollinearity in this study (Chatterjee and Yilmaz, 1992; Peng and Lai, 2012)
First, the correlation matrix of the exogenous latent constructs was assessed. Statistically, a correlation coefficient of 0.90 and above indicates the presence of multicollinearity between exogenous latent constructs (Pallant, 2010; Hair et al., 2010). Secondly, Tolerance and variance inflation factor (VIF) were examined to identify multicollinearity issue. Hair et al. (2011) recommended that multicollinearity is a concern if VIF value is higher than 5 and tolerance value is <0.20. Table 4 indicates that multicollinearity did not exist among the exogenous latent constructs as all VIF values were <5 and tolerance values exceeded 0.20 as suggested by Hair et al. (2011). Thus, multicollinearity is not an issue in the present study. Table 4 shows the VIF values, tolerance values and correlation matrix of all exogenous latent constructs.

### 4.9. Homoscedasticity Assessment

Assumption of homoscedasticity is also a concern to the researchers as how the values of the data are being spread out among the variables is very crucial in a study. Pallant (2010) defined homoscedasticity as the variance of the residuals about predicted DV scores should be the same for all predicted scores. If the assumption of homoscedasticity is unmet, the data is not appropriate for conducting a test of differences like ANOVA. In the present study, scatter plot will be used to test the homoscedasticity. It is expected to display a fairly even cigar shape along its length (Pallant, 2010). The data for this study met the assumptions of homoscedasticity as all the scatter plots show a cigar shape demonstrating both linearity in relationship between the variables and even spread of data for the study.

### 5. CONCLUSION

To conclude, this paper recapitulated the process of screening, editing and preparation of initial data before any further multivariate analysis. Thus, the study carried out non-response bias test, missing data detection and treatment, multivariate outlier detection and treatment, normality assessment, linearity assessment, CMV assessment, multicollinearity assessment, homoscedasticity assessment and descriptive analysis. All the assessment was conducted using IBM SPSS statistical software version 21.0 (SPSS). In brief, the data found to fulfill the multivariate assumptions.

### REFERENCES


