British Journal of Mathematics & Computer Science
16(6): 1-9, 2016, Article no.BJMCS.25707
ISSN: 2231-0851

SCIENCEDOMAIN *international*
www.sciencedomain.org

Factor Structure, Reliability and Validity of Attitudes of Biostatistics Scale

Muhammad Shafiq Harun[1*] and Sabri Ahmad[1]

[1]*School of Informatics and Applied Mathematics, University Malaysia Terengganu, 21030 Kuala Terengganu, Malaysia.*

Authors' contributions

This work was carried out in collaboration between both authors. Author SA designed the study, wrote the protocol and supervised the work. Author MSH carried out collecting data and performed the statistical analysis. Author MSH managed the analyses of the study and wrote the first draft of the manuscript. Author MSH also managed the literature searches and edited the manuscript. Both authors read and approved the final manuscript.

Article Information

DOI: 10.9734/BJMCS/2016/25707
Editor(s):
(1) Kai-Long Hsiao, Taiwan Shoufu University, Taiwan.
Reviewers:
(1) Radosław Jedynak, Kazimierz Pulaski University of Technology and Humanities, Poland.
(2) Hazem Ismail A. El Shekh Ahmed, Al Quds Open University, Palestine.
(3) Yaşar Barut, Ondokuz Mayıs University, Turkey.
Complete Peer review History: http://sciencedomain.org/review-history/14729

Short Research Article

Received: 16th March 2016
Accepted: 4th May 2016
Published: 22nd May 2016

Abstract

Confirmatory Factor Analysis (CFA) has been used for most of researchers nowadays to evaluate the goodness of fit of measurement model using structural equation modeling. The aim of this study to evaluate the factors used to validate the best model of four latent construct by using pooled confirmatory factor analysis (PCFA) technique on variable student's attitude toward Biostatistics and to assess dimensions of students' attitudes regarding Biostatistics courses. A survey adapted from Survey of the Attitudes toward Statistics (SATS) was employed to observe student's attitude toward a Biostatistics course. The data be collected through questionnaires distributed to first year students at a higher education institution. The data were analysed through four model which is model specification, model estimation, model evaluation, and model modification by using Analysis Moment of Structural (AMOS) in order to improve the validity of each latent construct. The result showed that the validity and reliability of all latent variables is achieved and the pooled CFA technique is more efficient.

**Corresponding author: E-mail: shafiqharunsalleh@gmail.com;*

Keywords: Pooled confirmatory factor analysis; factor structure; unidimensionality; validity; reliability and latent variables.

1 Introduction

The structural equation modeling (SEM) is a powerful multivariate statistical method in order to study the interrelationship among observed and/or latent variables. SEM is also a confirmatory method providing a comprehensive means for validating the measurement model of latent variables. This statistical procedure is called confirmatory factor analysis (CFA) and according to Hair et al. [1] describe CFA as statistical test to measures the latent variables are consistent with the researcher's belief of the nature of that latent variable. Nowadays, CFA is frequently used among researchers in order to achieve the validity and reliability of latent variables.The CFA method has the ability to assess the unidimensionality, validity and reliability for a latent variables. Basically, researchers or scholars need to perform CFA for all latent variable first before modeling their interrelationship in structural model.

There are two approach of CFA for the measurement model for each measurement model separately and the pooled CFA for measurement model at once [2]. However, pooled CFA is more efficient and highly suggested compare to run CFA separately for each latent variables [3,2]. Previous study state the limitation of CFA by analyses the measurement model separately due to the identification issues [3]. Thus, the proposed method namely Pooled CFA (PCFA) is no doubt to ease the scholar to carry out their research besides prone them to better understanding on the meant of empirical study [3]. The aim of this study to evaluate the factors used to validate the best model of four latent variable that are Value, Difficulty, Affect, and Cognitive Competence by using pooled confirmatory factor analysis (PCFA) technique on variable student's attitude toward Biostatistics and to assess dimensions of students' attitudes regarding Biostatistics courses.

2 Methodology

2.1 Pooled Confirmatory Factor Analysis (PCFA)

Confirmatory Factor Analysis (CFA) is a special form of factor analysis. It is employed to test whether the measures of a latent variable are consistent with the researcher's understanding of the nature of the variables. The CFA procedure replaced the older methods as exploratory factor analysis and Cronbach alpha reliability to determine variables reliability and validity. Recently, the more efficient and highly suggested method for assessing the measurement model was proposed using pooled confirmatory factor analysis [3]. This method combines all latent constructs in one measurement model and perform the CFA at once. The item deletion process and model respecification are made as usual.

This method is more preferred since it could handle the issue of identification problem. In other words, by using PCFA, the all latent variables examined simultaneously and the correlations between latent variables are also computed. So that, if the latent variable is correlated, then the multicollinearity problem is said be exist. The discriminant validity failed if the correlation between two exogenous latent variables is more than 0.85 (bivariate correlation). High correlation indicates that latent variables are redundant. In order to solve the variables redundancy, the researchers or scholars need to combine the exogenous latent variables to become one exogenous latent variable and run the PCFA again. Another solution is to drop one of these redundant exogenous latent variables before modeling the structural model.

This PCFA technique more efficient, better, and easier since researchers or scholars can assessing the multicollinearity and unidimensionality simultaneously compare running CFA for each latent variables separately. Once the PCFA procedure for measurement model is completed, the researchers or scholars need to compute other remaining measures which indicate the validity and reliability of the measurement model and summarize it. As has been discussed above, the requirement for unidimensionality, validity, and reliability needs to be addressed prior to modeling the structural model.

In order to evaluate the measurement model by PCFA technique, maximum likelihood estimation method have been used in this study. The maximum likelihood (ML) is traditional and most frequently used for estimating parameter in SEM. ML uses an iterative process to minimize the discrepancy between the sample covariance matrix and the reproduced covariance matrix, quantified by a fit function [4,5,6].

2.2 Unidimensionality

Unidimensionality is the degree to which items load only on their respective variables without having parallel correlational pattern [7]. Unidimensionality cannot be assed using factor analysis or Cronbach alpha internal consistency coefficient [8,7].

Unidimensionality is achieved when the measuring items have acceptable factor loadings for the respective latent variables. In order to ensure unidimensionality of measurement model, any observed variables with a low factor loading should be dropped. The deletion should be made one observed variables at a time with the lowest factor loadings to be deleted first. After an observed variables is deleted, the researchers need to respecify and run the new pooled measurement model. The process continues until the unidimensionality requirement is achieved [2].

2.3 Validity

Validity is the ability of instruments to measure what it supposed to be measured for a construct. Three types of validity are required for each measurement model are. Assessing validity is very important in order the get the best fit model before proceed with structural model.

2.3.1 Construct validity

The construct validity is achieved when the goodness of fit indexes for a latent variable achieved the required level. The goodness of fit indexes important in indicate how fit is the observed variable or items in measuring their respective latent variables. The goodness of fit indexes, their respective category, and the level of acceptance are discussed on Sub-Section 2.4.

2.3.2 Convergent validity

The convergent validity is achieved when all items in a measurement model are statistically significant. The convergent validity could also be verified through Average Variance Extracted (AVE). The value of AVE should be greater than 0.50 in order to achieve convergent validity.

2.3.3 Discriminant validity

The discriminant validity is achieved when the measurement model is free from redundant observed variables. AMOS will identify the pair of redundant observed variables in the model and reported in the Modification Index (MI). However, the certain cases the researchers could set the correlated pair as “free parameter estimates”. Another requirement for discriminant validity is the correlation between each pair of exogenous latent variables should be less than 0.85. The exogenous latent variables will having multicollinearity when the correlation between pair of exogenous latent variables greater than 0.85.

2.4 Reliability

Reliability is the extent of how reliable is the said measurement model in measuring the intended latent construct. The assessment of the reliability of a measurement model could be made using the following criteria:

2.4.1 Composite reliability

Composite Reliability (CR) is the measure of reliability and internal consistency of the observed variables representing the latent construct. A value of CR > 0.60 is required in order to achieve construct reliability [9]. CR was calculated by the formula (2.4.1):

$$\sum \mathrm{K}^2 \Big/ \left[\left(\sum \mathrm{K} \right)^2 + \left(\sum 1 - \mathrm{K}^2 \right) \right] \tag{2.4.1}$$

where K is the factor loading of each item and n is the number of item in a model.

2.4.2 Average variance extracted

Average Variance Extracted (AVE) is value explained average percentage of variation explained by the items in a construct. An AVE > 0.50 is required [10]. AVE was calculated by the formula (2.4.2):

$$\sum \mathrm{K}^2 \Big/ n \tag{2.4.2}$$

where K is the factor loading of each item and n is the number of item in a model.

2.5 Goodness of fit in measurement model

In structural equation modeling, there is a series of goodness of fit indexes that reflects the fit of the model to the data at hands [11]. At the moment, there is no agreement among the researchers and scholars which goodness of fit indexes should be reported since they have a lot of goodness of fit in structural equation modeling. Holmes-Smith et al. [12] recommends using at least one goodness of fit index form each category of fit model. There are 3 categories of goodness of fit which are absolute fits, incremental fits and parsimonious fits. The following Table 1 shows the type of goodness of fit indexes with literature support.

Table 1. Goodness of fit indexes

Name of category	Name of index	Acceptance level	Comment
Absolute fit	**Root Mean Square of Error Approximation (RMSEA) [13]** $RMSEA = \sqrt{\frac{\max[(T_T - df_T)/(N-1), 0}{df_T}}$	RMSEA < 0.08	Higher value of GFI as well as lower value of RMSEA indicate better model data fit
	Goodness of Fit Index (GFI) [14] $GFI = \frac{p}{p + 2[(T - df)/(N-1)]}$	GFI > 0.90	
Incremental fit	**Comparative Fit Index (CFI) [4]** $CFI = 1 - \frac{\max[(T_T - df_T), 0]}{\max(T_T - df_T), (T_B - df_B), 0}$	CFI > 0.90	Higher value of incremental fit indicate larger improvement over the baseline model in fit
	Tucker-Lewis Index (TLI) [5] $TLI = \frac{(T_B - df_B) - (T_T - df_T)}{(T_B / df_B - 1)}$	TLI > 0.99	
Parsimonious fit	**Chi Square/Degree of Freedom (Chisq/df) [15]** $Chisq/df = \frac{\chi^2}{df}$	Chisq/df < 5.0	Very sensitive to the sample size

**the index in bold are recommended in most of literatures*

2.6 Student attitude toward biostatistics course

This study focuses on student attitude toward Biostatistics course has 4 latent variables namely Value, Difficulty, Affect and Cognitive Competence that will be conducted for PCFA analysis. These four latent variables consists of 28 observed variables that has been developed for the specific population using questionnaire. Means that, the respondents should answer all of the questionnaire regarding their attitude toward Biostatistics course. A survey adapted from Survey of Attitudes toward Statistics (SATS) was employed to observed student's attitude toward a Biostatistics course. On the use of PCFA analysis will ascertain the researchers to determine whether the questionnaire developed is performed well or not for the respondents. If not, some of the questions will be removed and the remaining question will be proceeded for the subsequent analysis. In other words, the removal questions may not appropriate for the case study.

3 Results and Discussion

Based on the purpose of research, the PCFA procedure was conducted. All measurement models must be validated and accepted prior to modelling the structural model. For this study, there are have 4 dimensions which are Value (9 observed variables), Difficulty (7 observed variables), Affect (6 observed variables), and Cognitive Competence (6 observed variables). The factor loadings for each observed variables should be greater than 0.6. However, observed variables with factor loading which is greater that 0.5 is also accepted depend on the decision by the researcher if strong reason not to delete that observed variables. Table 2 shows the observed variables result remain after deleted low factor loading observed variables:

Table 2. Unidimensionality result

Latent variables	Number observed variables before remove	Number observed variables after remove
Value	9	7
Difficulty	7	7
Affect	6	6
Cognitive competence	6	5

Despite having the unidimensionality procedure, the model estimation, model evaluation and model modification should be apply in order to obtain the exactly result. The model evaluation is considered as the goodness of fit model. The model evaluation can be obtained based on the Root Mean Square of Error Approximation (RMSEA), Baseline Comparison (CFI, TLI, NFI) and Chisquare over degree of freedom. The result can be obtained as the table presented Tables 3 and 4.

Only Cognitive Competence latent variables is valid since for all category of goodness of fit index is achieved as the recommended by the literature (Table 3). For other three latent variables Value, Difficulty and Affect is not achieved their goodness of fit index for each category. Thus model modification is employed to remedy the multicollinerity problem.

Table 3. Goodness of fit before constraints

Latent variables	Absolute fit		Incremental fit			Parsimonious fit
	RMSEA	**GFI**	**CFI**	**TLI**	**NFI**	**Chisq/df**
Value	0.08	0.94	0.98	0.97	0.96	1.85
Difficulty	0.19	0.84	0.91	0.87	0.90	5.28
Affect	0.20	0.86	0.90	0.85	0.89	5.84
Cognitive competence	0.07	0.97	0.99	0.98	0.98	1.65

Based on the Table 4, the construct validity is achieved since the goodness of fit of indexes for all latent variables achieved the required level after apply the constraint or "free parameter estimate" that for model modification.

Table 4. Goodness of fit after constraint

Latent variables	Absolute fit		Incremental fit			Parsimonious fit
	RMSEA	**GFI**	**CFI**	**TLI**	**NFI**	**Chisq/df**
Value	0.07	0.96	0.99	0.98	0.97	1.50
Difficulty	0.08	0.97	0.99	0.98	0.97	1.70
Affect	0.09	0.96	0.98	0.97	0.97	2.10
Cognitive competence	0.07	0.97	0.99	0.98	0.98	1.65

Then, the convergent validity should be employed to validate all items in measurement models that are statistically significant. The Table 5 shows the convergent validity was achieved since the value of AVE for all construct were greater than 0.5. The items do correlate well with each other within their latent construct that is the latent factor is well explained by its observed variables.

The Fig. 1 shows the measurement model by PCFA after evaluate the goodness of fit test with value of correlation. This process is important to develop the discriminant validity for latent exogenous and endogenous variables. Hence, the constraint of double headed arrow is required to examine the strength of the relationship between these latent variables.

Based on the Table 6, all of the latent variables shows the correlation measure are below 0.85. Thus the discriminant validity is achieved and all of these latent variables could be used in a structural model for the further analysis. If the correlation value between to exogenous variable is higher than 0.85, one can conclude that the discriminant validity is not achieve [2]. For case like that, the exogenous latent variables are redundant of each other. Therefore, either one of the latent variables must be remove or drop in the subsequent analysis procedure.

Table 5. Convergent validity result

Latent variables	AVE
Value	0.633
Difficulty	0.662
Affect	0.620
Cognitive competence	0.613

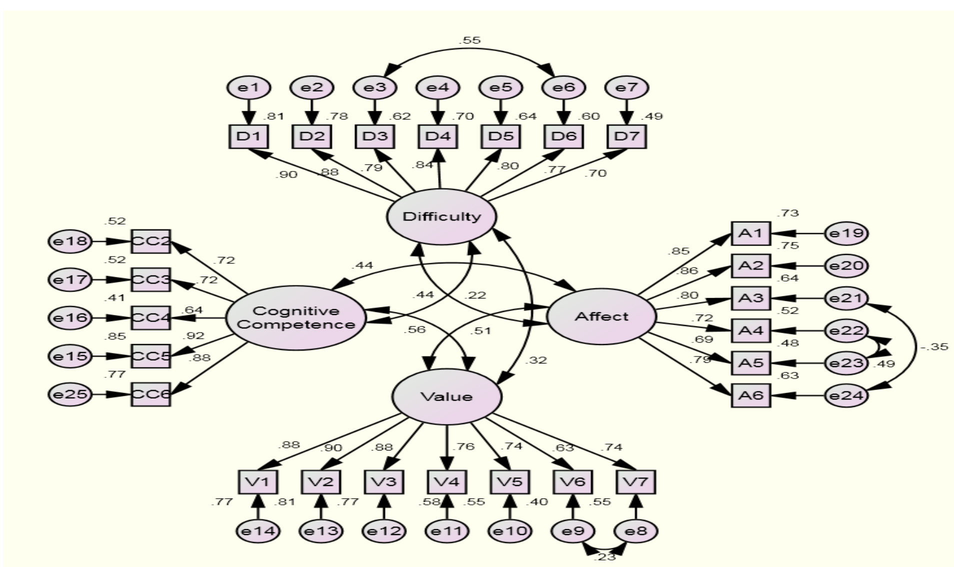

Fig. 1. The measurement model for student attitudes toward biostatistics

Table 6. Correlation among latent variables

			Estimate
Difficulty	<-->	Cognitive competence	0.443
Difficulty	<-->	Affect	0.221
Value	<-->	Affect	0.510
Value	<-->	Cognitive competence	0.556
Cognitive competence	<-->	Affect	0.442
Difficulty	<-->	Value	0.324

The diagonal values (in Bold) for Table 7 is the square root of AVE while other value are correlation between the respective latent variables. The discriminant validity for all latent variables is achieved when the diagonal value is higher than the value in its row or column. So based on Table 7, one can conclude that the discriminant validity for all latent variables is achieved than the observed variables correlate lower with items in other latent variables compare to observed variables within their latent variables. It means, the latent variables is better explained by some by its own observed variables than some other observed variables.

Table 7. Discriminant validity index summary

Latent variables	Value	Difficulty	Affect	Cognitive competence
Value	**0.80**			
Difficulty	0.32	**0.81**		
Affect	0.51	0.22	**0.79**	
Cognitive competence	0.56	0.44	0.44	**0.78**

The assessment for reliability for the measurement model has be made using two criteria which is Composite Reliability (CR) and also Average Variance Extraction (AVE). Table 8 shows the result for the assessment of reliability for the measurement model.

Table 8. Reliability result

Latent variable	Observed variable	Factor loading	CR	AVE
Value	V1	0.88	0.922	0.633
	V2	0.90		
	V3	0.88		
	V4	0.76		
	V5	0.74		
	V6	0.63		
	V7	0.74		
	V8	Item was deleted		
	V9	Item was deleted		
Difficulty	D1	0.90	0.932	0.662
	D2	0.88		
	D3	0.79		
	D4	0.84		
	D5	0.80		
	D6	0.77		
	D7	0.70		
Affect	A1	0.85	0.907	0.620
	A2	0.86		
	A3	0.80		
	A4	0.72		
	A5	0.69		
	A6	0.79		

Latent variable	Observed variable	Factor loading	CR	AVE
Cognitive competence	CC1	Item was deleted	0.886	0.613
	CC2	0.72		
	CC3	0.72		
	CC4	0.64		
	CC5	0.92		
	CC6	0.88		

The value of CR for all latent variables is greater than 0.6. That mean, the composite reliability was achieved in order to measure the reliability and internal consistency for a latent variables. Lastly for the average percentage of variation explained by measuring item for each latent variables is 63.3%, 66.2%, 62.0% and 61.3% respectively. Therefore, the measurement model is reliable in measuring the intended latent construct.

4 Conclusion

Using the factor Students Attitudes toward Biostatistics as a research model, the finding revealed all the validity and reliability of measurement model which PCFA procedure is achieved. In addition, it is clearly prove that PCFA is efficient and better than run CFA for each measurement model because it is time saving. The CFA procedure is very important before furthering the analysis. Hence, the reliability and validity applied to remedy the multicollinearity problem besides to improve the goodness of fit of the measurement model. The better model is depended the goodness of fit indexes of measurement. Thus, the requirement for unidimensionality, validity and reliability needs to be addressed prior to modelling the structural model.

Competing Interests

Authors have declared that no competing interests exist.

References

[1] Hair J, Black W, Rabin B, Anderson R. Multivariate data analysis (7th ed.). Englewood Cliffs, NJ: Prentice Hall; 2010.

[2] Awang Z. SEM made simple: A gentle approach to learning structural equation modeling: MPWS Rich Publication; 2015.

[3] Afthanorhan WM, Ahmad S, Mamat I. Pooled Confirmatory Factor Analysis (PCFA) using structural equation modeling on volunteerism program: A step by step approach. International Journal of Asian Social Science. 2014;4(5):642-53.

[4] Bentler PM. Comparative fit indexes in structural models. Psychological bulletin. 1990;107(2):238.

[5] Bentler PM, Bonett DG. Significance tests and goodness of fit in the analysis of covariance structures. Psychological bulletin. 1980;88(3):588.

[6] Tanaka JS, Huba GJ. A fit index for covariance structure models under arbitrary GLS estimation. British Journal of Mathematical and Statistical Psychology. 1985;38(2):197-201.

[7] Segars AH. Assessing the unidimensionality of measurement: A paradigm and illustration within the context of information systems research. Omega. 1997;25(1):107-21.

[8] Anderson JC, Gerbing DW. Structural equation modeling in practice: A review and recommended two-step approach. Psychological bulletin. 1988;103(3):411.

[9] Nunnally JC, Bernstein IH, Berge JM. Psychometric theory. New York: McGraw-Hill; 1967.

[10] Fornell C, Larcker DF. Evaluating structural equation models with unobservable variables and measurement error. Journal of Marketing Research. 1981;39-50.

[11] Bollen KA. A new incremental fit index for general structural equation models. Sociological Methods & Research. 1989;17(3):303-16.

[12] Holmes-Smith P, Coote L, Cunningham E. Structural equation modeling: From the fundamental to advance topics. Melbourne: Streams; 2006.

[13] Cudeck R, Browne MW. Cross-validation of covariance structures. Multivariate Behavioral Research. 1983;18(2):147-67.

[14] Jöreskog KG, Sörbom D. LISREL 7: A guide to the program and applications. Spss; 1989.

[15] Marsh HW, Hocevar D. Application of confirmatory factor analysis to the study of self-concept: First-and higher order factor models and their invariance across groups. Psychological Bulletin. 1985;97(3):562.

Peer-review history:
The peer review history for this paper can be accessed here (Please copy paste the total link in your browser address bar)
http://sciencedomain.org/review-history/14729