International Journal of Evaluation and Research in Education (IJERE)
Vol.5, No.2, June 2016, pp. 148~157
ISSN: 2252-8822

A Reliability and Validity of an Instrument to Evaluate the School-Based Assessment System: A Pilot Study

Nor Hasnida Md Ghazali
Faculty of Education and Human Development, Universiti Pendidikan Sultan Idris, Malaysia

Article Info

Article history:

Received Apr 15, 2016
Revised May 25, 2016
Accepted May 29, 2016

Keyword:

CIPP Model
Evaluation
Exploratory factor analysis
School-based assessment

ABSTRACT

A valid, reliable and practical instrument is needed to evaluate the implementation of the school-based assessment (SBA) system. The aim of this study is to develop and assess the validity and reliability of an instrument to measure the perception of teachers towards the SBA implementation in schools. The instrument is developed based on a conceptual framework developed by Daniel Stufflebeam, which is the CIPP (context, input, process and product) Evaluation Model. The instrument in the form of questionnaire is distributed to a sample of 120 primary and secondary school teachers. The response rate is 80 percent. The content validity is assessed by the experts and the construct validity is measured by Exploratory Factor Analysis (EFA). The reliability of the instrument is measured using internal consistence reliability, which is measured by alpha coefficient reliability or Cronbach Alpha. The finding of this pilot study shows that the instrument is valid and reliable. Finally, out of 71 items, 68 items are retained.

Corresponding Author:

Nor Hasnida Md Ghazali,
Faculty of Education and Human Development,
University Pendidikan Sultan Idris, Malaysia,
Tanjong Malim, 35900, Perak, Malaysia.
Email: hasnida@fppm.upsi.edu.my

1. INTRODUCTION

School-based assessment (SBA) is an assessment system which has been introduced to the Malaysian education system in 2011. It is a form of assessment conducted in schools following the procedures from the Malaysian Education Syndicate [1]. It is planned, administered, scored and reported by the students' subject teachers. This is the unique part of SBA compared to public examination. SBA is actually being proposed as a reform movement towards education system following the Tenth-Malaysian Plan (2011-2015) which includes the Government Transformation Program aspiration and the Economic New Model. SBA consists of two types of assessment, assessment for learning and assessment of learning. In the Malaysian context, SBA consists of four components which are school assessment, central assessment, psychometric assessment and physical activity, sports and co-curriculum assessment. The main objectives of the implementation of this new innovation are to get the overall picture of an individual's potential, to monitor individual's development and help to increase students' potential and to make meaningful reporting on individual learning [2]. In addition, SBA is a form of assessment which is focusing on the enhancement of the meaningfulness of assessment by focusing more on students' learning development rather than grade [3]. It also focuses on empowering schools and teachers with quality assessment. To ensure the quality of the assessment implementation, four aspects of quality assurance are conducted. They are the moderating, monitoring, tracking and mentoring processes which must be implemented by all government schools [4]. The theoretical framework is based on the CIPP Evaluation Model developed by Daniel Stufflebeam in 1970s [5]. This model was designed to evaluate programs, projects, personnel, products, institutions or systems from various disciplines such as education field, housing and community development,

transportation, safety and military personnel review systems [6]. Any educational program like tuition, extra classes, co-curriculum activities, seminar, workshops, teachers' recruitment policy, pedagogic strategy or examination and assessment system could also be evaluated using the CIPP Evaluation Model. When evaluating, it is important to find an evaluation which suits our main interests in the study. There are a wide variety of evaluation models with different classification - depending on their major evaluation methodology, target or purpose of evaluation and the responsibility and accountability of evaluation [7]. The main reason why this evaluation model is chosen for this study is because this model is based on the management-oriented evaluation approach which helps the decision-makers to plan, implement and evaluate programs [8]. Furthermore, it is widely used by the evaluators [9] and it covers wide variety of dimensions which could be chosen by the evaluators to best suit their studies [10].

According to Stufflebeam, evaluation involves decision-making, so all the main components of the evaluations (context, input, process and product) serve the decisions (planning, structuring, implementing and recycling) respectively in dynamic actions. Dynamic means that the information gained from any evaluation stages could be provided to any previous stages, so that modifications could be made [11]. Context evaluation gives an opportunity for the decision makers such as the ministry staff, administration group or school leaders to plan the program objectives either to confirm the present objectives, modify it or develop a new one [6]. Input evaluation gives an opportunity to decide on the structure of the programs such as something related to strategies, personnel, resources, procedures or a cost in achieving the programs objectives which have been derived earlier. Next, process evaluation involves making decisions on the implementation of the program such as the program's designs, strategies or action plans. Lastly is product evaluation which involves evaluating the outcomes of the programs. It involves recycling decisions whereby the outcomes of the programs is compared to the objectives of the programs whether to continue the program or not, for example. All of the four processes could be conducted for two main purposes – formatively or summatively.

Although there are various studies have been developed in evaluating SBA implementation, there is lack of studies and instruments that does the evaluation in various dimensions as proposed by the CIPP Evaluation Model. Recently, there are some researches which focus on one dimension of evaluation only such as looking at teachers' attitude towards SBA [12], teachers' leadership towards SBA [13], teachers' knowledge and best practises in SBA [14] or challenges faced by teachers in implementing SBA [15]. There is also research which look at the correlation between the dimensions [16],[17]. So, there is an urgent need to develop an instrument which is able to evaluate SBA in various dimensions of evaluation. Hence, the instrument could then be used to determine the interrelationship between dimensions of evaluation. However, the instrument used to assess teachers' perceptions and attitude about any particular concepts related to SBA need to be first evaluated before it can be administered. This could be done through pilot study. Actually, all studies have to be piloted before the real study is conducted. As [18] asserted that '*almost anything about a social survey can and should be piloted'*. Although pilot study might be a bit tedious, it is very important to check the validity, reliability and practicality of an instrument [19].

An instrument is valid when it is measuring what is supposed to measure [20]. Or, in other words, when an instrument accurately measures any prescribed variable it is considered a valid instrument for that particular variable. There are four types of validity; face validity, criterion validity, content validity or construct validity [20],[21]. Face validity is looking at the concept of whether the test looks valid or not on its surface [21]. Criterion validity is a concept which will be demonstrated in the actual study as to establish it needs 'a good knowledge of theory relating to the concept and a measure of the relationship between our measure and those factors' [20] whereas content validity is looking at the content of items whether it really measures the concept being measured in the study. Finally is the construct validity, which measures the extent to which an instrument accurately measures a theoretical construct that it is designed to measure.

Reliability on the other hand is defined as 'the extent to which test scores are free from measurement error' [20]. It is a measure of stability or internal consistency of an instrument in measuring certain concepts [21]. According to [22], there are various types of reliability depending on the number of times the instruments are administered and the number of individuals who provide information. There are test-retest reliability, alternate forms reliability, alternate forms and test-retest reliability, internal consistency reliability and inter-rater reliability. Test-retest reliability is a form of reliability achieved when the same instrument is administered to the same group of respondents on two different occasions and yet look at the correlation between the two sets of scores [23]. The higher the correlation value meaning that the instrument is more reliable. Alternate forms reliability is the extent to which scores from one sample are stable over an administration of two instruments of different versions of the instruments with the same concept but being administered twice at different at two different time intervals. Alternate forms and test-retest reliability combines both concept above. Internal consistency reliability is looking at the correlation between all items that make up the constructs to ensure that the items are measuring the same concept [20]. Lastly is the inter-

rater reliability, a concept looking at whether scores from one sample are consistent when more than one observer records the behaviour of respondents at the same time using the same instrument [22].

There is a relationship between validity and reliability. Any instrument can be reliable but not valid however, it cannot be valid if it is not reliable [21]. In other words, if an instrument is valid, it must be reliable. And, in general, checking for validity of an instrument is more difficult than checking for reliability because validity is measuring data related to knowledge whereas reliability only concerns with the consistency of scores.

2. RESEARCH METHOD

This pilot study was conducted at the primary and secondary schools in Melaka, a state at the south of Peninsular Malaysia. It is not the place of the actual research but it has a similar background to the actual study. The respondents were requested to sign on an informed consent form attached with the questionnaire which would serve as an evidence of their voluntary participation in the study. The data collection process was made in June 2012. Data is collected from four different types of schools – primary and secondary from urban and rural areas. Teachers are selected using the convenience sampling method. In general, data can be collected from an instrument such as a test, scale, observation procedure, and questionnaire or interview schedule [11]. Questionnaire could either be structured, semi-structured or unstructured and the items could either be a closed or an open type [19]. In this study, questionnaire is structured and the items are a closed type. All the items for evaluation dimensions are rated on a 5-point Likert scale from totally disagree to totally agree. In developing items for the questionnaire, initially the researcher determines the characteristics listed by Stufflebeam in the CIPP Model for context, input, process and product dimensions of evaluation. Then, constructs for several instruments on SBA, especially those developed with respect to the Asian countries are looked into. The researcher tries to match any suitable constructs with the dimensions suggested by Stufflebeam as there are no instruments which follow the dimensions suggested by Stufflebeam. After that, all the constructs go through the process of operationalizing as suggested by Cohen *et al.* It is the process of shifting a generalized purpose of questionnaire to a specific set of features to enable empirical data is collected from the items in the questionnaire. The questionnaire is divided into two sections: the main section on various dimensions of evaluation of SBA follows with the demographic information. There are 71 questions on the first section and 12 questions on the second section.

IBM Statistical Package for the Social Sciences (SPSS version 20.0) is used for the analysis of the data. In achieving content validity of an instrument, the researcher has made an extensive search of the literature from theories, previous instruments, models and past research findings for the concepts related to SBA implementation. Then, a professor who is an expert in this area is referred to. Next is to check for face validity. In order to assess for face validity, few respondents who are practising SBA in schools are asked to judge the items and then give comments on the questionnaire. A few changes are made upon comments such as ‘SBA is non-threatening’ is changed to ‘SBA is not burdening to me’. Finally, exploratory factor analysis (EFA) is used to check for construct validity. By definition, EFA is a technique used to explore the interrelationships among a set of variables [23]. An independent EFA using the principal components analysis (PCA) with a rotation called Direct Oblimin is conducted on the questionnaire. Eigenvalue or variance extracted by the factor greater than 1 is used. Next, in this study, reliability of the instrument is measured using internal consistence reliability as this study involves only one version of instrument and is administered once to all the respondents. It clearly shows that this study lacks test-retest reliability aspect as questionnaire is administered only once. In terms of practicality of the instrument, when pilot study is conducted, the respondents are asked to comment on the wording, timing and their understanding of the items. They are also asked to make suggestions on content that they feel are more suitable.

3. RESULTS AND ANALYSIS

150 questionnaires were distributed to the respondents and finally, 120 respondents responded. Nearly 90 percent respondents are Malay and nearly two-thirds are female. Nearly half of the respondents had 10 to 20 years of teaching experience and nearly half of them experience one year in practising SBA. Next, the reliability and validity of the instrument are presented below according to the evaluation dimensions (input, process and product) of SBA implementation. Context evaluation considers whether teachers are from urban-rural or primary-secondary type of school.

3.1. Reliability and Validity for Input Evaluation

Input dimension of evaluation consists of three constructs - material and personal needs in SBA, the appropriateness of personnel’s number and qualification and the suitability of physical infrastructure and

ICT. Each construct has three items so there are nine items altogether for input dimension. Cronbach's Alpha value is shown in Table 1. Cronbach's alpha for the first construct (material and personal needs in SBA) is 0.779, for the second construct (appropriateness of personnel's number and qualification) is 0.517 and for the third construct (suitability of physical infrastructure and ICT) is 0.809. As the value of Cronbach's alpha has to be above 0.7 for the items to be acceptable for the research purposes, then A18 from the second construct is deleted. The alpha value shows an increment from 0.517 to 0.675.

Table 1. Values of Cronbach's Alpha if item deleted and Overall Cronbach's Alpha for the Input Evaluation Constructs

Input Evaluation Constructs	Item	Cronbach's Alpha if item deleted	Overall Cronbach's Alpha Value
1. Material and personal needs in the SBA system	A15	0.689	0.779
	A16	0.771	
	A17	0.635	
2. Appropriateness of personnel's number and qualifications	A18	0.675	0.517
	A19	0.366	
	A20	0.209	
3. Suitability of physical infrastructure and ICT	A21	0.782	0.809
	A22	0.723	
	A23	0.706	

Table 2. Two-factor pattern matrix showing the factor loadings of each of the variables

Item	Component	
	1	2
A23	0.821	
A22	0.813	
A21	0.807	
A20	0.689	
A19	0.613	
A17		0.847
A16		0.840
A15		0.782

Only factor loadings more than 0.3 are counted towards any factors but still, meaningful interpretations have to be carefully been done when double loadings occur [20]. PCA is run after the deletion of A18. Two factors are formed as shown in Table 2. The researcher decides to continue with the three constructs as previously been hypothesized realizing the fact that factor analysis only indicates construct validity and not much on content validity. In addition, the researcher also feels that it is better to differentiate between the construct on 'personnel' and 'physical infrastructure'. So, the three constructs are maintained for input evaluation. The default setting is then changed as shown in Table 3, with the number of factors are specified to three. It clearly shows that the type of items that contributed to the factors is the same as have been presented earlier by the theory.

Table 3. Three-factor pattern matrix showing the factor loadings of each of the variables

Item	Component		
	1	2	3
A23	0.905		
A22	0.804		
A21	0.761		
A15		0.833	
A16		0.829	
A17		0.812	
A19			-0.852
A20			-0.843

3.2. Reliability and Validity for Process Evaluation

Process dimension of evaluation consists of ten main constructs altogether (belief, feeling, readiness, understanding, skill, in-house training, administration, moderation, monitoring and challenges). First is to look at the first three constructs which contributed to the attitude of teachers towards SBA. Teachers' attitude consists of three sub-constructs which are belief, feeling and readiness towards SBA. Looking at the

reliability of items in Table 4 and considering the three-factor pattern matrix of factor analysis on those items in Table 5, few decisions have been made.

i) For belief sub-construct, all the three items, A1, A2 and A5 are maintained and one item, item A7 from 'feeling' construct is added to it. Item A7 states that 'SBA can be effectively incorporated into existing lessons'. Item A6 is deleted to increase the Cronbach alpha value. The Cronbach alpha value for the four items, A1, A2, A5 and A7 is improved from 0.671 to 0.749.
ii) For feeling sub-construct, only one item, item A4 is maintained and item A3 is deleted. Since the Cronbach alpha is low, another two items are added. They are 'SBA is not boring' and 'SBA is compulsory'.
iii) For readiness sub-construct, A8 is deleted and the Cronbach alpha value for the three items, A9, A10 and A11 is improved from 0.581 to 0.654.

The fourth, fifth, sixth and seventh constructs are the teachers' understanding on SBA, the effect of SBA courses, IHT on SBA and encouragement by administrators. Considering the value of the Cronbach Alpha in Table 4 and the three-factor pattern matrix factor analysis of items in Table 6, few decisions have been made as follows:

i) For the fourth construct, 'teachers' understanding on SBA', all the three items, A12, A13 and A14 are maintained with the Cronbach alpha value is 0.644.
ii) For the fifth construct, 'effect of SBA courses', all the four items, B26i), ii), iii) and iv) are maintained with the Cronbach alpha value is 0.992.

Table 4. Values of Cronbach's Alpha if item deleted and Overall Cronbach's Alpha for the Process Evaluation Constructs

Process Evaluation Constructs	Item	Cronbach's Alpha if item deleted	Overall Cronbach's Alpha Value
1. Teacher's attitude: teacher's belief	A1	0.526	0.671
	A2	0.548	
	A5	0.642	
	A6	0.694	
2. Teacher's attitude: teacher's feeling	A3	0.227	0.523
	A4	0.488	
	A7	0.500	
3. Teacher's attitude: teacher's readiness	A8	0.654	0.581
	A9	0.429	
	A10	0.490	
	A11	0.455	
4. Teacher's understanding on SBA	A12	0.510	0.644
	A13	0.555	
	A14	0.579	
5. Effect of courses on improving skills of SBA	B26i	0.990	0.992
	B26ii	0.988	
	B26iii	0.990	
	B26iv	0.989	
6. IHT on SBA	B27i	0.968	0.973
	B27ii	0.955	
	B27iii	0.959	
7. Encouragement by administration	B28i	0.610	0.543
	B28ii	0.400	
	B28iii	0.278	
8. Moderation process	B29i	0.862	0.892
	B29ii	0.861	
	B29iii	0.818	
9. Monitoring process	B30i	0.649	0.793
	B30ii	0.792	
	B30iii	0.705	
10. Challenges	C1	0.837	0.836
	C2	0.832	
	C3	0.813	
	C4	0.818	
	C5	0.819	
	C6	0.821	
	C7	0.820	
	C8	0.821	
	C9	0.817	
	C10	0.827	
	C11	0.836	
	C12	0.818	

iii) For the sixth construct, 'IHT on SBA', all the three items, B27i), ii) and iii) are maintained with the Cronbach alpha value is 0.973.
iv) For the seventh construct, 'encouragement by administration', item B28i) is deleted and the Cronbach alpha value increases from 0.543 to 0.610.

Table 5. Three-factor pattern matrix showing the factor loadings of attitude construct

Item	Component		
	1	2	3
A7	0.866		
A5	0.719		-0.317
A1	0.644		
A2	0.639		0.380
A9		0.883	
A10		0.766	
A11		0.550	
A3	0.347	0.393	0.334
A4			0.906

Table 6. Three-factor pattern matrix showing the factor loadings of understanding, effect of SBA courses, IHT on SBA and administrators

Item	Component		
	1	2	3
A12		0.783	
A13		0.690	
A14		0.799	-0.310
B26i	0.977		
B26ii	0.994		
B26iii	0.972		
B26iv	0.992		
B27i	0.971		
B27ii	0.951		
B27iii	0.977		
B28ii			0.807
B28iii			0.839

The eighth, ninth and tenth constructs are the moderation process, the monitoring process and the challenges faced by the respondents. Considering the value of Cronbach alpha in Table 4, the three-factor pattern matrix factor analysis in Table 7 and the two-factor pattern matrix in Table 8, all items are maintained according to the theory.

Table 7. Three-factor pattern matrix showing the factor loadings of items on moderation, monitoring and challenges

Item	Component		
	1	2	3
B29i		-0.908	
B29ii		-0.864	
B29iii		-0.838	
B30i		-0.693	
B30ii		-0.708	
B30iii		-0.644	
C1		0.401	0.356
C2			0.880
C3	0.330		0.523
C4			0.551
C5			0.518
C6			0.832
C7			0.597
C8	0.750		
C9	0.640		
C10	0.816		
C11	0.414	-0.394	
C12	0.748		

Table 8. Two-factor pattern matrix showing the factor loadings of items on moderation, monitoring and challenges

Item	Component	
	1	2
C3	0.691	
C12	0.673	
C9	0.673	
C7	0.668	
C4	0.667	
C5	0.648	
C6	0.628	
C8	0.621	
C10	0.528	
C2	0.505	
C1	0.445	0.419
B29i		-0.884
B29ii		-0.829
B29iii		-0.765
B30i		-0.758
B30ii		-0.703
B30iii		-0.681
C11	0.353	-0.474

3.3. Reliability and Validity for Process Evaluation (School Improvement Constructs)

There are six items for the first construct (role of SBA) and six items for the second construct, (importance of SBA). Looking at the Cronbach alpha value in Table 9 and the two-factor pattern matrix in Table 10, two items are deleted from the construct 'role of SBA'.

Table 9. Values of Cronbach's Alpha if item deleted and Overall Cronbach's Alpha for the School Improvement Constructs

School Improvement Constructs	Item	Cronbach's Alpha if item deleted	Overall Cronbach's Alpha Value
1. Role of SBA	D32i	0.801	0.779
	D32ii	0.716	
	D32iii	0.746	
	D32iv	0.708	
	D32v	0.744	
	D32vi	0.747	
2. Importance of SBA	D33i	0.832	0.867
	D33ii	0.812	
	D33iii	0.865	
	D33iv	0.845	
	D33v	0.864	
	D33vi	0.843	

Table 10. Two-factor pattern matrix showing the factor loadings of items for the School Improvement Constructs

Item	Component	
	1	2
D33iv	0.906	
D33ii	0.886	
D33i	0.829	
D33iii	0.778	
D33vi	0.607	
D32ii	0.586	
D32iii	0.568	
D33v	0.425	
D32i		0.848
D32vi		0.622
D32iv		0.498
D32v		0.466

They are item D32ii) (reduces pressure in public exam) and item D32iii) (improves literacy and numeracy). They are then moved to the second construct, 'importance of SBA'.

3.4. Reliability and Validity for the Product Evaluation

Table 11 shows the values of reliability of items and Table 12 shows the three-factor pattern matrix factor analysis of items. Considering the reliability and validity of items, all eight items are maintained. So, all the eight items contribute to product evaluation constructs. Finally, Table 13 shows all the items on input, process and product evaluation before and after undergo reliability and validity process. Out of 71 items, 68 items are retained.

Table 11. Values of Cronbach's Alpha if item deleted and Overall Cronbach's Alpha for the Product Evaluation Constructs

Product Evaluation Constructs	Item	Cronbach's Alpha if item deleted	Overall Cronbach's Alpha Value
1. Student's attitude towards SBA	E34i	0.667	0.761
	E34ii	0.707	
	E34iii	0.658	
2. Student's knowledge in SBA	E35i		0.768
	E35ii		
3. Motivational source towards learning	E36i	0.722	0.722
	E36ii	0.622	
	E36iii	0.560	

Table 12. Three-factor pattern matrix showing the factor loadings of product evaluation

Item	Component		
	1	2	3
E35ii	0.921		
E35i	0.742	0.303	
E36i		0.798	
E36ii		0.791	
E36iii	0.351	0.685	
E34iii			0.855
E34ii			0.844
E34i			0.641

Table 13. All items on input, process and product evaluation constructs

Input evaluation construct	Items before the validity and reliability process	Items after the validity and reliability process
Material and personal needs in SBA system	A15, A16, A17	A15, A16, A17
Appropriateness of personnel's qualifications	A18, A19, A20	A19, A20
Suitability of physical infrastructure and ICT	A21, A22, A23	A21, A22, A23
Process evaluation construct		
Teacher's attitude: belief	A1, A2, A5, A6	A1, A2, A5, A7
Teacher's attitude: feeling	A3, A4, A7	A3, new item, new item
Teacher's attitude: readiness	A8, A9, A10, A11	A9, A10, A11
Teacher's understanding on SBA	A12, A13, A14	A12, A13, A14
Effect of SBA courses on improving skills of SBA	B26i, B26ii, B26iii, B26iv	B26i, B26ii, B26iii, B26iv
IHT on SBA	B27i, B27ii, B27iii	B27i, B27ii, B27iii
Encouragement by administration	B28i, B28ii, B28iii	B28ii, B28iii
Moderation process	B29i, B29ii, B29iii	B29i, B29ii, B29iii
Monitoring process	B30i, B30ii, B30iii	B30i, B30ii, B30iii
Challenges	C1 – C12	C1 – C12
School Improvement Construct		
Role of SBA	D32i, D32ii, D32iii, D32iv, D32v, D32vi	D32i, D32iv, D32v, D32vi
Importance of SBA	D33i, D33ii, D33iii, D33iv, D33v, D33vi	D32ii, D32iii, D33i, D33ii, D33iii, D33iv, D33v, D33vi
Product evaluation construct		
Student's attitude towards SBA	E34i, E34ii, E34iii	E34i, E34ii, E34iii
Student's knowledge in SBA	E35i, E35ii	E35i, E35ii
Motivational source towards learning	E36i, E36ii, E36iii	E36i, E36ii, E36iii
TOTAL ITEMS	71 items	68 items

3.5. Discussion

Assessment is a vital component in education. The interaction between assessment, curriculum and instruction is very important if we were to improve the teaching and learning process in school [24]. SBA is one of the main elements that contribute to this. Currently, the instrument to evaluate the implementation of SBA especially in Asian countries is still lacking. Therefore, the psychometric properties of an instrument to measure the perception of teachers towards SBA implementation are developed and assessed. According to [25], presenting the value of reliability and validity of a questionnaire is important so that other researchers are confident with the quality of the data they gain later. Since there is no validated instrument in the context of this study and also there is no such instrument which suits the objective of this study, this instrument has to be developed by the researcher. The instrument is developed based on literature reviews and past instruments on SBA especially those in Asian countries. In this study, the value of Cronbach Alpha is found to be between 0.610 and 0.992. This is considered quite acceptable as [20],[26] state that the value has to be more than 0.7 for a test to be internally consistent. In addition, the value of factor loadings towards factors are also quite high, hence it provides enough information on construct validity although meaningful interpretations by the researcher are also considered.

4. CONCLUSION

This is the study to provide a useful framework for evaluation of SBA in Malaysia. It is accepted that some SBA components will be aimed primarily at teachers, and the perception by pupils maybe less easy to identify. Furthermore, the success of this proposed framework is dependent upon the quality of the evaluations undertaken. In order to gain a more meaningful formative and summative evaluation, perceptions from different samples such as the students, head teachers, administration groups and the ministry officers are greatly needed. If evaluative approaches are not developed excessively, the interrelationship between factors which highly contributed to SBA implementation will remain elusive. Furthermore, investing in the professional development of teachers that have little impact on students' development might not really help in improving the performance of students.

ACKNOWLEDGEMENTS

My deepest appreciation and unlimited thankfullness addressed to my family, my supervisor, Prof Daniel, my collegues and all staffs of faculty.

REFERENCES

[1] Lembaga Peperiksaan Malaysia, "Panduan dan Peraturan PBS. Putrajaya," 2011. Available at: http://www.scribd.com/faridooi/d/81532201-Panduan-Dan-Peraturan-Pbs-2011. (Accessed: 22 November 2011).
[2] Lembaga Peperiksaan, "Pentaksiran Berasaskan Sekolah," 2010. Available at: http://buletinkpm.blogspot.co.uk/2012/11 /pentaksiran-berasaskan-sekolah-pbs.html (Accessed: 7 November 2011).
[3] C. Noraini, *et al*., "Ideal vs Reality: Evidences from senior teacher experiences on the Malaysia SBA system," *Proceedings of the Malaysian Educations Dean Council. UIAM, 23-25 September. INSTED,* 2013. Available at:http://www.iium.edu.my/medc2013/documents/ProsidingMEDC.pdf (Accessed: 8 October 2014).
[4] Lembaga Peperiksaan, "Panduan Pengurusan PBS. KPM," 2012. Available at: http://www.moe.gov.my /lp/files/pbs/bahan/Buku%20Panduan%20Pengurusan%20Pentaksiran%20Berasaskan%20Sekolah%20%28PBS %29 %202012.pdf (Accessed:26 November 2013).
[5] D. L. Stufflebeam, "The relevance of the CIPP Evaluation Model for Educational Accountability," *Annual Meeting of the American Association of School Administrators,* 1971. Available at: http://eric.ed.gov/PDFS/ED062385 .pdf (Accessed: 3 September 2012).
[6] D. L. Stufflebeam, "The CIPP Model for Evaluation," in Kellaghan, T and Stufflebeam, D. L., *International Handbook of Educational Evaluation*, Dordrecht: Kluwer Academic Publishers, pp. 31-62, 2003.
[7] A. Armstrong & E. Ogren, "Evaluation Models and Strategies," *Evaluation and Training and Services Australia,* Melbourne, 1986.
[8] K. Hakan & F. Seval, "CIPP Evaluation Model Scale: Development, reliability and validity," *Procedia and Social and Behavioral Sciences*, vol. 15, pp. 592-599, 2011.
[9] G. Zhang, *et al*., "Using the context, input, process and product evaluation model (CIPP) as a comprehensive framework to guide the planning, implementation and assessment of service-learning programs," *Journal of Higher Education Outreach and Engagement*, vol/issue: 15(4), pp. 57-84, 2011.
[10] Y. Azizi, *et al.,* "Sejauhmana Model Stufflebeam (KIPP) boleh membantu dalam Penilaian Program Pembelajaran," 2010. Available at: http://eprints.utm.my/2256/1/AziziYahaya_Sejaumanakah_Model _Stufflebeam_%28 KIPP %29.pdf (Accessed: 2 March 2012).

[11] S. Isaac & W. B. Michael, "Handbook in Research and Evaluation: A Collection of Principles, Methods, and Strategies Useful in the Planning, Design and Evaluation of Studies in Education and the Behavioral Sciences," 2nd edn. California, EdITS Publishers, 1982.
[12] J. Salmiah, "Acceptance towards SBA among agricultural integrated living skills teachers: challenges in implementing a holistic assessment," *Journal of Technical Education and Training (JTET)*, vol/issue: 5(1), pp. 44-51, 2013.
[13] Y. Boon & M. Shaharuddin, "Kepemimpinan Guru Besar Dalam Pelaksanaan Pentaksiran Berasaskan Sekolah (PBS) Di Sekolah Kebangsaan Daerah Kota Tinggi, Johor," Fakulti Pendidikan. Universiti Teknologi Malaysia, 2011. Available at:http://eprints.utm.my/11929/1/Kepemimpinan_Guru_Besar_Dalam_Pelaksanaan_ Pentaksiran _Berasaskan _Sekolah.pdf (Accessed: 8 March 2012).
[14] Chan Y. F. & Gurnam K. S., "School-based Assessment among ESL Teachers in Malaysian Secondary Schools," *Journal of the Malaysian Education Deans' Council*, vol. 9, pp. 1-18, 2012.
[15] S. A. Lukman & A. A. Uwadiegwu, "SBA as an innovation in Nigerian educational system: The Implementation Challenges," *Knowledge Review*, vol/issue: 25(1), 2012.
[16] K. Koh & R. L. Velayutham, "Improving teachers' assessment literacy in Singapore schools," 2009. Available at: http://www.nie.edu.sg/nie_cma/attachments/topic/14e8a745eaKV/NIE _research_brief_09_002.pdf (Accessed: 12 November 2013).
[17] L. Cheng, *et al.*, "Impact and consequences of SBA: students' and parents' views of SBA in Hong Kong," *Language Testing*, vol/issue: 28(2), pp. 221-250, 2011.
[18] A. N. Oppenheim, "Questionnaire Design, Interviewing and Attitude Measurement," London, Pinter Publishers, 1996.
[19] L. Cohen, *et al.*, "Research Methods in Education," 6th edn. London, Routledge, 2007.
[20] D. Muijs, "Doing Quantitative Research in Education with SPSS," London, SAGE Publications Ltd, 2011.
[21] S. L. Jackson, "Research Methods and Statistics, A Critical Thinking Approach," USA, Thomson Wadsworth, 2003.
[22] J. W. Creswell, "Educational Research: planning, conducting and evaluating quantitative and qualitative research," New Jersey, Pearson Education, Inc., 2002.
[23] J. Pallant, "SPSS Survival Manual: A step by step guide to data analysis using SPSS for Windows," 3rd ed. New South Wales, Allen and Unwin, 2007.
[24] S. Young & C. Giebelhaus, "Formative Assessment and Its Uses for Improving Student Achievement," Education Data Management Solutions, STI. 2005. Available at: www.cbohm.com/news/STI/STI_White_Paper.pdf (Accessed Nov 2011).
[25] M. Miller, "RES 600: Graduate Research Methods: Reliability and Validity," Western International University, 2012. Available at: http://michaeljmillerphd.com/res500_lecturenotes /Reliability_and_Validity.pdf.
[26] J. O. Nunnally, "Psychometric theory," New York, McGraw-Hill, 1978.

BIOGRAPHY OF AUTHOR

Nor Hasnida Md Ghazali is a senior lecturer at Faculty of Education and Human Development, Universiti Pendidikan Sultan Idris, Tanjong Malim, Perak, Malaysia. She completed Degree in Mathematics Education from University Kebangsaan Malaysia in 1998; Master in Mathematics Education from University Kebangsaan Malaysia in 2010 and Doctor of Philosophy in Assessment and Evaluation in Education from University of Southampton, United Kingdom in 2015.