Australian Journal of Basic and Applied Sciences, 5(12): 1098-1103, 2011
ISSN 1991-8178

Confirmatory Factor Analysis (Cfa) for Testing Validity And Reliability Instrument in the Study of Education

[1]Hamdan Said, [2]Badrullah Bakri Badru and [3]Shahid. M

[1,2]Faculty of Education.
[3]Department of Physics University of Technology Malaysia Johor Bahru 81310 Malaysia.

Abstract : The exploit of factor analysis is to ordeal the hypotheses about the dormant traits that underlie a set of measured variables. The traditional factor analysis approaches such as Pearson correlation and Cronbach's Alpha have some limitations. The aim of this paper is to draw on the application of Confirmatory Factor Analysis (CFA) in Structural Equation Modeling (SEM), to test the validity and reliability of instruments in the field of education. To trounce the drawbacks of Pearson correlation and Cronbach's Alpha in the measurement of validity and reliability, Confirmatory Factor Analysis (CFA) using Structural Equation Modeling (SEM), have been used to test the validity and reliability of the instruments. Various tests i.e. Regression Weights, Standardized Regression Weights, Convergent Validity, Variance Extracted, Construct Reliability, and Discriminant Validity have depicted improved result with better validity and reliability.

Key words: Validity, Reliability Cfa, Sem, Regression Weights, Standardized Regression Weights, Convergent Validity, Variance Extracted, Construct Reliability, Discriminant Validity.

INTRODUCTION

Education system in general comprises a classroom management, handling assignments, effectual communication skills and conducting different question answer session and quiz with students (Shulamn, L.S., 1987). Teachers' beliefs are considered as part of knowledge for teaching, which plays a vital role in giving impact on teaching (Lantin, A.J.P. and A.K.M. Sangalang, 2009). Point-Biserial Correlation, Pearson Correlation, KR-20, Cronbach-Alpha, etc are the Various statistical tests ,which are used to estimate the validity and reliability of instruments using in any educational systems (Siniscalco, M.T., 2005; Gay, L.R., 2006; Creswell, J.W., 2008). Various instruments recently have been used to measure the validity and reliability by using Confirmatory Factor Analysis (CFA) (Cabrera, P., 2010; Jung, J., *et al*., 2010; Légaré, F., *et al*, 2011; Agung, I.G.N., 2008; Meihan, L., 2011; Martin, N.K., 2010). Structural equation modeling (SEM) is a confirmatory technique used for exploratory purposes (B.B. James, 2006; R.B. Kline. 2005). SEM includes two components i.e. CFA and structural model. CFA depicts the pattern of observed variables for those latent construct hypothesized model. Confirmatory factor analysis plays the role of validating and finding the reliability of any measurement in most social science studies (Hernandez, R., 2010).

Now, the hitch is, what will be the appropriate statistical test used to test the validity and reliability of research instruments in the field of education, which uses an ordinal scale in data collection. If we use an interval or ratio scale, then Pearson r test can be used. However, if the scale used is nominal or ordinal, then the use of the test statistics will be wrong.

The added problem is associated with the number of samples. In general, samples which are used in testing the instrument are relatively smaller than the actual number of samples size acquired during data collection. Now the question is whether this relatively small amount of sample is representative to make inferences from the results of the testing instrument?

In order to takeover these tribulations; we have used a Confirmatory Factor Analysis (CFA) in Structural Equation Modeling (SEM) to test the validity and reliability of the instruments in the field of education. We have used AMOS 18.0 programs relating to the application of SEM within the framework of a CFA model.

Theory:

Reliability:

Reliability is the degree of consistency of an instrument. In other words, a reliable instrument is that which gives identical score at all times (Kerlinger, F. N., 2000). Creswell, JW., (2008) divides reliability into five types, namely:

(i) Test-retest reliability: It decribes, how far a score of one sample is stable at different testing times.
(ii) Alternate forms reliability: It involves the use of the same instrument to measure the linkage concept or variable in a group of individuals.

Corresponding Author: Shahid. M, Department of Physics University of Technology Malaysia Johor Bahru 81310 Malaysia.
E-mail: [bluefiber08@gmail.com]

(iii) Alternate forms and test-retest reliability: It is a sort of reliability that takes into account the level of rate stability over time and the equality of items.
(iv) interrater reliability: It is a procedure, that is used when making behavioral observations. It involves observations made by individuals against the behavior of an individual or several individuals.
(v) Internally consistent reliability: It deals the scores indicating internal reliability of all items on an instrument.

In the intervening time, Gay, LR, Mills, GE, and Airasian, P. (2006) have put forward a sixth type, Split-Half reliability. This deals the size of the internal consistency test; involve a division into two parts.

Validity:

Validity is the measure of the accuracy of an instrument used in a study(Linn, R.L, 2000; Stewart, C.D., 2009). As with the reliability, validity also consists of several types, namely
(i) Content validity: It is estimated by testing the validity of the content of the instrument by rational analysis or through professional judgment.
(ii) Criteria Validity: It requires the availability of external criteria that can be used as the basis of test score instrument.
(iii) Construct Validity: It is the validity of theoretical involving building variables to be measured. An instrument is said to have construct validity if the items are arranged in a matter of instruments to measure every aspect of thinking of a variable to be measured by these instruments. Construct validity testing of the instrument is rarely carried out among students, but is often done is to test the validity of the criteria.

Methodology:

Application of Confirmatory Factor Analysis (CFA) in Structural Equation Modeling (SEM) to Test the Validity and Reliability of the Instrument:

Different trials were conducted in the construct validity using CFA (Confirmatory Factor Analysis). In constructs validity, four sizes have been used, i.e. Convergent Validity, Variance Extracted, Construct Reliability, and Discriminant Validity (Arbuckle, J.L., 2010; Dimitrov, D.M., 2003; Ferdinand, A., 2002; Ghozali, I., 2004; Hair, *et al*., 1998; Hisyam, 2010; Hwang, W.Y., 2004; Idris, R., 2010; Lawson, A.B, 2010). Convergent Validity intended to see how big indicator Converge or shares in a single construct. An indicator is said to converge if it has a factor loading value is high and significant. In addition, it has a standardized factor loading estimate greater than 0.5. The construct validity is determined by the average value AVE (Average Variance Extracted). AVE values got hold of the formula:

$$\text{AVE} = \frac{Sum\ \ of\ Standardized\ Loadning\ Square}{Sumof\ \ Standardized\ Loadning\ Square + measurement\ error} \qquad (1)$$

Measurement error = 1 – (Stadardized Loading)^2

Construct Reliability (CR) is intended to determine the consistency of construct validity indicator. Construct Reliability was calculated by the formula:

$$\text{CR} = \frac{Square\ of\ Total\ Standardized\ Loadning}{Square\ of\ Total\ \ Standardized\ Loadning + measurement\ error} \qquad (2)$$

Discriminant Validity test shows how much variance is in the indicators that are able to explain variance in the construct. Discriminant Validity (DV) value obtained from the root of AVE value as:

$$\text{DV} = \sqrt{\text{AVE}} \quad \text{---------------} \qquad (3)$$

Testing Validity and Reliability using Ordinal Scale:

Case I:

In order to test the validity of motivation level, an instrument comprising 10 item questions, with the number of sample of 20 students was considered. The instrument questionnaire had four choices: (1) Strongly Disagree (2) Disagree (3) Agree (4) strongly Agree.

The test results were compiled and enumerated using SPSS and tabulated in Table 1 and Table 2.

Table 1: Reliability Statistics for case 1.

Cronbach's Alpha	N of Items
.882	8

Result depicted that the value Cronbach's Alpha value for the whole item is valid (after invalid items excluded) for 0.882, which means that the instrument has a high level of consistency (above 0.85). On the other hand, the value of Cronbach's Alpha for deleted Item was found to be greater than 0.444. Thus it can be wraped up that the instrument with point 1 to point 8 in Table 2 has a high consistency, or fit for use in data collection.

The exceeding results were compared with Confirmatory Factor Analysis (CFA) that is in the Structural Equation Modeling (SEM) using Amos 18.0 program.

Table 2: Item-Total Statistics.

No	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Item_1	20.7000	28.116	.408	.889
Item_2	20.6000	26.253	.611	.871
Item_3	20.7000	25.695	.642	.868
Item_4	20.3500	24.555	.796	.852
Item_5	20.3500	28.134	.519	.879
Item_6	20.5000	24.474	.765	.855
item_7	20.7000	23.274	.737	.858
Item_8	20.6000	23.726	.716	.860

After meting out the results obtained with the Amos 18.0 are tabularized in Table 3.

Table 3: Regression Weights: (Group number 1 - Default model).

No			Estimate	S.E.	C.R.	P	Label
Item 1	<---	Motivation level	1.000				
Item 2	<---	Motivation level	2.161	1.777	1.216	.224	
Item 3	<---	Motivation level	2.034	1.666	1.221	.222	
Item 4	<---	Motivation level	2.703	2.096	1.290	.197	
Item 5	<---	Motivation level	1.158	1.032	1.123	.262	
Item 6	<---	Motivation level	3.037	2.390	1.270	.204	
Item 7	<---	Motivation level	3.638	2.854	1.275	.202	
Item 8	<---	Motivation level	3.629	2.876	1.262	.207	
Item 9	<---	Motivation level	-.143	.636	-.224	.823	
Item 10	<---	Motivation level	-.531	.752	-.706	.480	

Regression Weights in Table 3 shows that the 10 indicators has a P value not significant due to greater than the value 0.05. Thus it can be stated that the item 10 did not meet the test of construct validity, so it is not commendable item in the collection of data. It is perceived that conclusions obtained under the Tables 1 and 2 conflicts with Table 3. This indicates that the validity criterion is not sufficient to conduct the research.

Case II:

In order to test the validity of precision level of the policy instruments, data has been collected from 345 Students. Results of correlation were enumerated and itemized in Table 4 and 5.

In Table 4 and 5 the value of Cronbach's Alpha for the whole items was found to be 0.902, which means that the instrument has a high level of consistency (above 0.85). In addition, the value of Cronbach's Alpha for Item deleted was too high (above 0.85). Thus it can be stated that the instrument with a crumb item 1 to item 20 has high consistencies, or fit for use in data collection. The above fallouts were compared with Confirmatory Factor Analysis (CFA) in the Structural Equation Modeling (SEM) by means of Amos 18.0 program.

After dispensation the data via Amos 18.0, results are tabulated in Table 6 and 7.

Regression Weights in the Table 6 shows that all 20 indicators have a significant P value being smaller than 0.05 (mark *** indicates figures that are much smaller than 0.05). But in the Standardized Regression Weights in Table 7, there are eight indicators that have factor loading smaller than 0.5, which are X14, X19, X115, X116, X117, X118, X119, and X110. Therefore, these eight indicators have to be detached, and then scrutinize again. Accordingly the upshots of Table 4 and 5 are yet again dissimilar from the outcomes of Table 6 and 7. In Table 4 and 5, all items are declared invalid, but in tables 6 and 7 there are eight items that are not valid. In addition, the reliability level of instruments based on the schedule 4 is 0.902, while calculated construct reliability was 0.899. These results indicated that the calculations of the validity criteria are not strong enough to declare that the instruments valid and reliable. If data collection is done using an instrument base on the schedules 4 and 5, then results will be biased.

Consequently Confirmatory Factor Analysis (CFA) in the Structural Equation Modeling (SEM) gave well again results in testing the validity and reliability of the instrument. Besides this it is apparent that testing with the CFA should be done after the completion of data collection.

Table 4: Reliability Statistics caseII.

Cronbach's Alpha	N of Items
.902	20

Table 5: Item-Total Statistics case II.

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
X11	74.8841	67.643	.505	.898
X12	74.8377	67.293	.557	.897
X13	74.8841	66.888	.559	.897
X14	75.0029	69.154	.470	.899
X15	74.6174	68.940	.566	.897
X16	74.6174	67.411	.651	.895
X17	74.6493	68.031	.595	.896
X18	74.7420	67.419	.599	.896
X19	75.3623	68.650	.366	.903
X110	75.0783	67.921	.509	.898
X111	75.2899	64.805	.664	.893
X112	75.2377	66.048	.621	.895
X113	74.8754	66.918	.629	.895
X114	75.1739	66.708	.599	.896
X115	74.6928	68.789	.424	.900
X116	75.4754	68.558	.416	.901
X117	75.2957	67.970	.491	.898
X118	75.0493	66.547	.543	.897
X119	74.9826	66.732	.532	.898
X120	74.6551	69.791	.457	.899

Table 6: Regression Weights: (Group number 1 – Default model).

No			Estimate	S.E.	C.R.	P	Label
X11	←-	Policy	1.000				
X12	←-	Policy	.976	.109	8.977	***	par_1
X13	←-	Policy	1.094	.116	9.465	***	par_2
X14	←-	Policy	.731	.092	7.985	***	par_3
X15	←-	Policy	.885	.091	9.751	***	par_4
X16	←-	Policy	1.126	.106	10.629	***	par_5
X17	←-	Policy	1.064	.103	10.292	***	par_6
X18	←-	Policy	1.106	.111	9.967	***	par_7
X19	←-	Policy	.614	.118	5.200	***	par_8
X110	←-	Policy	.892	.108	8.234	***	par_9
X111	←-	Policy	1.191	.136	8.746	***	par_10
X112	←-	Policy	1.064	.124	8.606	***	par_11
X113	←-	Policy	1.055	.110	9.584	***	par_12
X114	←-	Policy	.974	.119	8.212	***	par_13
X115	←-	Policy	.752	.106	7.116	***	par_14
X116	←-	Policy	.603	.111	5.438	***	par_15
X117	←-	Policy	.754	.108	6.995	***	par_16
X118	←-	Policy	.947	.126	7.509	***	par_17
X119	←-	Policy	.935	.125	7.496	***	par_18
X120	←-	Policy	.664	.085	7.779	***	par_19

Table 7: Standardized Regression Weights: (Group number 1 – Default model)

No		Variable	Estimate
X11	←-	Policy	.573
X12	←-	Policy	.580
X13	←-	Policy	.617
X14	←-	Policy	.500
X15	←-	Policy	.686
X16	←-	Policy	.779
X17	←-	Policy	.744
X18	←-	Policy	.711
X19	←-	Policy	.314
X110	←-	Policy	.536
X111	←-	Policy	.614
X112	←-	Policy	.594
X113	←-	Policy	.662
X114	←-	Policy	.571
X115	←-	Policy	.443
X116	←-	Policy	.337
X117	←-	Policy	.443
X118	←-	Policy	.500
X119	←-	Policy	.496
X120	←-	Policy	.501

Conclusions:

On the basis of calculations, it can be perceived that, Validity is a measure of consistency of questioned items of an instrument, so the questioned items are strongly believed to be able to measure what is to be measured.

Reckoning s of the validity criteria is not strong enough to declare that the instrument is valid and reliable. Reliability is an evenness of an instrument, so firmly believed that the instrument is capable of providing a steady data (fixed), although given at different times to the same respondents.

Confirmatory Factor Analysis (CFA) in the Structural Equation Modeling (SEM) gives better results in testing the validity and reliability of an instrument. The test results can be indicated by; Regression Weights, Standardized Regression Weights, Convergent Validity, Variance Extracted, Construct Reliability, and Discriminant Validity. Also, testing with the CFA should be carried out after a data collection.

REFERENCES

Agung, I.G.N., 2008. Simple quantitative analysis but very important for decision making in business and management. International Conference of Business and management Research (ICMBR), 27-29 Agustus, Indonesia.

Arbuckle, J.L., 2010. Amos 18.0 User's Guide. USA: Amos Development Corpororation.

Cabrera, P., 2010. Author Guidelines for Reporting Scale Development and Validation Results in the Journal of the Society for Social Work and Research, 1(2): 99-103.

Creswell, J.W., 2008. Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research. New Jersey: Pearson Prentice Hall.

Dimitrov, D.M., 2003. Validation of Cognitive Structures: A Structural Equation Modelling Approach. Multivariate Research, 38(1): 1-23.

Ferdinand, A., 2002. Structural Equation Modelling dalam Penelitian Manajemen. Semarang: Undip.

Gay, L.R., G.E. Milss and P. Airasian, 2006. Educational Research: Competencies for Analysis and Applications. New Jersey: Pearson Prentice Hall.

Ghozali, I., 2004. Model Persamaan Struktural: Konsep dan Aplikasi dengan Program Amos 16.0. Semarang: Badan Penerbit - Undip.

Hair, *et al*., 1998. Multivariate Data Analysis. 7^{Th} edition. New Jersey: Prentice Hall.

Hernandez, R., 2010. A Short Form of the Subjective Well-Being Scale for Filipinos. Educational Measurement and Evaluation Review, 1: 105-115.

Hisyam., 2010. Pengaruh Kualitas dan Biaya Jasa Terhadap Kepuasan dan Loyalitas Mahasiswa. Makassar: Endeh Ofset.

Hwang, W.Y., C.B. Chang, G.J. dan Chen, 2004. The Relationship of Learning Traits, Motivation and Performance-Learning Response Dynamics. Computers and Education, 42: 267-287.

Idris, R., S.R. Ariffin, N.M. dan Ishak, 2010. Hierarki Model Pengukuran Confirmatory Factor Analysis (CFA) ke Atas Instrumen Kemahiran Generik. Jurnal Kemanusiaan bil., 16.

James, B.B., F.K. Stage, F.K., J. King, A. Nora, 2006. Reporting Structural Equation Modeling and Confirmatory Factor Analysis Results: A Review. The journal of education research, Heldref Publications, 99(6): 323-337.

Jung, J., *et al*., 2010. Validation of the "SmoCess-GP" instrument - a short patient questionnaire for assessing the smoking cessation activities of general practitioners: a cross-sectional study. BMC Family Practice, 11: 9.

Kerlinger, F.N., and H.B. Lee, 2000. Foundations of behavioral research (4th ed.). Holt, NY: Harcourt College Publishers.

Kline, R.B., 2005. Principles and practice of structural equation modeling (2nd Ed.). New York: Guildford Press.

Lantin, A.J.P. and A.K.M. Sangalang, 2009. A Belief Scale on Cooperative Learning. The Assessment Handbook, 2: 23-37.

Lawson, A.B., L. Willoughby, K. dan Logossah, 2010. Developing an instrument for measuring e-commerce dimensions. Journal of Computer Information Systems. Winter.

Légaré, F., *et al*., 2011. Developing a theory-based instrument to assess the impact of continuing professional development activities on clinical practice: a study protocol. Implementation Science, 6: 17.

Linn, R.L., N.E. Grondlund, 2000. Measurement and Assessment In Teaching . Eighth edition. New Jersey: Merril an imprint of Prentice Hall.

Martin, N.K. D.A. dan Sass, 2010. Construct validation of the Behavior and Instructional Management Scale. Teaching and Teacher Education XXX, 1-12.

Meihan, L., W.N. dan Chung, 2011. Validation of the psychometric properties of the health-promoting lifestyle profile in a sample of Taiwanese women. Qual Life Res., 20: 523-528.

Shulamn, L.S., 1987. Knowledge and teaching, foundation of the new reform. Harvard educational Rev., 57: 1-22.

Siniscalco, M.T., N. dan Auriat, 2005. Quantitative Research Methods in Educational Planning. Paris: UNESCO.

Stewart, C.D., 2009. A Multidimensional Measure of Professional Learning Communities: The Development and Validation of the Learning Community Culture Indicator (LCCI). Disertation. Brigham Young University.