The Likert scale analysis using parametric based Structural Equation Modeling (SEM)

Zainudin Awang*[1], Asyraf Afthanorhan[2], Mustafa Mamat[3]
[1,2,3] Faculty of Economics and Management Sciences, Universiti Sultan Zainal Abidin, Kampus Gong Badak, 21300 Kuala Terengganu, Malaysia.

Abstract

The Likert scale is commonly used in survey research using primary and secondary data to measure the respondent attitude by asking insofar to which they agree or disagree with a particular questions. In generals, Likert scale would be preferred in the questionnaire development stage to ascertain the researchers conducting their research needed. However, the researchers nowadays are abuse to understand the nature of measurement scale in data analysis and thus causes the finding obtained are meaningless. This article is aimed to compare the performance of two categories of measurement scales which are 5 point and 10 points of Likert scales using the same sample size and research subject that would pave the way to understand the real different between both of these ranges using Structural Equation Modeling (SEM). Moreover, this study also interested to clarify briefly between two types of measurement scale namely ordinal and interval data. The findings reveal that 10 points of Likert scale is more efficient than 5 points of Likert scale in operating of measurement model.

Keywords: *Likert Scale, Structural Equation Modeling (SEM), Ordinal, Interval, Parametric*

1. Introduction

The Likert scale is commonly used in survey research especially from social science, management, marketing, education, tourism, healthcare and other disciplines to measure the respondents attitude by asking insofar to which they agree or disagree with a particular question or statement presented. A typical scale that frequently apply by majority of researchers might be “strongly agree, somewhat agree, not sure/undecided, somewhat disagree, and strongly disagree. At the outset, survey data using Likert scale may seem easy to analyze or to identify the factors involve in the study, but there are another important issues that should be addressed for a data analyst to consider it. This is because the implementation of the Likert scale in analysis has become one of the main interested technique for each researchers and scholars lately.

However, most of the researchers from various field abuse the true nature of the Likert scale in the questionnaire development stage. Therefore, their findings would be troublesome or probability to derive the true value is risky due to the requirement of the scale needed is limited. Thus, the main problem of this research is to let the scholars understand the real strength of Likert scale using structural equation modeling on 5 points and 10 points of Likert scale using the same quantity of data sets and model. Generally, the researchers prefer to choose the short ones than the long scales as they believe the result obtained would not be affected. Consequently, the 5 points of Likert scale was frequently implemented if it is related with the survey research. Additionally, most of the researchers believe the label for each measurement scale is required in developing the questionnaire. In fact, their aim of the study was more prone towards the parametric technique as it is only being rationale for interval and ratio scale. Therefore labeling term for each scale was exactly as ranking order, but then, it is being ignored by applied researchers as they thought it can help the respondents to make a choice for each questions presented. By doing so, the researchers keep analyze the data obtained using parametric technique without concern the sensitivity of the statistical assumptions.

Lately, Structural Equation Modeling (SEM) has become one of the prominent statistical method as it is take into account of the multiple variables simultaneously and being free from the measurement error that associated with every variable. In SEM, the achievement of fitness of measurement model must be ensured during the Confirmatory Factor Analysis (CFA; Zainudin, 2015). The fitness of measurement model was very sensitive to the characteristics or pattern of the data sets. Therefore, the good fitness of

* corresponding author email: drzainudin888@gmail.com

measurement model was actually represents the data obtained was compatible with the theories conveyed (Henseler, Hubona, & Ray, 2016). In order to obtain the good fitness of measurement model, the researchers must concern with their data sets that is the data was collected based on respondents perception. In the questionnaire, the Likert scale was considered as the measurement scale to assess the degree of the respondents' opinion. If the measurement scale presented more choices, then, it will accelerate the researchers to decide their choice.

In the questionnaire development stage, Likert scale might be coding for 5 point, 7 points and 10 points to represent of how much higher of respondent to agree of the particular question. The original idea for the Likert scale is found in Rensis Likerts 1932 articles in Archive of Psychology titled "A Technique for the Measurement Attitudes". In his research, he expand the present knowledge of the procedures namely Likert scale developed by Thurstone (Edmondson, 2005). In scale of measurement, the 5 points typically been employed to determine the respondents agree or disagree on questions contemplated. However, the scale used with 5 points is inadequate at determining of characters of respondents intention especially when the researchers decide to attempt parametric test in statistical inference. Because 5 points of Likert scales was represents of 5 choices undermine the strength of parametric techniques.

Consequently, this study interest to make a comparison between 5 point and 10 points of Likert scale using SEM approach. The outcome of each factors would portray the real different for each strength of scale of measurement. However, majority of researchers less sensitive on the scale used, instead more likely concentrate on the analysis adapted. In fact, the probability to attain the true value is closely related on the scale used in determining of the parameter estimate and hypothesis testing for inference testing. In addition, the fitness of measurement model and reliability of construct also affected.

The structure of paper is as follows. First, relevant issues on scale of measurement and sampling technique are discussed. This is followed by methodology and generation of scale items used. Once the data analysis is focused on CFA approach, path analysis is discussed. Then after psychometric properties of the scales, results, theoretical, practically, and implication of scale are presented. The paper also concludes the recommendation, future research and scope of limitations.

2. Likert Scale in the form of rank order

Mostly, the researchers today would using the primary data as a main source for them to carry on the particular research required. Therefore, they would design their questionnaire based on their previous literature to support their decision to ask respondents. However, the way they design the questionnaire can harm the result at the data analysis stage. This is because most of them are still confused to distinguish between ordinal and interval scale in data analysis. In fact, the wrong identification data can causes misinterpretation finding.

Ranking procedure require the respondents to order stimuli with respect to some designated property of interest. For instances, the researchers would classify for each 5 point scales in their questionnaire design as presented below:

Figure 1: Likert Scale of Ordinal order

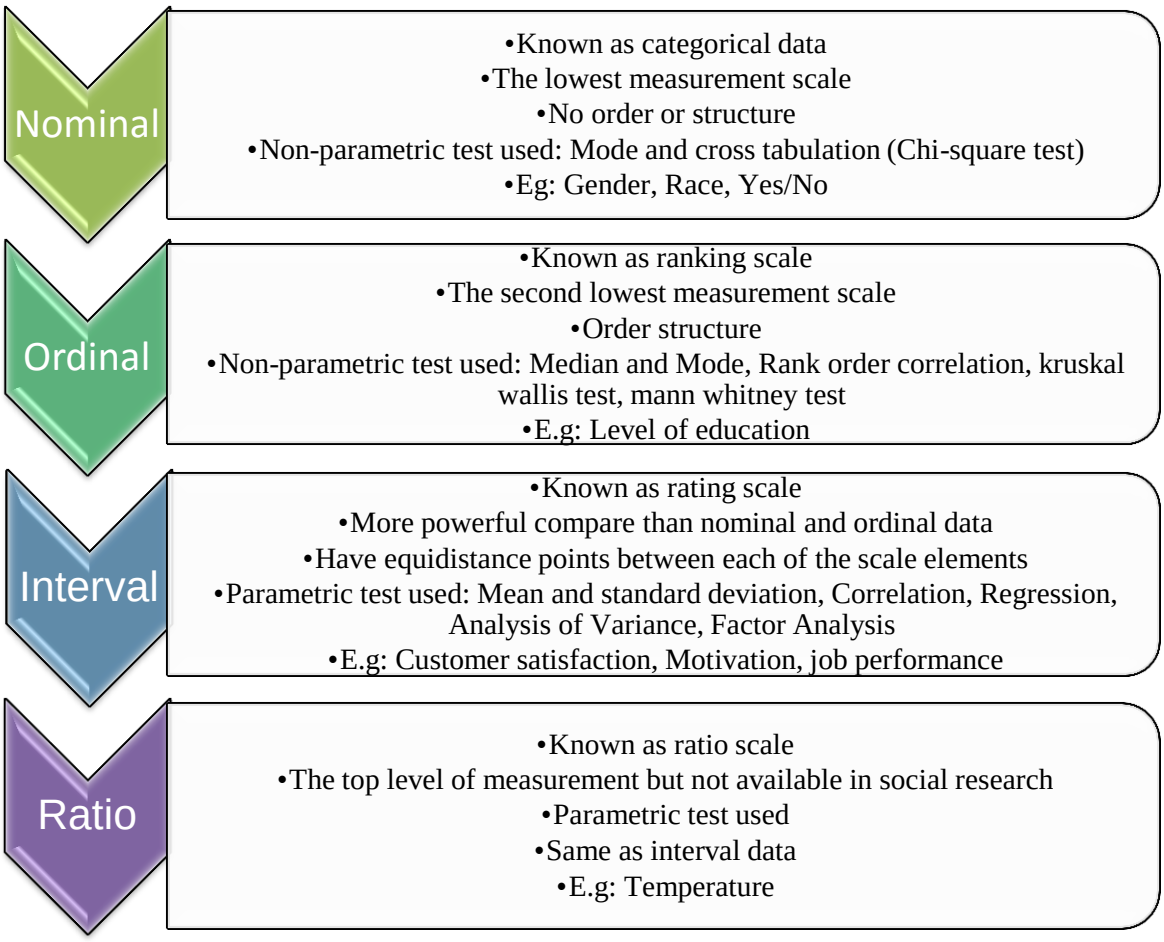
This scale of measurement applied is inappropriate to be applied since the researcher rename of each scale listed. This step would be recognized as ordinal data due to the rank of each scale. In parametric test, the researchers usually intend to analyze on how much of attitude of respondent instead to ask rank of their opinion. Parametric test can be conducted for interval and ratio data, thus, nominal and ordinal is not eligible to be test as parametric analysis.

Figure 1 is the sample of ordinal scale that classify of each scale. This scale is only useful for the non-parametric test such as descriptive frequency, mean, Mann Whitney test, Kruskal Wallis and so forth. So, the major problem here is induced the researchers are inconsiderate to learn the statistical methodology that is prior in empirical research.

In statistics discipline, there are four scale of measurement that should be understand by researchers and scholars so that they would know on how to differentiate of all scale provided. Indeed, all scale can be analyzed using method required to complement their analysis. Nevertheless, many infamous researchers claim that interval data is more powerful in determining of the attitude of respondent on each issues faced (Velleman & Wilkinson, 1993; knapp, 1990; korotayev, 2004). Therefore the interval order often gain attraction of researchers to conduct the parametric analysis as presented Figure 2.

Scale of measurement

Basically, data can be divided into numerical and categorical data. Numerical data contains numbers that we can manipulate using ordinary arithmetic operations, in a while, categorical data can be sorted into categories. (Lau, Phang & Zainudin, 2012). Usually, data is classified as nominal, ordinal, interval or ratio as explained below:



Source: Lau, Phang & Zainudin (2012), Statistics Edition three

2.1. Treating ordinal scales as interval scales

Today, the use of scale of ordinal scale has become a common practice in statistical analysis. Most of the researchers do not really understand of requirement for each scale measurement. For instances, researchers today postulates ordinal scale can be used in parametric test as well such as ANOVA, t-test, regression analysis, and so forth. In fact, those approach can be well functioned with interval or ratio scale only. Other than that, those estimation provided for inference testing are vague.

To add, this issue has been continuing controversy since Harvard psychologist S.S. Stevens (1946) advances his ideas concerning the connection between measurement scales and statistical analysis regarding of traditional descriptive and inferential statistics. Marcus-Roberts and Roberts (1987) put forth that it is always appropriate to calculate means for ordinal scales but inappropriate to make certain statements about such means. This statement justify that the inferential statistics dependent on ordinal is not comprehensive to reflect the actual research that involving of large sample size.

Knapp (1990) remarked that if the researcher convince that their data applied are ordinal, they should be able to carry on their research using traditional statistics that leads to fruitful result. This statement also be agreed by Marcus-Roberts & Roberts, (1987) that if the scale used is ordinal, nonparametric test should be employed in the inference stage, only because of the distribution-free nature of such test but because they tend to be more appropriate for hypotheses that are meaningful for ordinal scale.

The appropriate method will provide a meaningful hypothesis as explain by (Marcus-Roberts & Roberts, 1987).Therefore, treating of ordinal scales as interval scales in inferential stage is improper to be applied since this scales is limited to the nonparametric test. Plus, the meaningless hypothesis would not contribute significant impact on the social research. Thus, the researchers have to construct their questionnaire as interval scales if

interested to apply the parametric test. In SEM, the researchers permanently using maximum likelihood estimator that is one of normality theory. Using this, maximum likelihood is admissible with interval scales. However, Muthens (1998) offered a new estimator as Weighted Least Square Maximum Variance (WLSMV) in the integration of SEM was rationale as well with ordinal scale. Such finding enable the researchers to excel their analysis with SEM. In order to make sure the readers understand distinguish between parametric and non-parametric approach, we listed several approaches at the Table 1 and different abilities between them at Table 2 as follows:

Table 1: Comparison Parametric and Non-Parametric Approach

Parametric Approach (Means)	Non-Parametric Approach (Medians)
Paired t-test	McNemar test
1-sample t-test	Median test, Wilcoxon sign test
2-independent sample t-test	Mann-Whitney test
Pearson Correlation	Spearman Correlation, Tau Equivalent, Rank Correlation, Ordinal Correlation
Analysis of Variance (ANOVA)	Kruskal-Wallis test, Mood's median test
Factorial DOE with one factor and one blocking variable	Friedman test
Multivariate ANOVA (MANOVA)	Permutation MANOVA
Binary Logistic	Cochran test
Regression Analysis	Robust Regression Analysis, Quantile Regression Analysis, Ordinal Regression

Table 2: Comparison abilities between Parametric and Non-Parametric Test

Parametric test	Non-Parametric test
There are numerous statistical assumption or stringent assumption about the parameter means	There is no stringent assumption or they relaxes the statistical assumptions
Parametric center on the mean differences and differences between medians	Non-Parametric center on order or ranking. The data are changes from exact scores to rank and different signs
The population must have the same variance (homoscedasticity)	No same variance (heteroscedasticity)
High statistical power and efficiency	Low statistical power and efficiency
High sensitive to detect the large sample size	Less sensitive to detect the large sample (preferred for small samples)

2.2. Forced Measure (interval scale should be identical and independent)

By using Structural Equation Modeling (SEM) as a main method in this study, the authors would analyze two type of ranges which is for 5 points and 10 points to obtain the result. Usually, the researchers are interested to apply 5 points of measurement scale, instead to apply 10 points of Likert scale. Most of them believe the result cannot be affected even adapt short scale. In fact, the long scale (e.g: 10 points) is better than short scale (e.g: 5 points) in determining of how much agree or disagree of respondent on particular questions.

For instance, if the enumerator ask the respondents to evaluate of the question consists of 5 points, the respondent would rate the particular question in the range of 0% to 100%. If they decide to agree of that questions is 90%, the enumerator would be claim that 90% was exactly represents of 5 points of the Likert scale which is the highest point in scale of measurement in the questionnaire designed. This situation is contrary to the questionnaire that posits of 10 points for each items in which the enumerator would rate the question as 9 point that represents of 90% in scale of measurement. Thereby, it can be inferred that the equidistance is not occurred in measurement scale when correlate with short scale.

In fact, the scale of measurement is supposedly reflect the actual intention of respondent towards the question submitted. If the 5 points is sustain in scale of measurement, the analysis provided will be not reliable even the standard error produced is much lower than of 10 points. The short scale of measurement inhibit the actual scale of

respondent intention and thus pushing them to rate the question based on scale presented. Thus, this situation might create a force measure in inferential statistics that provide a meaningless outcome induced there is no freedom for respondents to make a choice of their intended.

In this respect, the expected value for mean and variance of IID random variables should be identical and independent specifically as X_1, X_2,…, X_n . Means that, the scale used should be independent and identical for constructing the linear formula as stipulated in z-test and t-test in order to prevent occurrence of force measure. Moreover, the common method variance that tendency to harm the construct validity, measurement error and covariance among latent variables resolved.

2.3. Sampling Technique

Sampling is the process of selecting a sample from a population. Since the information obtained from the sample is used to generalize or to make a conclusion about the population, the sample must be selected in such a way that it will accurately represent its population (Lau, Phang, & Zainudin, 2012).

Sampling technique can be classified broadly into two categories namely non-probability sampling technique and the probability sampling technique. The non-probability sampling techniques encompasses of convenience sampling, judgmental sampling, snowball sampling, and quota sampling. In a while, probability sampling encompasses of simple random sampling, systematic sampling, stratified sampling, cluster sampling and multi-stage sampling. Both of these sampling carry the different character of methodology in inferential statistics. In generals, non-probability sampling is appropriate for nonparametric test whereas probability sampling is useful for parametric test.

However, the researchers fail to meet the requirement of parametric test especially during the data collection stage that is associated with the sampling technique. At this stage, they collect the data or distribute the questionnaire based on non-probability sampling and supposedly nonparametric test should be conducted in empirical research. Against on that, they constantly attempt parametric analysis since the evidence clear that parametric analysis is better than non-parametric analysis. Therefore, the finding produced is against their theoretical concept. In order to compare between 5 point and 10 points as suggested in this paper, the data was generated into two different of measurement scale that consist same quantity of data sets. Both of these data sets will be tested at the same confirmatory model such that to examine which kind of measurement scale was preferable for those model. In this stage, CB-SEM was ideal represents of second generation modeling that is able to operate multiple construct measurement (refers Figure 2).

3. Finding

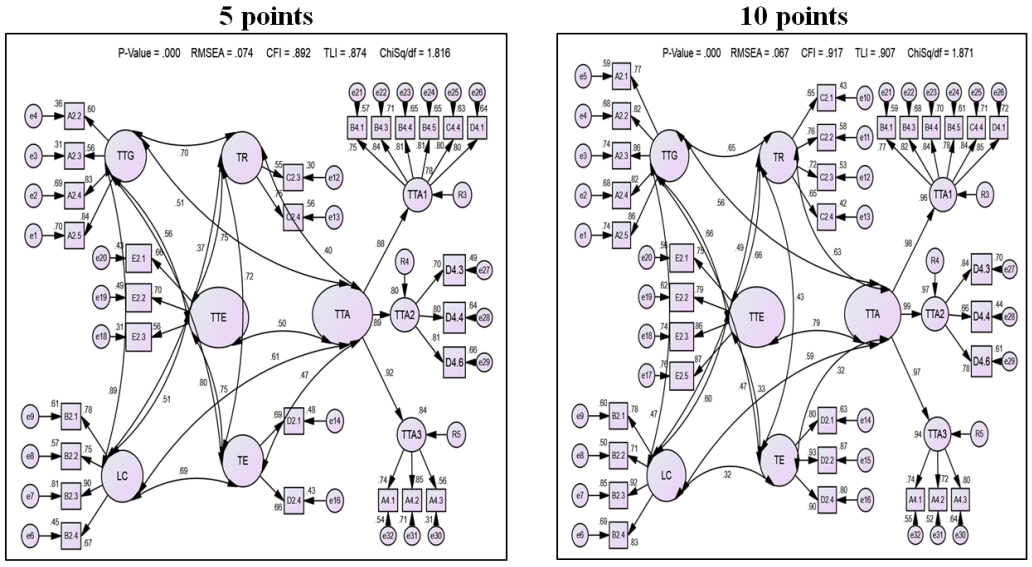


Figure 2: Comparison model between 5 points and 10 points

Figure 2 shows the result of construct measurement that consist of different measurement scales. Based on that, the items retained in the model among them was not equivalent although the technical procedure to specify the construct measurement was identical. Followed by those factor loadings, the next assessment such as Composite Reliability (CR), Average Variance Extracted (AVE) and discriminant validity was performed.

3.1. The 5 points of measurement scales

Table 3: Factor loading of 5 points of measurement scale

Construct	TTA	TTA1	TTA2	TTA3	TR	TTG	LC	TTE	TE
TTA1	.88								
TTA2	.89								
TTA3	.92								
B4.1		.75							
B4.3		.84							
B4.4		.81							
B4.5		.81							
C4.4		.80							
D4.1		.80							
D4.3			.70						
D4.4			.80						
D4.6			.81						
A4.1				.74					
A4.2				.85					
A4.3				.58					
C2.3					.55				
C2.4					.75				
A2.2						.60			
A2.3						.58			
A2.4						.83			
A2.5						.84			
B2.1							.78		
B2.2							.75		
B2.3							.90		
B2.4							.67		
E2.1								.66	
E2.2								.70	
E2.3								.56	
D2.1									.69
D2.4									.66
CR	.925	.915	.815	.772	.598	.810	.860	.677	.626
AVE	.804	.643	.595	.536	.433	.523	.607	.413	.456

Table 4: Discriminant Validity

	TTA	TR	TTG	LC	TTE	TE
TTA	.896					
TR	.48	.658				
TTG	.51	.70	.723			
LC	.61	.37	.89	.779		
TTE	.50	.75	.56	.51	.642	
TE	.47	.72	.80	.69	.75	.543

3.2. The 10 points of measurement scales

Table 5: Factor loadings of 10 points of measurement scales

Construct	TTA	TTA1	TTA2	TTA3	TR	TTG	LC	TTE	TE
TTA1	.98								
TTA2	.99								
TTA3	.97								
B4.1		.77							
B4.3		.82							
B4.4		.84							
B4.5		.78							
C4.4		.84							
D4.1		.85							
D4.3			.84						
D4.4			.66						
D4.6			.78						
A4.1				.74					
A4.2				.72					
A4.3				.80					
C2.1					.65				
C2.2					.76				
C2.3					.72				
C2.4					.65				
A2.1						.77			
A2.2						.82			
A2.3						.86			
A2.4						.82			
A2.5						.86			
B2.1							.78		
B2.2							.71		
B2.3							.92		
B2.4							.83		
E2.1								.75	
E2.2								.79	
E2.3								.86	
E2.5								.87	
D2.1									.80
D2.2									.93
D2.4									.90
CR	.986	.923	.806	.798	.754	.915	.886	.890	.910
AVE	.960	.668	.583	.569	.506	.683	.662	.671	.772

Table 6: Discriminant validity

	TTA	TR	TTG	LC	TTE	TE
TTA	.997					
TR	.63	.711				
TTG	.56	.65	.826			
LC	.59	.49	.47	.813		
TTE	.79	.66	.66	.60	.819	
TE	.32	.43	.47	.32	.33	.878

Based on the results revealed, the construct measurement that consist of 10 points of measurement scales was seemed more relevant under the confirmatory approach. Because the factor loadings appear in the model was higher than 0.60 (Zainudin, 2015), value of CR was greater than 0.70 (Nunally & Bernstein, 1994; Afthanorhan, Ahmad, & Mamat, 2014), value of AVE higher than 0.50 (Hair et al., 2009) and discriminant validity that associated with latent variable correlation was satisfied (correlation must be less than the square root of AVE; Fornell & Larcker, 1981). Therefore, the construct measurement of 10 points (refer Table 5 and Table 6) is able to proceed to the next stage that is the relationships between latent variables can be confirmed. In contrast, the results of construct measurement with 5 points of Likert scale (refer Table 3 and Table 4) was seemed fail to satisfy the SEM requirement as this approach can be admissible when the construct measurement was satisfied in terms of its reliability and validity. Other than that, the path analysis for estimating the relationships among latent variables cannot be executed.

In SEM, the first priority is to ensure the construct involve in the model was reliable and valid. Thereby, this method is frequently viewed as confirmatory tool (Bollen, 2014) that is the model must be have a strong theory. The necessity of strong theory should be compatible with the data available to avoid more items deleted from the

model. In this case, we confirm that the 5 points of Likert scale was not meet with such requirement of parametric based SEM.

4. Conclusion

The plethora of discussion and debate around measurement scale called for a fresh look at this scale as the use of Likert scale become a common practice in social science and management research. As important aspect endeavor, we provide an answer the question "What point of scale is preferable?" This answer is 10 points of Likert scale as justified by our recent findings. It is particularly helpful for researchers who have educated in measurement scale in the past, and who interest to add knowledge or understanding of the measurement scale. With this, they know how to distinguish which one is called ordinal or interval and further they more convince to choose the best scale for their research.

Finally, recent research confirms that 10 points of Likert scale serves a promising scale under parametric based SEM. Both measurement and structural models can be assessed with 10 points of Likert scale that is expected more success in determining the construct validity. We anticipate that once social science and management's researchers' interest in developing questionnaire based of Likert scale becomes more pronounced, 10 points of Likert scale will face an additional substantial gain popularity and surely the danger of common method variance in the model can be evaded.

5. References

[1] Afthanorhan, W. M. A. B. W., Ahmad, S., & Mamat, I. (2014). Pooled Confirmatory Factor Analysis (PCFA) using structural equation modeling on volunteerism program: A step by step approach. International Journal of Asian Social Science, 4(5), 642-653.
[2] Bollen, K. A. (2014). Structural equations with latent variables. John Wiley & Sons.
[3] Edmondson, R.D. (2005), "Likert Scales: A History", University Florida, USA, 127-133.
[4] Fornell, C., & Larcker, D. F. (1981). Structural equation models with unobservable variables and measurement error: Algebra and statistics.Journal of marketing research, 382-388.
[5] Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2009). Análise multivariada de dados. Bookman Editora.
[6] Henseler, J., Hubona, G., & Ray, P. A. (2016). Using PLS path modeling in new technology research: updated guidelines. Industrial Management & Data Systems, 116(1), 2-20.
[7] Knapp, T. R. (1990). Treating ordinal scales as interval scales: an attempt to resolve the controversy. Nursing research, 39(2), 121-123.
[8] Korotayev, A. (2004). World religions and social evolution of the old world Oikumene civilizations: A cross-cultural perspective. E. Mellen Press.
[9] Kya, L. T., Ngor, P. Y., & Awang, Z. (2012). Statistics for UiTM.
[10] Likert, R. (1932). A technique for the measurement of attitudes. Archives of psychology.
[11] Marcus-Roberts, H. M., & Roberts, F. S. (1987). Meaningless statistics.Journal of Educational and Behavioral Statistics, 12(4), 383-394.
[12] Muthén, B. O. (1998). Mplus technical appendices. Los Angeles, CA: Muthén & Muthén, 2004.
[13] Nunally, J. C., & Bernstein, I. H. (1994). Psychonometric theory. New York.
[14] Velleman, P. F., & Wilkinson, L. (1993). Nominal, ordinal, interval, and ratio typologies are misleading. The American Statistician, 47(1), 65-72.
[15] Zainudin Awang. (2015). A Handbook on SEM. MPWS Publisher.