

Linking Organisational Culture, Leadership Styles, Human Resource Management Practices and Organisational Performance: Data Screening and Preliminary Analysis

Amer Ahmad AlAnazi
Universiti Utara Malaysia

Faridahwati Mohd. Shamsudin
Universiti Utara Malaysia

Johanim Johari
Universiti Utara Malaysia

This paper explored the data collected regarding the study linking organisational culture, leadership styles, human resource management practices and organisational performance. Two hundred and fifty six Saudi Arabia Private firms participated in the study. The data was analyzed using Statistical Package for the Social Sciences (SPSS) application package version 22. An initial data screening and preliminary analysis was performed in order to ensure that the assumptions of multivariate analysis have not been violated. Specifically, the assessment of missing values, outliers, normality test and multicollinearity test were performed in the present study. We concluded that the data was fit for further multivariate analysis.

INTRODUCTION

The importance of the initial data screening in multivariate data analysis technique cannot be over-emphasized because the validity of inferences drawn from statistical test results accuracy of the of results largely depend on the whether the key assumptions of multivariate analysis have been violated or not. Although the validity of inferences drawn from statistical test results depends on how well data meet the key assumptions of multivariate analysis, however, literature indicates that many research have been carried our without reporting whether such assumptions are violated or not (Hoekstra, Kiers, & Johnson, 2012). Study also suggests that initial data screening in multivariate data analysis technique is intended to identify and correct some methodological errors or at least to minimize their impact on the overall study results (Van den Broeck, Cunningham, Eeckels, & Herbst, 2005).

The purpose of this paper was to explore the data collected regarding the study linking organisational culture, leadership styles, human resource management practices and organisational performance. Specifically, our goal was to confirm that the key assumptions of multivariate analysis have been violated before conducting the main analyses for the study. Toward this end, the remainder of this paper is organized as follows. In section 2, we the research method employed for this study was highlighted. This is followed by presentation of the and discussion of results in section 3, which are based on assessment of

missing values, assessment of outliers, normality test, as well as multicollinearity test. Finally, based on the results, conclusion was drawn In the section 4.

METHODS

Sample and Procedure

The sample for this study was conducted among private firms in Saudi Arabia. Based on the determined sample size of 346, the present study distributed 400 questionnaires to Chief Executive Officers (CEOs)/Directors of private organizations in Saudi Arabia. To achieve higher response rate in the present study, several attempts have been made, including phone call reminders, Short Message Services (SMS) to those respondents who were yet to complete their survey after one month from the date it was sent to them (Dillman, 2000; Sekaran & Bougie, 2010). Hence, these efforts yielded 302 returned questionnaires, out of 400 questionnaires that were distributed to the target participants. This gives a response rate of 76%. Of these 302 questionnaires, 32 were invalid and considered unusable as significant part of those questionnaires were not completed by the participants; and the remaining 270 useable questionnaires were used for the analyses. This accounted for 68% valid response rate. Therefore, the valid response rate of 68%, based on Al-Marri, Ahmed, and Zairi (2007) satisfied all the statistical requirements and proved to be very useful for purpose of testing hypothesized relationships.

Measures

Human Resource Management Practices

We used 10 items taken from the works of (Arthur, 1994) and MacDufie (1995) to measure the two dimensions of HRM practices, namely, Commitment-based HRM practices, as well as performance - based HRM practices. Specifically, for each dimension, five items were adapted. However, for the sake of parsimony, HRM practices have been conceptualized as second order the construct. Ratings were completed on a seven-point Likert-type scale ranging from 1 (*strongly disagree*) to 7 (*strongly agree*). Sample items include: "Our organization offers formal training and development of individual skills". "In our organization, employees work strictly within the assigned tasks". "In our organization, managers are the ones who make the decisions".

Nitaqat Status

This measure developed for this study aims to determine the current status of the respondent's company within the general Nitaqat scheme. The section consists of two items. The first item directly asks the respondent about the Nitaqat status of the company. The choices of answers are: "Red", "Yellow", "Green", "Platinum", and "I Do Not Know". The second item asks the respondent to indicate an approximate percentage of Saudis employed. The second item serves as a confirmation of the current Nitaqat status of the company. The choices of answers are: "Less than 6%" (corresponds to Red Nitaqat status), "7-11%" (corresponds to Yellow Nitaqat status), "12-39%" (corresponds to Green Nitaqat status), "Over 40%" (corresponds to Platinum Nitaqat status), and "Not Sure."

Organizational Culture

Organizational culture was assessed using an Organizational Culture Assessment Instrument (OCAI), which reflects four dimensions of organizational culture (hierarchy culture, market culture, clan culture, and adhocracy culture), will be used to measure the kind of organizational culture that prevails in the respondent's organization. More specifically, twenty items were adapted from OCAI to measure aspects of organizational culture. Hierarchy culture was measured with five items (e.g., "The organization emphasizes permanence and stability. Efficiency, control and smooth operations are important" and "Loyalty and truth hold our organization together is loyalty and mutual trust, commitment to the organization runs high."). Market culture was measured with five items (e.g., "The management style in the organization is characterized by individual risk-taking, innovation, freedom and uniqueness" and "The management style in the organization is characterized by hard-driving competitiveness, high demands and

achievement"). Clan culture was measured with five items (e.g., "The management style in the organization is characterized by security of employment, conformity, predictability and stability in relationships" and "The glue that holds the organization together is the emphasis on achievement and goal accomplishment"). Finally, adhocracy culture was measured with five items (e.g., "The organization defines success on the basis of having the most unique or the newest products. It is a product leader and innovator" and "The management style in the organization is characterized by teamwork, consensus and participation"). For each dimension, participants will be asked to state the extent to which they agreed with each statement by using a 7-point Likert type scale (1 = strongly disagree, 7 = strongly agree).

Leadership Style

Leadership style was measured using the Leadership Behavior Questionnaire (LBQ IV) developed by Pearce (1997) and validated by Pearce and Sims (2002). The original questionnaire includes four key dimensions of leadership: Directional, Transactional, Transformational, and Empowering. Pearce and Sims showed high internal consistency of the items (Cronbach's alpha over .9), and the subsequent empirical studies (e.g., Wood, 2005) demonstrated high levels of reliability as well. In total, the section includes 12 items corresponding to the four leadership styles considered in this research. For each style, items from Pearce and Sims (2002) will be selected. In addition, a general question regarding leadership styles as defined by Pearce and Sims (2002) is provided. It defines each leadership style and asks the respondent to select the one which matches the leadership style of their organization most.

Organizational Performance

Organizational performance was assessed using a twelve-item based on the work of Delaney and Huselid (1996). Sample items include: "How would you compare your organization's performance over the past 3 years in terms of quality of products, service, and programs?" "Compared to other organizations that do the same kind of work, how would you compare your organization's performance over the last 3 years in terms of marketing?" Ratings were completed on a seven-point scale ranging from 1 (*substantially worse*) to 7 (*substantially better*).

RESULTS AND DISCUSSION

Prior to conducting the main analyses, the data collected were subjected to preliminary analyses in terms of missing values, statistical outliers, normality, linearity, homoscedasticity, and multicollinearity. This was to ensure that the statistical assumptions necessary for multivariate analysis were satisfied. Results pertaining to each of the four key assumptions are reported below.

Missing Values

Missing value "is one of the most pervasive problems in data analysis... its seriousness depends on the pattern of missing data, how much is missing, and why it is missing" (Tabachnick & Fidell, 2007, p. 62). It is statistically important to check for missing values before conducting any analytic procedures because some statistical packages (e.g., SmartPLS) will not work even with a single data missing. Furthermore, overlooking cases with missing values could lead to the loss of vital information, which subsequently minimizes the statistical power and increases standard errors (Peng, Harwell, Liou, & Ehman, 2006). While there is no universally acceptable cut-off in the literature regarding the percentage of missing value in a dataset for valid statistical analysis (Schafer, 1999), for instance, argued that a missing rate of 5% or less is immaterial. On the other hand, Bennett (2001) asserted that statistical analysis is likely to be valid when the amount of missing value is 10% or less. To determine if there was any value missing in the dataset, the Statistical Package for the Social Sciences (SPSS) was used. The results of the missing values analysis are provided in Table 1.

TABLE 1
RESULT OF REPLACED MISSING VALUES

No	Variable	Number of replaced missing values
1	HRC04_1	1
2	HRC05_1	1
3	Attitudes_1	12
Total number of replaced missing values		14 out of 19,170 data points
Percentage of replaced missing values		.07%

Note. Percentage of missing values was obtained by dividing the total number of randomly missing values for the entire data set by total number of data points multiplied by 100.

As shown in Table 1, of the 19,170 data points in the SPSS dataset, 14 were randomly missed, thereby accounting for .07%. In particular, commitment-focused HR practices and performance-focused HR practices had 1 missing value each. Finally, attitudes had 12 missing values. Despite the fact that .07% missing value in a dataset is still valid for a meaningful statistical analysis (Schafer, 1999), it was decided that the missing value be replaced using mean substitution (Tabachnick & Fidell, 2007). Need to explain what mean substitution is.

Statistical Outliers

According to Tabachnick and Fidell (2007), a statistical outlier is “a case with such an extreme value on one variable (a univariate outlier) or such a strange combination of scores on two or more variables (multivariate outlier) that it distorts statistics” (p. 72), and thus leading to Type I error and/or Type II error. In this study, two forms of statistical outliers were examined, namely: univariate outliers and multivariate outliers. The former were examined based on standardized values. According to Tabachnick and Fidell (2007) a case with standardized values of ± 3.29 ($p < .001$) or more is considered to univariate outliers. In line with this threshold, it was found that none of the case was identified to have standardized values above the threshold of ± 3.29 ($p < .001$) or more. Hence, no potential univariate outliers were detected. Furthermore, to be sure that that assumption of statistical outliers has not been violated in the present study, multivariate outliers were assessed using Mahalanobis distance (D^2). According to Tabachnick and Fidell (2007) Mahalanobis distance (D^2) refers to 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). The results for the assessment of multivariate outliers are presented in Table 2.

As shown in Table 2, the assessment of multivariate outliers was based on 65 observed variables, with recommended threshold of chi-square is 104.72 ($p = 0.001$). According to Hair, Black, Babin, and Anderson (2010) a data case with Mahalanobis desistance value greater than the chi-square value is considered to have multivariate outlier. Following Hair et al. (2010), fourteen multivariate outliers were detected and subsequently deleted from the dataset (Table 4). Hence, after deleting fourteen multivariate outliers, the final dataset used for the main analyses was 256.

TABLE 2
SUMMARY OF CASES WITH MULTIVARIATE OUTLIERS

Respondents ID	Mahalanobis distance (D^2)
4	118.07185
67	119.13717
161	116.75945
176	126.26486
190	128.29393
196	113.53889
198	126.4059
199	143.90992
210	110.84645
213	121.42591
226	111.67267
242	127.76731
244	111.22221
247	141.40957

Note. N = 65; df = 64; $X^2 = 104.72$; p = .001; $D^2 = \geq X^2$

Normality

Even though PLS-SEM does not rely on the restrictive assumption of multivariate normal distribution like covariance-based structural equation modeling (Hair et al., 2010; Hair, Ringle, & Sarstedt, 2011), checking the behavior of data collected before conducting the main analyses is very crucial. This is because conducting PLS-SEM analysis with non-normal data can inflate bootstrap standard errors, and, thus, underestimate the structural model relationships(Chernick, 2008; Hair, Hult, Ringle, & Sarstedt, 2014). To confirm that the normality assumption has not been violated, first, the skewness and kurtosis statistics were checked. According to Kline (2011), the assumption of multivariate normal distribution becomes a major concern when the skewness and kurtosis statistics are greater than ± 3 and ± 10 , respectively. Table 3 summarizes the descriptive statistics of the normality test.

TABLE 3
DESCRIPTIVE STATISTICS FOR NORMALITY TEST

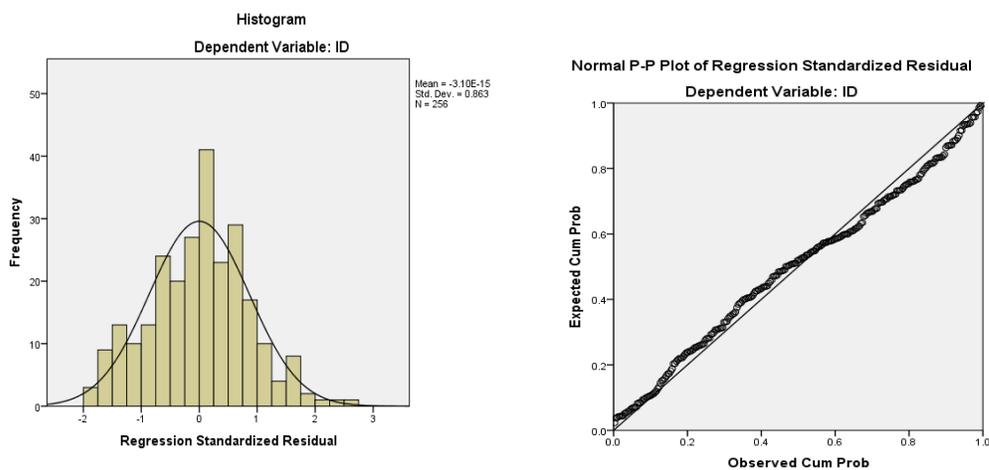
	N	Mean	SD	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	SE	Statistic	SE
OP01	256	5.05	1.586	-.953	.152	.584	.303
OP02	256	5.64	1.269	-.804	.152	.241	.303
OP03	256	5.65	1.253	-.465	.152	-.850	.303
OP04	256	5.33	1.467	-.835	.152	.456	.303
OP05	256	5.08	1.464	-.718	.152	.446	.303
OP16	256	4.22	1.806	-.480	.152	-.696	.303
OP07	256	4.49	1.650	-.516	.152	-.286	.303

OP08	256	4.97	1.406	-.832	.152	.787	.303
OP09	256	5.05	1.442	-.667	.152	.341	.303
OP10	256	4.98	1.393	-.714	.152	.844	.303
OP11	256	5.02	1.357	-.899	.152	1.056	.303
OP12	256	4.60	1.488	-.371	.152	-.211	.303
NS01	256	5.04	1.395	-.767	.152	.643	.303
NS02	256	4.97	1.269	-.725	.152	.822	.303
CC01	256	5.27	1.117	-.310	.152	.000	.303
CC02	256	5.40	1.119	-.539	.152	.196	.303
CC03	256	5.26	1.065	-.295	.152	-.127	.303
CC04	256	5.16	1.070	-.102	.152	-.462	.303
CC05	256	5.16	.995	-.254	.152	.146	.303
CH01	256	5.12	1.039	-.139	.152	-.397	.303
CH02	256	4.91	1.219	-.233	.152	-.155	.303
CH03	256	5.06	1.110	-.377	.152	.116	.303
CH04	256	5.37	1.009	-.279	.152	-.267	.303
CH05	256	5.30	1.047	-.557	.152	.963	.303
CM01	256	5.10	1.038	-.048	.152	-.325	.303
CM02	256	5.27	1.029	-.334	.152	-.004	.303
CM03	256	5.16	1.109	-.504	.152	.303	.303
CM04	256	5.11	1.036	-.250	.152	-.151	.303
CM05	256	5.18	1.018	-.330	.152	.535	.303
CA01	256	5.18	1.013	-.390	.152	.504	.303
CA02	256	5.34	1.062	-.465	.152	.535	.303
CA03	256	4.96	.996	-.380	.152	1.017	.303
CA04	256	4.97	.978	-.327	.152	.151	.303
CA05	256	4.75	1.218	-.276	.152	.090	.303
DL01	256	5.76	1.504	-1.518	.152	2.054	.303
DL02	256	5.79	1.461	-1.500	.152	2.214	.303
DL03	256	5.43	1.794	-1.252	.152	.620	.303
TSL01	256	5.21	1.853	-.865	.152	-.303	.303
TSL02	256	5.42	1.926	-1.085	.152	-.087	.303
TSL03	256	5.54	1.701	-1.205	.152	.712	.303
TFL01	256	4.05	1.971	.031	.152	-1.217	.303
TFL02	256	3.77	1.899	.023	.152	-1.048	.303
TFL03	256	4.04	2.007	-.145	.152	-1.153	.303
EMP01	256	4.24	2.175	-.252	.152	-1.382	.303
EMP02	256	5.23	1.801	-.931	.152	-.025	.303
EMP03	256	5.18	1.894	-.810	.152	-.571	.303
HRC01	256	4.55	1.383	-.751	.152	.425	.303
HRC02	256	4.72	1.301	-.686	.152	.502	.303

HRC03	256	4.85	1.163	-.182	.152	-.206	.303
HRC04	256	4.97	1.269	-.356	.152	-.039	.303
HRC05	256	5.13	1.401	-.776	.152	.388	.303
HRP01	256	4.95	1.467	-.438	.152	.468	.303
HRP02	256	5.40	1.323	-1.073	.152	1.620	.303
HRP03	256	5.22	1.359	-.709	.152	.620	.303
HRP04	256	5.52	1.244	-1.198	.152	2.331	.303
HRP05	256	5.77	1.190	-1.218	.152	2.161	.303
LGCY01	256	2.85	1.632	.768	.152	.056	.303
LGCY02	256	2.89	1.689	.735	.152	-.293	.303
LGCY03	256	2.98	1.819	.795	.152	-.287	.303
LGCY04	256	3.30	1.839	.692	.152	-.448	.303
LGCY05	256	3.13	1.757	.789	.152	-.274	.303
LGCY06	256	2.98	1.823	.849	.152	-.366	.303
LGCY07	256	2.43	1.626	1.307	.152	1.232	.303
LGCY08	256	2.76	1.717	.882	.152	.040	.303
Attitudes	256	4.26	1.336	-.217	.152	-.239	.303

As shown in Table 3, there was no evidence of violation of the assumption of multivariate normal distribution since the skewness and kurtosis statistics were all below the thresholds of ± 3 and ± 10 , respectively. To further verify whether the data collected follows the normal distribution curve, both histogram and the normal probability plot (P-P Plots) of the regression standardized residual were considered in this study.

**FIGURE 1
NORMAL PROBABILITY PLOTS**

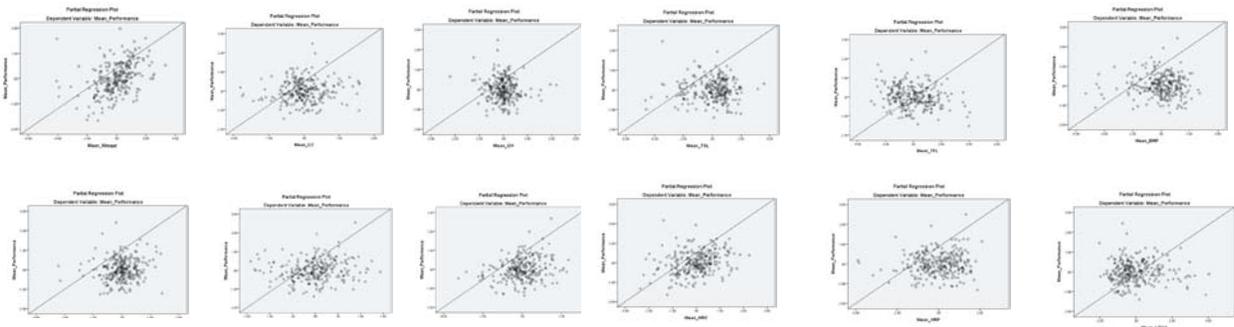


As demonstrated in Figure 1 and Figure 2, the data collected conformed to normal distribution curve. Hence, it can be concluded that the assumption of multivariate normal distribution was not violated.

Linearity

Linearity assumption states that the relationships between predictor variables and dependent variables are linear in nature (Casson & Farmer, 2014). Because PLS path modeling belongs to the family of regression analysis, it is important to confirm if the linearity assumption is met, so as to avoid underestimation of the relationship between the independent variables and dependent variables. Linearity assumption is usually confirmed through partial regression plot between each predictor variable and the predicted variable (Hair et al., 2010). Furthermore, most of the residuals should be scattered around zero point, as well as having a straight-line relationship with predicted dependent variable scores (Pallant, 2010). The results of linearity test (Figure 2) showed that the relationships between the independent and the dependent variable were linear because most of the residuals depicted straight-line relationships with predicted dependent variable scores, and also scattered around zero point. Thus, it can be concluded that the linearity assumption had been verified before conducting the main analysis.

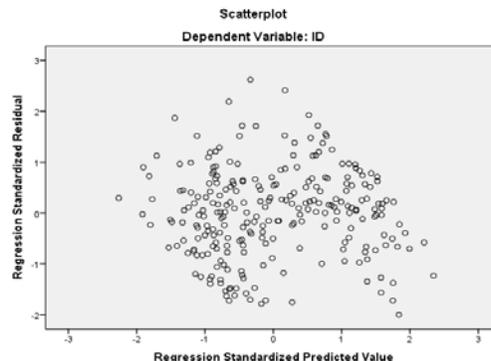
**FIGURE 2
PARTIAL REGRESSION PLOTS**



Homoscedasticity

Homoscedasticity (i.e. equality of variances) is defined as a situation where the variance on a criterion variable appears to be constant over a range of predictor variables (Hair et al., 2010). It is also imperative to confirm whether the assumption of homoscedasticity has been satisfied before undertaking the main analysis because when the variance on a criterion differs over a range of independent variables, heteroscedasticity becomes a major concern. Hence, heteroscedasticity can seriously distort findings of the study, thereby increasing the chance of committing Type 1 error (Hair et al., 2010). Similar to linearity, homoscedasticity assumption was and tested by analyzing a residual plot, and that scattered around zero point. As depicted in Figure 3, the assumption of homoscedasticity was not violated since the residual plots mostly scattered around zero point.

**FIGURE 3
SCATTER PLOT OF THE RESIDUALS**



Multicollinearity

In order to confirm whether the assumption of multicollinearity had been satisfied, tolerance and variance inflation factor (VIF) values were checked. Statistically, tolerance value is defined as 1 minus the proportion of the variance that is explained (R^2 value) for the regression of one independent variable on all remaining independent variables (Allison, 1999). On the other hand, VIF refers to the reciprocal of the tolerance ($1/1-R^2$) and it indicates the magnitude of inflation in the estimated regression coefficients by the existence of correlation among the predictor variables in the model (Allison, 1999; Jani, 2014). According to Hair et al. (2010), tolerance values above .10 and VIF values less than 10 indicate no multicollinearity among the independent variables. Table 6 summarizes the results of multicollinearity test among independent variables. Table 4 shows that the tolerance values ranged from .296 to .865, and VIF values ranged from 1.156 to 3.374, thereby suggesting that multicollinearity was not a major concern in the present study.

TABLE 4
RESULTS OF MULTICOLLINEARITY TEST

Exogenous Latent Variables	Collinearity Statistics	
	Tolerance	VIF
Nitaqat status	.791	1.264
Clan culture	.365	2.736
Hierarchy culture	.325	3.076
Market culture	.296	3.374
Adhocracy culture	.376	2.661
Directional leadership	.540	1.850
Transactional leadership	.641	1.561
Transformational leadership	.745	1.343
Empowering leadership	.677	1.478
Commitment focused HR practices	.805	1.243
Performance focused HR practices	.865	1.156
Legitimacy	.678	1.475

CONCLUSION

In spite of the fact that initial data screening and preliminary analyses add to the credibility of multivariate technique, literature indicate that many research that have been carried out in management related discipline did not report whether the key assumptions of multivariate analyses have been violated or not. Having performed data screening and preliminary analyses, we found that data collected for this study has not violated multivariate assumptions, including assessment of missing values, assessment of outliers, normality test, as well as multicollinearity test. Hence, we concluded that the data was fit and suitable for further multivariate analyses.

REFERENCES

- Al-Marri, K., Ahmed, A. M. M. B., & Zairi, M. (2007). Excellence in service: an empirical study of the UAE banking sector. *International Journal of Quality & Reliability Management*, 24(2), 164-176.
- Allison, P. D. (1999). *Multiple regression: A primer*. Thousand Oaks, CA: Pine Forge Press.

- Arthur, J. B. (1994). Effects of human resource systems on manufacturing performance and turnover. *Academy of Management Journal*, 37, 670-687. doi: 10.2307/256705
- Bennett, D. A. (2001). How can I deal with missing data in my study? *Australian and New Zealand Journal of Public Health*, 25(5), 464-469. doi: 10.1111/j.1467-842X.2001.tb00294.x
- Casson, R. J., & Farmer, L. D. M. (2014). Understanding and checking the assumptions of linear regression: a primer for medical researchers. *Clinical & Experimental Ophthalmology*, 42, 590-596. doi: 10.1111/ceo.12358
- Chernick, M. R. (2008). *Bootstrap methods. A guide for practitioners and researchers* (2nd ed.). Hoboken, New Jersey: Wiley.
- Delaney, J. T., & Huselid, M. A. (1996). The impact of human resource management practices on perceptions of organizational performance. *Academy of Management Journal*, 39, 949-969. doi: 10.2307/256718
- Dillman, D. A. (2000). *Mail and Internet surveys: The tailored design method* (2nd ed.). New York: John Wiley & Sons, Inc.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis* (7th ed.). Upper Saddle River, New Jersey: Prentice Hall.
- Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2014). *A primer on partial least squares structural equation modeling (PLS-SEM)*. Thousand Oaks: Sage Publications.
- Hair, J. F., Ringle, C. M., & Sarstedt, M. (2011). The Use of Partial Least Squares (PLS) to Address Marketing Management Topics: From the Special Issue Guest Editors. *Journal of Marketing Theory and Practice*, 18(2), 135-138.
- Hoekstra, R., Kiers, H. A., & Johnson, A. (2012). Are assumptions of well-known statistical techniques checked, and why (not)? *Frontiers in psychology*, 3, 1-9. doi: 10.3389/fpsyg.2012.00137
- Jani, P. N. (2014). *Business statistics: Theory and applications*. Delhi: PHI Learning Private Limited.
- Kline, R. B. (2011). *Principles and practice of structural equation modeling*. New York, NY: Guilford Press.
- MacDuffie, J. (1995). Human resource bundles and manufacturing performance: Organizational logic and flexible production systems in the world auto industry. *Industrial and Labor Relations Review*, 48, 197-221.
- Pallant, J. (2010). *SPSS survival manual: A step by step guide to data analysis using SPSS* (4th ed.). New York, NY: Open University Press.
- Peng, C.-Y. J., Harwell, M., Liou, S.-M., & Ehman, L. H. (2006). Advances in missing data methods and implications for educational research. In S. S. Sawilowsky (Ed.), *Real data analysis* (pp. 31-78). Charlotte, North Carolina: Information Age Publishing.
- Schafer, J. L. (1999). Multiple imputation: a primer. *Statistical Methods in Medical Research*, 8(1), 3-15. doi: 10.1177/096228029900800102
- Sekaran, U., & Bougie, R. (2010). *Research methods for business: A skill building approach* (5th ed.). New Jersey: John Wiley and Sons.
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston, MA: Allyn & Bacon/Pearson Education.
- Van den Broeck, J., Cunningham, S. A., Eeckels, R., & Herbst, K. (2005). Data cleaning: detecting, diagnosing, and editing data abnormalities. *PLoS Med*, 2(10), e267.