

Confirmatory Factor Analysis of the School-Based Assessment Evaluation among Teachers

Nor Hasnida Md Ghazali¹

Norazilawati Abdullah²

Mahizer Hamzah³

Norfishah Mat Rabi⁴

¹Senior Lecturer, Faculty of Education and Human Development, Sultan Idris Education University, Malaysia
hasnida@fppm.upsi.edu.my

²Senior Lecturer, Faculty of Education and Human Development, Sultan Idris Education University, Malaysia
nora@fppm.upsi.edu.my

³Senior Lecturer, Faculty of Education and Human Development, Sultan Idris Education University, Malaysia
mahizer@fppm.upsi.edu.my

⁴Senior Lecturer, Faculty of Education and Human Development, Sultan Idris Education University, Malaysia
norfishah@fppm.upsi.edu.my

Doi:10.5901/mjss.2016.v7n5p114

Abstract

The school-based assessment (SBA) system is a holistic assessment system that is conducted in schools by subject teachers in assessing the students cognitive (intellect), affective (emotional and spiritual) and psychomotor (physical) aspects. In order to evaluate the implementation of SBA, a measurement scale was validated. The aim of this paper was to explore different factor structures of the SBA evaluation scale by using the second-order Confirmatory Factor Analysis. Questionnaire was used as an instrument for data collection. 776 primary and secondary school teachers were selected as respondents using stratified random sampling. Results indicated that the SBA evaluation model was a valid and reliable scale. The input measurement model was validated with two factors ('personnel qualifications' and 'physical infrastructure'), process measurement model was validated with six factors ('attitude', 'understanding', 'skills', 'challenges', 'moderation' and 'monitoring') and product measurement model was validated with two factors ('students' attitude' and 'students' motivation'). This study provides support for using a valid instrument in evaluating the implementation of SBA in schools. Furthermore, CFA procedures used supported the conceptual framework set out earlier. Thus, it presents clearly the importance of evaluation of any system to follow all the dimensions outlined in the evaluation model by Stufflebeam.

Keywords: school-based assessment, evaluation scale, confirmatory factor analysis, exploratory factor analysis

1. Introduction

According to the aspiration of the Malaysia National Education Philosophy, education in Malaysia is supposed to be an on-going effort towards further developing the potentials of individuals in a holistic and integrated manner so that well-balanced individuals can be produced. In order to put on such an effort, reform has to be formulated in our education system. Recently, the assessment system in Malaysian education has been reviewed and the inference is that previously, the assessment was only focusing on summative type where public examinations were implemented to all students in Year six, nine and eleven (Ong, 2010). In the recent years, formative assessment has been introduced to certain subjects at certain level of schooling in all government schools.

A traditional concept of assessment has been found to be less effective in improving students' learning. This is because it focuses more on public examination which has rendered the students to become examination-oriented (William, 2001) and the assessment only evaluates the students purely on their academic achievement based on knowledge and skills in a very time-limited situation (Fan, 2011). It is also seen as negatively affecting students' emotion and confidence levels (Stiggins, 2005) hence, it produced more passive students and teachers (Mercurio, 2008). Similarly, the Malaysian public examination is a method that orientates the public to focus on the examination (Cheah, 2010). As such, a new system of assessment which is capable in determining the full potential of students and improving

students' learning is greatly needed. This is why formative assessment is becoming more and more popular these days. In general, related to SBA, there are two main forms of assessment which are formative SBA and summative SBA. Formative SBA is an assessment to promote students' learning and it is school-based (Lembaga Peperiksaan Malaysia, 2011). It is conducted in line with the teaching and learning process using various methods of gathering information such as worksheet, observation, quiz, check list, assessment report, homework or test. By contrast, a summative SBA is an assessment which is also school-based providing a record of a student's overall achievement at the end of the month, semester or year using monthly or semestered testing (Harlen, 2004; Lembaga Peperiksaan Malaysia, 2011).

SBA, which is now being implemented in Malaysia has two main objectives, which are to gain an overall picture of an individual potentials and also to have a meaningful reporting on individual learning (Nor Hasnida, 2015). It includes both types of assessment; formative and summative (Lembaga Peperiksaan, 2010). Furthermore, SBA which focuses more on formative assessment rather than summative has been conducted in countries like Australia, New Zealand, Hong Kong, Finland, United Kingdom, USA, Canada, Africa, Sweden, Scandinavia and Singapore (Assessment Support Material, 2001). Australia has implemented SBA in the late 1960s (Mercurio, 2008) while Finland and Sweden had it in the early 1970s (Darling-Hammond and McCloskey, 2008). Malaysia has taken an astonishing decision when SBA has been formally implemented in all the government schools since 2011, with the Year one students becoming the first batch of students to undergo the SBA enactment. The need to have a valid and reliable measurement model to evaluate the implementation of the system is becoming increasingly important. Hence, the instruments used to assess the teachers' perception about a particular concept need to be evaluated first before administering. This is to make sure that the questionnaire used is valid and reliable or in other words, it is measuring what it is supposed to be measured and that the extent to which test scores are free of measurement error (Muijs, 2011). There is a valid and reliable instrument developed previously but it is on teachers' attitude towards SBA (Nor Hasnida *et al.*, 2012). Validity and reliability of the questionnaire are the most important things to consider when dealing with measurement (Barroon and Abd Rahman, 2015). And, the relationship between the two is that, any test can be reliable without being valid but it cannot be valid if it is not reliable (Jackson, 2003). There are various types of reliability but in this study, three types of reliability are considered which are internal reliability, construct reliability (CR) and average variance extracted (AVE), whereas in validity aspect, there are convergent validity, construct validity and discriminant validity. Internal reliability is a concept referring to the degree to which all of the items are measuring the same underlying construct (Pallant, 2007) whereas construct reliability is a concept to assess the extent to which a measuring instrument accurately measures a theoretical construct that it was designed to measure (Jackson, 2003). Construct validity is the extent to which a set of items actually reflect the theoretical latent construct those items are designed to measure (Hair *et al.*, 2006) whereas discriminant validity is a concept where individual measured items should represent only one latent construct.

When a questionnaire is valid and reliable, a researcher will have confidence in the results obtained using those questionnaires during data collection. Hence, the purpose of this study is to develop an instrument to evaluate teachers' perception on the factors concerning SBA implementation by exploring the different factor structures of the evaluation scale by using the second-order Confirmatory Factor Analysis.

1.1 Theoretical Underpinning of the Study

Following the CIPP Model developed by Daniel Stufflebeam, any system could be evaluated in four dimensions – context, input, process and product (Stufflebeam, 1971a). In evaluating a system, Stufflebeam believes that one can either look at one dimension at a time or at all the dimensions simultaneously (Stufflebeam and Shienkfield, 1985). Furthermore, each dimension in evaluation should contribute to a certain decision or conclusion as evaluation involves decision-making process (Stufflebeam, 1971a). Thus, context, input, process and product evaluation should serve planning, structuring, implementing and recycling decisions respectively. Or, in other words, context evaluation should be acquiring information like 'Were important needs addressed?', meanwhile input evaluation should be asking, 'Was the effort guided by a defensible plan?'. Furthermore, process evaluation should be acquiring, 'Was the service design executed competently?' and product evaluation should be providing us with the overall conclusion whether or not the effort thrive, ('Did the effort succeed?') (Stufflebeam, 2003). In this study, context evaluation includes two factors which identified as school type (urban and rural schools) and school category (primary and secondary schools). Input evaluation includes three 1st-order factors ('material and personal needs', 'personal qualifications' and 'physical infrastructure'), process evaluation includes ten 1st-order factors ('attitude', 'understanding', 'courses', 'in-house training', 'administration', 'challenges', 'moderation', 'monitoring', 'role of SBA' and 'importance of SBA') and product evaluation includes three 1st-order factors ('students' attitude, knowledge and motivation').

All the factors involved in this implementation of SBA are supported by learning theories such as behaviorism,

Piaget learning theory, constructivism, multiple intelligences theory and brain research. In general, these learning theories agree to some extent that the implementation of SBA in classroom could improve students' learning. In other words, formative and summative assessments in SBA seem to fit all the above learning theories. There is a model that explains how the interaction between staff and students in a dynamic relationship produces an effective assessment practice. It is developed by Christ Rust and his colleagues from the Oxford Center for Staff and Learning Development (Rust *et al.*, 2005) based on the social constructivist approach. The model shows that a constructive assessment process needs a well-defined explicit assessment criteria followed by an active engagement of the students and teachers with the criteria. Active engagement with feedback is also important because feedback lies at the heart of assessment process. The assessment process also includes marking and moderation process by the teachers. The advantage of this model is due to its relevant application to both assessments i.e. formative and summative and there is a dynamic relationship between elements in assessment practices which represents the true picture of what really happens in the schools in the corporeal world. However, the problem with this model is that it is based solely on the social constructivist approach where the approach itself might have certain weaknesses.

2. Research Methods

Survey questionnaire are distributed using postal mail and by-hand to the primary and secondary schools in ten major districts in Kelantan, a state in the north-east of Peninsular Malaysia. Teachers are selected as respondents because they are the most involved and the most concerned with the system compared to other parties. A total of 776 usable questionnaires were obtained for analysis. This sample size has met the suggested recommendation by Kline (2005) as he suggested that a sample size of more than 200 participants is enough to run SEM analysis. Similarly, 500 participants are regarded as a minimum sample size required for a study involving more than seven latent constructs with some constructs that have less than three items (Hair *et al.*, 2010). The issues of uni-dimensionality, reliability and validity for all measurement models are determined. Uni-dimensionality is achieved when the factor loading of items for the respective latent construct is 0.5 or more (Zainuddin, 2012). Three types of reliability are considered, they are internal reliability, construct reliability (CR) and average variance extracted (AVE), whereas in validity aspect, there are also three categories of validity determined namely convergent validity, construct validity and discriminant validity are determined. The requirements are shown in Table 1.

Table 1. Requirement for the reliability and validity of the measurement model

Type	Requirement
Internal reliability	Cronbach Alpha ≥ 0.7
Construct reliability (CR)	$CR \geq 0.6$; $CR = (\sum K)^2 / [(\sum K)^2 + (\sum 1-K^2)]$; (K=factor loading of every item)
Average variance extracted (AVE)	$AVE \geq 0.5$
Convergent validity	$AVE \geq 0.5$; $AVE = \sum K^2 / n$ (K=factor loading of every item; n=number of items in a model)
Construct validity	Fitness index is achieved as: $GFI \geq 0.90$; $CFI \geq 0.90$; $RMSEA \leq 0.08$; $Chisq/df < 5.0$
Discriminant validity	Free from redundant items; Correlation coefficient between each pair of latent construct ≤ 0.85

3. Statistical Analysis

In this study, AMOS version 18 and SPSS version 21 are used to facilitate the result analysis. AMOS software is used in assessing the relationship between latent and observed variables of a measurement model. The technique used is called a confirmatory factor analysis. It is a theory-driven technique which determines the goodness-of-fit between the model and the sample data (Byrne, 2010). This type of analysis is preferable when the researcher has had some knowledge about the latent structure. In this study, maximum likelihood estimation method is used in generating parameter estimates of the measurement models. This estimation method is more practical due to its ability to deal with complex models and also its robustness to non-normality data (Brown, 2006). There are a few fit indices used in this study to discern how well the specified model reproduces the covariance matrix among the indicator items (Hair *et al.*, 2006). They are grouped under three main groups of measures; practical fit measures (chi-square statistics or X^2/df), absolute fit indices (GFI, AGFI or RMSEA) and incremental fit indices (TLI or CFI). According to Hair *et al.* (2010), a study should report at least

three fit indices with at least one from each category. In addition, the accepted values listed in Table 2 have to be fulfilled if we were to gain a good or perfect fit model.

Table 2. Goodness-of-fit indices

Goodness-of-fit Index	Acceptable Value	Comments
Chi-square (χ^2)	$p > 0.05$ (non-significant)	Indicates exact fit of the model. Value is sensitive to large sample size
Practical indices of fit:		
Normed chi-square (χ^2/df)	[2.00, 5.00]	This is to reduce the sensitivity of χ^2 to sample size $\chi^2/df < 3.0$: good fit
Absolute fit index:		
The Goodness-of-Fit Index(GFI)	[0.00, 1.00]	GFI = 1.00: perfect fit GFI > 0.9: good fit
The Adjusted GFI (AGFI)	[0.00, 1.00]	Values close to 1.00: good fit GFI > 0.9: good fit
Root Mean Square Error of Approximation (RMSEA)	RMSEA \leq 0.08	RMSEA < 0.05: good fit RMSEA 0.05 - 0.08: adequate fit Values up to 0.10: poor fit
Incremental fit indices:		
Normed Fit Index (NFI)	NFI \geq 0.90	NFI = 1.00: perfect fit Values close to 0.00: poor fit
Comparative Fit Index (CFI)	CFI \geq 0.90	0.00 > CFI > 1.00 for acceptance
Tucker-Lewis Index (TLI)	TLI > 0.90	0.0 > TLI > 1.00 for acceptance

4. Results

Nearly two-third (74.7 percent) of the participants are females and one-third (24.6 percent) are males. The majority (93.6 percent) of them are Malays. Nearly half of them have had 10 to 20 years of teaching experience. Overall, most of them have experienced practising SBA in the range of 0 to 3 years.

4.1 Input Evaluation

The Input evaluation as a 2nd-order measurement model is proposed to measure personnel, resources and procedures in achieving SBA objectives (Stufflebeam, 1971a). Three factors are involved, known specifically as material and personal needs ('mat'), appropriateness of qualification ('appr') and suitability of physical infrastructure and ICT ('suit'). These factors are measured by three items, two items and three items respectively as shown in Figure 1 (initial model). A total of eight items are used to measure input evaluation. The model yields a Chi-square (χ^2) statistic of 157 with 756 on 17 degrees of freedom. The model was over-identified but with hierarchical model, the higher-order structure would be just-identified. To resolve just-identification issue, equality constraints are placed on particular parameters to yield a more accurate estimate. Reviewing the goodness of fit statistics, it shows that $\chi^2/df=9.280$; GFI=0.952; AGFI=0.898; NFI=0.928; CFI=0.935; TLI=0.892 and RMSEA=0.103. This measurement model provides a poor fit and thus, modification such as deleting a construct or items is later conducted to gain a better fit. It follows with a determination of modification indices values to correlate the measurement error between items. According to Arbuckle and Wothke (1999), these have to be done by considering a theoretical or common sense to avoid producing an absurd parameter estimate.

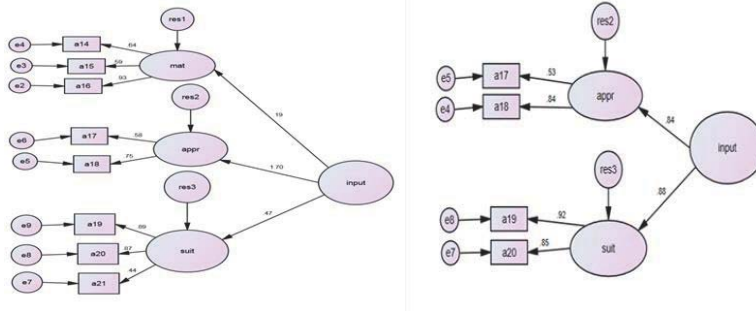


Figure 1: Measurement model for input evaluation (initial model [left] and final model [right])

Table 3. Input Evaluation Items and their descriptions

Original Item	Item Label	Item Deleted
Material and personal needs:		
It is easy to implement if teachers are supplied with a complete assessment document	a14	Deleted
A teaching assistant is needed to help teachers in assessment	a15	Deleted
Training of the personnel involved should be properly planned and implemented	a16	Deleted
Appropriateness of qualification:		
Teachers are sufficiently qualified to implement assessment activities	a17	Maintained
Teachers' skills in assessment are adequate for implementing assessment activities	a18	Maintained
Suitability of physical infrastructure and ICT:		
The physical equipment in classroom is suitable to conduct SBA activities	a19	Maintained
The space can be adapted to assessment activities	a20	Maintained
The ICT hardware is suitable to conduct SBA activities	a21	Deleted

For final measurement model (Figure 1), four items are left to measure input evaluation. List of remaining items are as listed in Table 3. These remaining items (a17, a18, a19 and a20) have factor loadings ranging from 0.53 to 0.92 indicating the meaning of the factors that have been preserved. Reviewing the benefit of fit statistics, this final measurement model indicates a very good fit (as in Table 6). Finally, the issues of uni-dimensionality, validity and reliability have been addressed and are shown in Table 5.

4.2 Process Dimension

The process evaluation as a 2nd-order measurement model is proposed to measure the process implemented in achieving the objectives of the programme (Stufflebeam, 1971a). There are twelve major constructs proposed – belief, feeling, readiness, understanding, skills, in-house training, administration, moderation, monitoring, challenges, role and the importance of SBA with a total of fifty-two items. When this measurement model is run, the result shows that it does not fit the implementation process. Therefore, the principal component analysis (PCA) and confirmatory factor analysis (CFA) technique are conducted. The models have also been modified based on theory. Finally, four measurement models are produced – process1 with three 2nd-order constructs which identified as attitude, understanding and courses (skills), process2 with two 2nd-order constructs which were moderation and monitoring, process3 with two 2nd-order constructs identified as role and importance (crucial) of SBA and last of all, is the 1st-order construct which is challenges. Model modification has been applied to get the most fitted models. The final measurement models for process dimensions are shown in Figure 2 and the issues of uni-dimensionality, validity and reliability are addressed in Table 5.

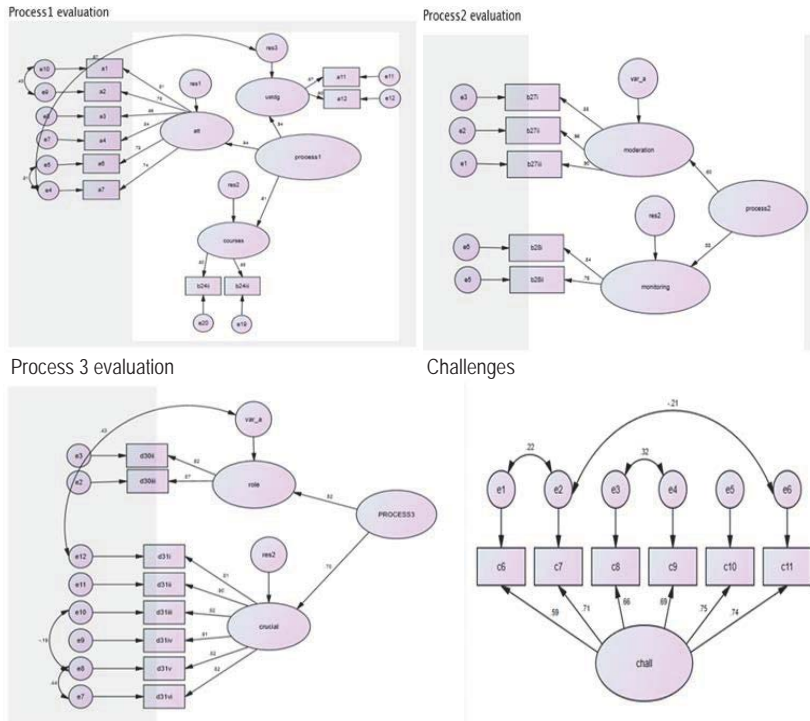


Figure 2: Four measurement models for process evaluation (final model)

4.3 Product Dimension

The product evaluation as a 2nd-order measurement model is proposed to measure the program outcomes. Three factors are taken into consideration and they are students' attitude towards SBA ('att'), students' knowledge in SBA ('know') and students' motivation towards learning ('mot'). These factors are measured by three items, two items and three items respectively. A total of eight items are used altogether. The model yields a Chi-square (χ^2) statistic of 138 with 876 on 17 degrees of freedom. Reviewing the goodness of fit statistics, it shows that $\chi^2/df=8.169$; GFI=0.960; AGFI=0.915; NFI=0.971; CFI=0.974; TLI=0.958 and RMSEA=0.096. This measurement model presents a poor fit, hence model modification such as deleting a construct or items was conducted to achieve a better fit. This follows with a determination of modification indices values to correlate the measurement error between items. As the consequence, there are five items left as shown in Table 4.

Table 4. Product Evaluation Items and their descriptions

Original Item	Item Label	Item Deleted
Students' attitude towards SBA:		
Students practice SBA in their study	e32i	Maintained
Students are motivated to prepare for SBA	e32ii	Maintained
Practising SBA in their study is NOT burdening them at all	e32iii	Deleted
Students' knowledge of SBA:		
SBA tasks are part of teaching and learning of their respective subjects	e33i	Deleted
Feedback from teachers helps them develop skills that may not be reflected in public exams	e33ii	Deleted
Students' motivational towards learning:		
SBA encourages students to read more books than before	e34i	Maintained
Students are becoming more interested in my subject than before	e34ii	Maintained
SBA helps students to understand more on their strengths and weaknesses	e34iii	Maintained

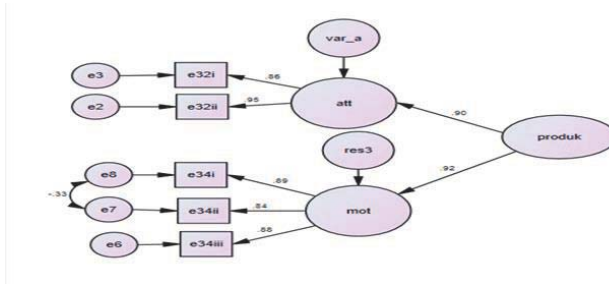


Figure 3: Measurement model for product evaluation (final model)

For final product measurement model (Figure 3), five items remain to measure product evaluation. The remaining items (e32i, e32ii, e34i, e34ii and e34iii) have quite a high factor loading ranging from 0.84 to 0.95 indicating the meaning of the factors has been preserved. Reviewing the goodness of fit statistics, it shows that the measurement model indicates a very good fit (as shown in Table 6). Finally, the issues of uni-dimensionality, validity and reliability are addressed in Table 5.

Table 5. The CFA results reporting for the measurement model

2 nd -order factor	1 st -order factor	Standardized factor loading	Item	Standardized factor loading	CR Alpha (>0.7)	CR (>0.6)	AVE (>0.5)						
Input evaluation	<i>Appr</i>	0.925	a17	0.523	0.600	0.635	0.478						
			a18	0.826									
	<i>Suit</i>	0.824	a19	0.924	0.752	0.813	0.604						
			a20	0.839									
Process1	<i>Attitude</i>	0.998	a1	0.768	0.908	0.907	0.498						
			a2	0.735									
			a3	0.763									
			a4	0.787									
			a6	0.753									
			a7	0.813									
	<i>Understanding</i>	0.962	a11	0.870	0.679	0.673	0.429						
			a12	0.601									
	<i>Skills</i>	0.614	b24ii	0.810	0.813	0.902	0.698						
			b24iii	0.849									
Process2	<i>Moderation</i>	0.623	b27i	0.827	0.915	0.831	0.788						
			b27ii	0.942									
			b27iii	0.891									
	<i>Monitoring</i>	0.581	b28i	0.815	0.758	0.763	0.521						
			b28ii	0.734									
	<i>Challenges</i>	-	c4	0.630	0.889	0.882	0.455						
			c5	0.662									
			c6	0.658									
			c7	0.668									
			c8	0.641									
			c9	0.655									
			c10	0.726									
			c11	0.703									
			c12	0.722									
			Product evaluation	<i>Attitude</i>				0.924	e32i	0.855	0.861	0.867	0.687
									e32ii	0.894			
				<i>Motivation</i>				0.958	e34i	0.829	0.885	0.889	0.727
e34ii	0.796												
e34iii	0.928												

Table 6 shows the characteristics of the six final measurement models. In general, the fitness indices values are identified to be the most well-fitting; all un-standardized estimates are statistically significant given critical values more than 1.96; all

standard errors are in good order; all standardized estimates are above the moderate strength and the multivariate kurtosis value has improved and has achieved the required level. All multivariate kurtosis values is less than 50.0 indicated a multivariate normality distribution of data set. However, there is a high correlation value between process3 and product ($r=0.939$) and also between process3 and process1 ($r=0.923$). It displays a *multi-collinearity* phenomenon, so process3 model has been deleted.

Table 6. Final characteristics of the measurement models

Measurement Model	Fit indices value	Number of items	Multivariate kurtosis
Input Evaluation	$X^2 = 1.751$; $df = 1$; $X^2/df = 1.751$; $GFI = 0.999$; $AGFI = 0.989$; $NFI = 0.999$; $CFI = 0.999$; $TLI = 0.996$; $RMSEA = 0.031$	4	6.104
Process Evaluation			
a) Process1	$X^2 = 122.722$; $df = 30$; $X^2/df = 4.091$; $GFI = 0.969$; $AGFI = 0.943$; $NFI = 0.975$; $CFI = 0.981$; $TLI = 0.971$; $RMSEA = 0.064$	10	41.606
b) Process2	$X^2 = 11.067$; $df = 4$; $X^2/df = 2.767$; $GFI = 0.994$; $AGFI = 0.979$; $NFI = 0.995$; $CFI = 0.997$; $TLI = 0.993$; $RMSEA = 0.048$	5	39.925
c) Process3	$X^2 = 65.893$; $df = 16$; $X^2/df = 4.118$; $GFI = 0.979$; $AGFI = 0.952$; $NFI = 0.988$; $CFI = 0.991$; $TLI = 0.984$; $RMSEA = 0.064$	8	40.738
d) Challenges	$X^2 = 9.930$; $df = 6$; $X^2/df = 1.655$; $GFI = 0.996$; $AGFI = 0.985$; $NFI = 0.994$; $CFI = 0.998$; $TLI = 0.994$; $RMSEA = 0.029$	6	12.251
Product Evaluation	$X^2 = 3.458$; $df = 3$; $X^2/df = 1.153$; $GFI = 0.998$; $AGFI = 0.991$; $NFI = 0.999$; $CFI = 1.000$; $TLI = 0.999$; $RMSEA = 0.014$	5	12.123
TOTAL		38	

5. Discussion

Literatures have been reviewed to look for gaps in the existing SBA implementation research. Most evaluation processes look only at some dimensions which do not give a fully rounded indication of the effectiveness of the system implemented such as looking at teachers' attitude towards SBA (Majid, 2011), teachers' leadership (Boon and Shaharuddin, 2011), teachers' knowledge and best practises in SBA and few more. To date, studies that combine all the four dimensions of evaluation are non-existent. Therefore, in this study, the psychometric properties of an instrument is developed and measured. Selecting a validated instrument is easy but to get an instrument which suits the study objectives and the study context is quite difficult. In this case, it is to develop and validate an instrument to measure teachers' perception towards SBA implementation in the Malaysian context. Finding a validated instrument for this purpose is not easy as so many factors need to be considered in this context. Evidence has shown that the final model on the evaluation of SBA consisted of five factors (input, process1, process2, challenges and product evaluation). The model is hierarchical, so there are first and second-order factors involved. Input comprises of two first-order factors (personnel's qualification and physical infrastructure and ICT), process1 consists of three 1st-order factors (attitude, understanding and skills), process2 constitutes of two 1st-order factors (moderation and monitoring process), challenges consists of six strongly loading items and product is made up of two 1st-order factors (students' attitude and students' motivational towards learning). As all the values of fitness indices are the most well-fitting, all un-standardized estimates are statistically significant, all standard errors are in good order and all standard estimates are above the moderate strength, this result implies good reliability and validity of the instrument. Hence, the questionnaire is suitable to assess the perception of school teachers on the SBA system implementation in schools in the Malaysian context.

6. Conclusion

The psychometric properties of a new extended SBA evaluation scale for assessing teachers' perception of the SBA system are presented. The instrument was developed after reviewing relevant literatures and consulting experts' in measurement and evaluation. The findings demonstrate that the instrument has adequate psychometric properties (valid

and reliable) and is fit to be used for the main study as it was tested with quite a large sample size and has been analyzed using CFA. Furthermore, CFA procedures used in this study supported the conceptual framework set out earlier. Thus, it presents clearly the importance of evaluation of any system to follow all the dimensions outlined in the evaluation model by Stufflebeam. Hence, the findings of this present study have expanded the existing body of knowledge on the development of a measurement scale to evaluate the SBA system.

Nevertheless, this study has several limitations. First, the samples were taken only from teachers and not from other stakeholders, and therefore the development and validation of instruments might be limited. Furthermore, data comes only from the perceptions of the teachers without observing their real practices. Secondly, items included in the survey have been deleted during CFA procedure. Deletion of the items is needed to make sure that the models would be fit and yet considering the hypothesized models are acceptable but there might be other variables which are more influential than those we have chosen. In addition, there might also be other models which may fit the data that we have not tested. Finally, the sample of this study has only been collected at the government schools in one of the states in the north-east of Malaysia. Although the education system might not have much difference between each country, the cultural difference might limit the generalization of findings to other states.

The model reported here might be useful in the Asian context especially in those countries that are becoming more and more interested in making assessment system further aligned to classroom learning, providing effective feedback and validly describing students' learning. This current instrument could be such a great value for them. On the contrary, for those countries which are still examination-oriented in their assessment system, it is expected that there will be a higher disagreement and discrepancies among the teachers in accepting SBA as the teachers might see this new way assessment as relevant but not to the extent to improving students' attitude, knowledge and motivation towards learning. The results of this survey indicates that the knowledge of students of SBA are not consistent with the official Malaysian government policy concerning the objectives of the National Education Assessment System on improving students' learning. Certainly, based on this survey, we would expect practitioners or teachers to be exposed more vigorously to some form of professional development, so that they are equipped with enough skills especially on the use of feedback. If the students do not understand the function of feedback in improving their learning, it would be difficult to achieve the desired objectives in the SBA implementation.

References

- Arbuckle, J. L. & Wothke, W. (1999). *Amos 4.0 User's Guide*. Small Waters Corporation: United States of America.
- Assessment Support Material. (2001). Assessment Support Material VCE: English/ESL. Australia. Retrieved from http://www.vcaa.vic.edu.au/vce/Studies/english/English_ASM.doc (November 12, 2011).
- Barroon, I. A. & Abd Rahman, A. (2015). Reliability and Validity of A Questionnaire to Evaluate Diabetic Patients' Intention to Adopt Health Information Technology: A Pilot Study. *Journal of Theoretical and Applied Information Technology*, 2(72), 253-258.
- Boon, Y. and Shaharuddin, M. (2011) *Kepemimpinan Guru Besar Dalam Pelaksanaan Pentaksiran Berasaskan Sekolah (PBS) Di Sekolah Kebangsaan Daerah Kota Tinggi, Johor*. Fakulti Pendidikan. Universiti Teknologi Malaysia. Retrieved from http://eprints.utm.my/11929/1/Kepemimpinan_Guru_Besar_Dalam_Pelaksanaan_Pentaksiran_Berasaskan_Sekolah.pdf (Accessed: 8 March 2012).
- Brown, T. (2006). *Confirmatory factor analysis for applied research*. USA: The Guildford Press.
- Byrne, B. M. (2010). *Structural Equation Modelling with AMOS. Basic concepts, applications and programming*. Mahwah: New Jersey: Earlbum.
- Cheah, U. (2010) 'Assessment in primary Mathematics Classrooms in Malaysia' *Tsukuba International Conference: Innovation of Mathematics Teaching and Learning through Lesson Study - Connection between Assessment and Subject Matter*. Tsukuba, 17-21 February. Retrieved from http://www.criced.tsukuba.ac.jp/math/apec/apec2009/doc/pdf_20-21/CheahUIHock-paper.pdf (Accessed: 11 November 2011).
- Darling-Hammond, L. & McCloskey, L. (2008). Assessment for Learning Around the World: What Would it Mean to be Internationally Competitive?. *Phi Delta Kappan*, 90(4), 263-272.
- Fan, L. (2011). *Performance Assessment in Mathematics: Concepts, Methods, and Examples from Research and Practice in Singapore Classrooms*. Singapore: SG, Pearson.
- Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E. & Tatham, R. L. (2006). *Multivariate Data Analysis. 6th edn*. New Jersey: Prentice-Hall International, Inc.
- Hair, J. F., Black, W. C., Babin, B. J. & Anderson, R. E. (2010). *Multivariate Data Analysis. 7th edn*. Englewood Cliffs, NJ: Prentice Hall.
- Harlen, W. (2004) *A systematic review of the evidence of the impact on students, teachers and the curriculum of the process of using assessment by teachers for summative purposes*, Research Evidence in Education Library. London: EPPI-Center, Social Science Research Unit, Institute of Education.
- Jackson, S. L. (2003) *Research Methods and Statistics, A Critical Thinking Approach*. USA: Thomson Wadsworth.
- Kline, R. B. (2005). *Principles and practice of SEM*. New York: Guilford Press.

- Lembaga Peperiksaan Malaysia. (2011). Panduan dan Peraturan PBS. Putrajaya. [Online] Retrieved from <http://www.scribd.com/faridooi/d/81532201-Panduan-Dan-Peraturan-Pbs-2011>. (November 22, 2011).
- Lembaga Peperiksaan. (2010). Pentaksiran Berasaskan Sekolah. Retrieved from <http://buletinkpm.blogspot.co.uk/2012/11/pentaksiran-berasaskan-sekolah-pbs.html> (November 7, 2011).
- Lembaga Peperiksaan. (2012). Panduan Pengurusan PBS. KPM. Retrieved from <http://www.moe.gov.my/lp/files/pbs/bahan/Buku%20Panduan%20Pengurusan%20Pentaksiran%20Berasaskan%20Sekolah%20%28PBS%29%202012.pdf> (November 26, 2013).
- Majid, F. (2011) 'SBA in Malaysian Schools: The concern of the English teachers', *Journal of US-China Education Review*, 8(10), pp. 1-15.
- Mercurio, A. (2008). Re-imagining school-based assessment at the upper secondary level, SACE Board of South Australia. Retrieved from <http://www.iaea2008.cambridgeassessment.org.uk/ca/digitalAssets/180434-Mercurio.pdf> (November 5, 2011).
- Muijs, D. (2011) *Doing Quantitative Research in Education with SPSS*. London: SAGE Publications Ltd.
- Nor Hasnida M. G., Baharin Y. and Afian A. M. (2012) 'Why Do We Need to Change?. Teachers' Attitude Towards School-Based Assessment System' *SCR London's First International Conference on Social Science and Humanities in the Islamic World*, London, 28-30 May. Available at: <http://bu.univ-ouargla.dz/production%20scientifique/international/2012/PAPER%20LONDON%20CONFERENCE1.pdf> (Accessed: 8 March 2012).
- Nor Hasnida, C. M. G. (2015). An Evaluation of The Implementation of The School-Based Assessment System in Malaysia. (Doctoral Dissertation). Retrieved from <http://eprints.soton.ac.uk/381724/1.hasCoversheetVersion/THESIS%20HASNIDA%20%28P%29.pdf>
- Ong, S. L. (2010) 'Assessment profile of Malaysia: high stakes external examinations dominate', *Assessment in Education: Principles, Policy & Practice*, 17(1), pp. 91-103.
- Pallant, J. (2007) *SPSS Survival Manual: A step by step guide to data analysis using SPSS for Windows, 3rd ed.* New South Wales: Allen and Unwin.
- Rust, C., ODonovan, B. & Price, M. (2005). A social constructivist assessment process model: how the research literature shows us this could be best practice. *Assessment & Evaluation in Higher Education*, 30(3), 231-240.
- Stiggins, R. (2005). From Formative Assessment to Assessment for Learning: A Path to Success in Standards-Based Schools. *Phi Delta Kappan*. 87(4), 324-328.
- Stufflebeam, D. L. & Shinkfield, A. J. (1985). *Systematic evaluation: A self-instructional guide to theory and practice*. Boston: Kluwer Academic Publishers.
- Stufflebeam, D. L. (1971a). The relevance of the CIPP Evaluation Model for Educational Accountability. Annual Meeting of the American Association of School Administrators. Retrieved from <http://eric.ed.gov/PDFS/ED062385.pdf> (September 3, 2012).
- Stufflebeam, D. L. (1971b). Educational evaluation and decision-making. Itasca, Illinois: F.E. Peacock Publisher, Inc.
- Stufflebeam, D.L. (2003). The CIPP Model for evaluation. 2003 Annual Conference of the Oregon Program Evaluators Network (OPEN), Oregon.
- William, D. (2001). What is wrong with our educational assessments and what can be done about it. *Education Review*, 15(1).
- Zainuddin, A. (2012). *A Handbook on SEM Structural Equation Modelling. 6th edn.* Kampus Kota Bharu: UiTM Kelantan.