

Evaluation of the School-Based Assessment System: an Evidence via Structural Equation Modelling

C.M.G. Nor Hasnida

Fakulti Pendidikan dan Pembangunan Manusia, Universiti Pendidikan Sultan Idris,
Tanjong Malim, 35900, Perak DarulRidzuan, Malaysia

Abstract: The school-based assessment (SBA) system is a holistic assessment system conducted in school by subject teachers in assessing the students' cognitive (intellect), affective (emotional and spiritual) and psychomotor (physical) aspects which is in line with the National Philosophy of Education and the Standards-based School Curriculum. It is becoming recognized as an important means of assessing students which could be considered as an effort in improving human capital development holistically, as well as to lessen the negative influence of examination-oriented education systems on students. Although SBA is still in a relatively early stage of development in an upper middle-income developing country like Malaysia, it is becoming increasingly important that SBA is evaluated considering its impact on students' achievement. In order to evaluate SBA, a self-administered questionnaire is designed based on the CIPP evaluation format. The data are derived from a sample of 776 primary and secondary school teachers who have been sampled using a stratified random sampling of schools. A technique called structural equation modeling (SEM) with the use of AMOS version 18 software is used for data analysis. The result indicates that there is a reasonable fit with the SBA evaluation model with an interrelationship between the dimensions of evaluation (context, input, process and product). Theoretical, methodological and practical implications are discussed.

Key words: School-based assessment • Evaluation • The CIPP model • SEM

INTRODUCTION

Education systems around the world are going through reforms in an effort to improve quality of life. Reforms are aiming to raise the bar of students' performance and to close the gap especially among the lower performing group of students [1]. It is also aimed in providing students with competencies and higher order skills which suits the 21st century challenges [2]. Reforms include the way students are assessed. There is a large body of literature which focuses on the negative impact of the traditional concept of assessment on student learning. A sole focus on public examination could negatively affect the society such as exerting unconstructive influence on students' emotional strength as a barrier to success [3], distort the school curricula and produce less reliable and invalid results [4] or regulate incapability in assessing skills such as problem-solving skills, orally expressing thoughts or school behavior [5].

Recently, the trend of assessment system in some countries is changing with the introduction of formative assessment. Malaysia, as an upper middle income developing country has also started to implement SBA to a few subjects in 2002 and, has finally implemented it in total in 2011 with the Year One students made the first move into it [6]. Most literature on formative assessment are from developed countries as they have implemented it since quite a long time ago. As such, in Australia, the emergence of SBA started in the late 1960s [7] and in Finland and Sweden, its implementation has begun in the early 1970s [8]. Therefore, there is a need for more research on the execution of SBA in developing countries. School programs that are implemented by the means of this system need evaluation to avoid meaningless changes and reforms [9]. Even, evaluation itself is an essential part of improvement for any educational practices and procedures. However, the evaluation of SBA is rarely undertaken in a systematic and focused

manner although its importance is widely acknowledged. As such, the research evidence on SBA evaluation is looking only at some dimensions which do not give a full rounded indication of the effectiveness of the system. Some researches focus only on the process of evaluation itself including looking at teachers' attitude [10, 11] or product evaluation to identify the benefits of SBA [12]. Other researches look at both evaluation dimensions; process and product as in Annie [13] which determines students' perception of and reaction to feedback or relating process and product [14, 15] as well as the input and product [16]. To date, studies that are related to all the four evaluation dimensions of SBA are non-existent.

Looking at the Malaysian National Philosophy of Education which was formulated in 1988, it strictly stated that the education has to be an on-going effort in developing potentials of students in a holistic manner based on a firm belief in and devotion to God [17]. This means that education has to have a variety of strategies and approaches to make sure students' potentials are developed in a holistic and integrated manner to produce a well-balanced community [18]. Furthermore, according to the fifth chapter of the five year plan in the Tenth Malaysian Plan (2011-2015), the aim of the Plan is to improve the education of the people starting from early education and basic education, followed by tertiary education to the working environment. Three core strategies involved are to reform the education system to improve students' performance, to increase people's skills for employability and reforming the labor market. The interest of this study seems to be consistent with the first core strategy of the Plan, which is focusing on reform or improvement of the education system based on the National Philosophy of Education [19]. Furthermore, the strategies of the Plan also include a focus on the involvement of students in sports and curriculum, inculcating values and ethics as well as improving the use of ICT and communication amongst students.

Frameworks for Evaluating SBA: A number of conceptual frameworks have been developed in evaluating programs. Among the evaluation frameworks identified as the most applicable to educational programs are Tyler's concept, Kirkpatrick's model, Stake's model and Stufflebeam's CIPP model. According to the CIPP model developed by Daniel Stufflebeam, any system or project could be evaluated in four dimensions – context, input, process and product [20]. A program can be evaluated in a single dimension or in a few dimensions altogether [21]. CIPP model has been used to evaluate

various educational programs and projects from various disciplines [22]. Furthermore, each dimension in the evaluation should serve a particular decision as evaluation involves decision-making process [23]. Hence, the context, input, process and product evaluation should support the planning, structuring, implementing and recycling of decisions respectively. Context evaluation refers to the systematic process of getting information on the establishment of new program objectives or to translate needs into objectives and also the modification of existing objectives or confirmation of the present objectives in order to guide the decisions planning [23]. In other words, it emphasizes on assessing needs, assets and problems within a defined environment [24]. It is like asking, 'What should we do to evaluate this program?', 'Which objectives should be obtained?' [25], 'Have the important needs addressed accordingly?' [22] or 'What is the context of the program and how will it match with the target population?' [26]. Methods of collecting data for this purpose could be surveys, interviews, document reviews [27] or through focus group, website, journals or standardized testing [26]. However, Clinton believes that the most used method is the existing documentation. Input evaluation focuses on assessing the strategies, personnel, resources, procedures or a prospective cost assessment in achieving the program objectives which has been derived earlier [23]. This leads to structuring decisions by the decision-makers and a few appropriate questions which should be considered for this evaluation are for instance, 'Was the effort guided by a defensible plan?', 'Which strategies or procedures should be tried out?' or 'Was the effort guided by a defensible plan or budget?' [22] and 'What input have been received from the ministry?' or 'What professional development did the providers receive?' [26]. The data for input evaluation could be accumulated using surveys, interviews, websites or journals with the majority of data come from the existing documentation [26].

The next step is to carry out the process evaluation. During process evaluation, decision-makers determine the processes implemented to achieve the goal of a program. The processes should be monitoring challenges, identifying program adjustments, getting additional information for changes, documenting the process or running regular monitoring on the activities [28]. Important questions to be asked for example, 'Are we doing it correctly?' and 'How adequately are these strategies or procedures working?' [25]. Process evaluation involves implementing decisions as it guides the decision-makers on how to reinforce the program's

implementation, to assist program replication or to demonstrate the reasons why the objectives of the program could not be achieved [23]. Interview is recognized as the best means of collecting data for process evaluation [26]. The final dimension is the product evaluation, also known as 'outcome evaluation'. It serves as the program recycling decisions in determining and examining the specific outcomes of the program. In this regard, the decision-makers should consider a few aspects like, 'Should the program be continued or not?', 'How effective are the goals and objectives being accomplished?' [25] or 'Did the effort succeed?' [22]. In short, product evaluation is like comparing the outcomes of a program with its objective which has been set earlier. All the evaluations dimensions mentioned above are in a dynamic relationships to each other whereby information from any components of evaluation could be provided to the previous components so that the changes could be made should the need arises [25].

The Assessment System in the Malaysian Educational

Context: In Malaysia, assessment system is controlled by the Malaysian Examination Syndicate (MES) centralized in Putrajaya. MES is responsible for the development of a national curriculum and all procedures related to examination such as preparing test, setting examinations rules and regulations, revising syllabi [29]. Previously, assessment system was focusing on public examination but formative assessment has been introduced in recent years. In 2011, SBA which consists of both assessments, formative and summative has been implemented formally on the Year One students. SBA in the Malaysian educational context consists of two main components known as the academic components (school and central assessment) and the non-academic components (physical activities, sports, co-curricular assessment and psychometric assessment). It is a form of assessment which is planned, administered, scored and reported by the subject teachers following the guidelines from MES [30]. The main objectives of its implementation are to get an overall picture of individual's potential, to monitor individual's development and to help them to increase their potentials as well as to make a meaningful reporting on individuals. In other words, SBA is assessing the process and product of each and every student in formative and summative way by practicing both, the assessment for learning and assessment of learning concept.

The Present Study: Based on the CIPP model of evaluation that promotes the interrelationships between dimensions of evaluation, this study examines dimensions of evaluation such as context, input, process and product that have a causal link to one another. In this study, context evaluation focuses on the environment where the changes happen involving two factors; school type (urban and rural schools) and school category (primary and secondary schools). The other three evaluation dimensions are described in more detail as below:

- Input evaluation. This study focuses on three first-order factors: material and personal needs in SBA, appropriateness of personnel qualifications and suitability of physical infrastructure and ICT. (Example of items to be rated by teachers are: 'Teachers are sufficiently qualified to implement assessment activities' or 'The ICT hardware is suitable to conduct SBA activities').
- Process evaluation includes twelve first-order factors: belief, feeling, readiness, understanding, courses, in-house training, administration, challenges, moderation, monitoring, role and importance of SBA. (Examples of items are: 'SBA is efficient in improving students' learning' or 'SBA is not burdening me')
- Product evaluation includes three first-order factors: students' attitude, knowledge and motivational toward learning. (Examples of items are: 'Students practice SBA in their study' or 'SBA encourages students to read more books than they had before').

The factors involved in each evaluation dimension are formed based on the literature especially the three assessment models such as the formative assessment model [31], the Logic model [32] and the SCAP model [33]. Formative assessment model concentrates on four main elements: identifying the gap, feedback, student involvement and learning progressions [34]. Based on this model, identifying the gap refers to the process of identifying the 'zone of proximal development' of the students to enable the teachers to scaffold or support the students by adapting their teaching or having collaborative processes with students during formative assessments. Feedback is a descriptive and criterion-based information from the teachers on the students' current learning and methods for improvement or in short, it could be viewed as an information about the gap [35]. Feedback, being the core of formative assessment with various roles in influencing students learning [36] has to

be well understood by teachers to avoid practicing it in an inefficient way. Student involvement focuses on the development of students' self and peer-assessment skills, whereas learning progressions are a series of progressions that should be learned by students in achieving their goals. Students' learning progressions are important as they are the basis for the other three main components of the model and developing them is not an easy task for teachers as they have to be well-trained on those skills. In conclusion, this model proves that the establishment of learning goals and criteria for success as an input, followed by the process of formative assessment which goes along in a cycle has rendered the ultimate goal in learning as the product to be achieved.

Aim of this Paper: The aim of this paper is to evaluate the implementation of SBA in schools in preparing a summative report on the effectiveness of the system. The model proposed studies the interrelationship between all the evaluation dimensions (context, input, process and product). It is hypothesized that process is positively associated with product and input is positively associated with both process and product.

MATERIALS AND METHODS

This cross-sectional survey study was conducted in February 2013. In choosing samples of schools, stratified random sampling is used. Overall a total of 2500 questionnaires have been distributed to the respondents in primary and secondary schools in one of the state in the north-east of Peninsular Malaysia. There are 826 completed questionnaires received from the respondents but 50 questionnaires have been discarded due to having had a lot of unanswered items or same responses written on the same page. The final total of questionnaires obtained is 776 and therefore the effective response rate was approximately 31.04 percent. The number of samples gathered seems adequate to apply a stable maximum likelihood estimation procedure in SEM in addressing the research objectives as suggested by Hair *et al.* [37]. The questionnaire has been developed by the researcher accordingly to suit the CIPP model interpretations based on theories, models and instruments from previous studies. It consists of two main sections, the demographic characteristics of the respondents and the items related to context, input, process and product evaluations.

The items on evaluation are ranked on a five-point *Likert* scale ranging from 'strongly disagree' to 'strongly agree'. The interrelationships between dimensions are analyzed using SEM. This is the most suitable method for this study as SEM is a multivariate technique which simultaneously examines a series of interrelated dependence relationship among variables [38]. Furthermore, SEM is also capable in correcting measurement error by providing estimation of error variance which could not be conducted by traditional multivariate procedures [39].

RESULTS

Demographic Description: Data indicates that nearly two-thirds (74.7%) of the respondents are females and the remaining are males. The majority of the respondents are Malay (93.6%). Nearly half of them (49.7%) are below 40 years of age. Most of them are degree holders (71.4%), a few with masters (3.4%) and none with Phds. Teaching experiences of them varied with most of them (49.1%) have 10 to 20 years experiences. Since SBA has been implemented in 2011 with the Year One students made as the first participants, based on the data collected in early 2013, nearly half (49.9%) of the respondents have had a year experience in practicing SBA and a similar percentage of respondents (53.0%) have at least attended courses on SBA once.

Confirmatory Factor Analysis: When using SEM, it is necessary to assess how well have the items representing their underlying latent constructs. Thus, a measurement model of input, process and product evaluation have gone through the process of Confirmatory Factor Analysis (CFA) to address the issues of validity and reliability of the models. CFA is used when the models have already been developed based on literature so the analysis of data is done based on the specified model. However, if the model is not specified and analysis is conducted based on data, exploratory factor analysis (EFA) is addressed. All the three models are run and they have resulted in a poor fit. Therefore, EFA has been conducted and yielded six new measurement models. Based on modification indices, some paths of covariance between error terms are added. Finally, all the six measurement models have shown an excellent fit to the data, valid and reliable. Characteristics for each of the measurement model tested are presented in Table 1.

Table 1: Characteristics of each of the final measurement models

Measurement Model	Fit indices value	Number of items	Multivariate kurtosis
Input Dimension	$X^2 = 1.751$; $df = 1$; $X^2/df = 1.751$; GFI = 0.999; AGFI = 0.989; NFI = 0.999; CFI = 0.999; TLI = 0.996; RMSEA = 0.031	4	6.104
Process Dimension			
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 Dimension	$X^2 = 3.458$; $df = 3$; $X^2/df = 1.153$; GFI = 0.998; AGFI = 0.991; NFI = 0.999; CFI = 1.000; TLI = 0.999; RMSEA = 0.014	5	12.123
Total		38	

Structural Equation Modelling: This study involves a two-stage approach to SEM analysis which includes assessing the CFA and analyzing the SEM. This approach is good as it could avoid bad measures because validation for each measurement model is assessed during CFA before proceeding with the full structural model [38]. SEM technique which is theory-driven, is used to determine the interrelationships between evaluation dimensions in a full structural model. Structural model (Fig. 1) is assembled based on theoretical interrelationships among the constructs of grounded in empirical research [39].

There are nine hypothesized causal paths altogether in the initial structural model with 38 observed variables. Initially, there are 68 observed variables but after going through procedures to gain best measurement models, they are reduced to 38. When this model is evaluated, it shows that the overall X^2 is 2692.337 with degrees of freedom equals to 639. The indices of fit shows $X^2/df = 4.213$, CFI = 0.909, GFI = 0.836, NFI = 0.884, TLI = 0.899, RMSEA = 0.065 and ECVI = 3.781. It clearly shows that the model is not fit. Then, Process3 is deleted from the model since the value of correlation between Process3 and Process1 is 0.923 and between Process3 and Product is 0.944, thus rejects discriminant validity. Then, the structural parameter estimates are reviewed. The paths are determined. All the non-significant paths are deleted one at a time to check for the best solution [39]. A new model is produced with two path deleted (Product <- Process2; CR = 1.212 and Product <- Input; CR = 1.091). The model is estimated again and still not fit. According to Byrne

[39], the deletion of construct can change the number of sample moments and degree of freedom and then causes a big difference in X^2 .

Next, to further improve the fitness of the model with the five statistically significant paths, the monitoring process and one item from challenge construct are deleted due to their low estimate parameter value and the overlap of item content. Next, one item from attitude construct is deleted due to low parameter estimates and it is negatively skewed. It then followed by the deletion of understanding construct but maintained one item ('educational transformation through SBA leads to changes from examination-oriented to an assessment which is more integrated'). The item is maintained and grouped together with attitude construct as theoretically, the content of item suits the construct. The structural model is then being evaluated again and this seems to be the final structural model with 26 observed variables left as shown in Figure 2.

Although there is no firm rules when to stop the model, it depends to the researcher in getting the best parsimonious model by determining the items, the constructs and the interrelationship between them. The final structural model yields several noteworthy results which collectively support the adequacy of the model. The chi-square test results with a statistically significant discrepancy between the model and the data ($X^2 = 1066.796$, $p = 0.000$; CMIN/df = 3.756). The overall fit statistics shows a good fit characteristics (CFI = 0.943, TLI = 0.935, RMSEA = 0.060 and ECVI = 1.568).

The diagram illustrates a structural equation model with the following components and standardized path coefficients:

- INPUT** (Latent Variable):
 - Observed Variables: *appr* (loading .85, error *e0* variance .15), *sult* (loading .82, error *e7* variance .18), *res1* (loading .70, error *res1* variance .30).
- PRODUCT** (Latent Variable):
 - Observed Variables: *att* (loading .84, error *e10* variance .16), *mot* (loading .80, error *e18* variance .20).
 - Path from **INPUT**: *att* (loading .83, error *e10* variance .16).
- CHALL** (Latent Variable):
 - Observed Variables: *c7* (loading .45, error *e25* variance .55), *c9* (loading .76, error *e23* variance .24), *c10* (loading .74, error *e22* variance .26), *c11* (loading .73, error *e21* variance .27), *c8* (loading .67, error *e81* variance .33).
- PROCESS1** (Latent Variable):
 - Observed Variables: *courses* (loading .70, error *e74* variance .30), *b24ii* (loading .88, error *e77* variance .12), *b24iii* (loading .72, error *e78* variance .28).
 - Path from **INPUT**: *courses* (loading .54, error *e74* variance .30).
- PROCESS2** (Latent Variable):
 - Observed Variables: *mod* (loading .87, error *e63* variance .13), *b27i* (loading .88, error *e50* variance .12), *b27ii* (loading .91, error *e49* variance .09), *b27iii* (loading .90, error *e48* variance .10).
 - Path from **PROCESS1**: *mod* (loading .78, error *e63* variance .13).
- attitude** (Latent Variable):
 - Observed Variables: *a1* (loading .78, error *e03* variance .22), *a2* (loading .74, error *e02* variance .26), *a3* (loading .76, error *e01* variance .24), *a4* (loading .50, error *e00* variance .50), *a5* (loading .57, error *e06* variance .43), *a7* (loading .54, error *e07* variance .46).
 - Path from **CHALL**: *attitude* (loading .76, error *e57* variance .24).
 - Path from **PROCESS1**: *attitude* (loading .63, error *e57* variance .24).
 - Path from **PROCESS2**: *attitude* (loading .32, error *e57* variance .24).

885

All of the unstandardized estimates for the five structural paths are statistically significant, all values of standardized estimates are sound and are less than 1.00. As Chin [40] states that the value of standardized estimates is meaningful when it is more than 0.30, or at least 0.20. Furthermore, all covariance and error variance are statistically significant. In this model, the squared multiple correlation which represented the proportion of variance that was explained by the predictors of the construct shows 58.5% for product and 70.7%, 22.1%, 18.0% for process1, process2 and challenges respectively. For example, this structural relations explain quite a large amount of variance in product, meaning that product is well explained by the two factors of process evaluations which are teachers 'attitude and skills' and 'challenges'.

DISCUSSION AND CONCLUSIONS

SBA is a vital component in education as assessment is an integral part of the educational process. Understanding the interaction between assessment, curriculum and instruction is important in order to produce better impact in educational improvement. The findings of the present study have expanded the existing body of knowledge on the effectiveness of SBA implementation. Since there is an urgent need to know the effectiveness of this assessment system, this study has attempted to evaluate the system using CIPP model as a framework. Overall, our findings suggest that the model is partly helpful in explaining the interrelationships between dimensions of evaluation in SBA. Therefore, to some extent, CIPP model is suitable to be used as a means of explaining an evaluation process of SBA. Elements of CIPP model which are applicable in the western world are transferable to the Malaysian context. Furthermore, results from SEM also indicate that the model explains quite a high percentage of the variance. There is a strong positive and significant relationship between teachers 'attitude and skills' and students 'attitude and motivation' whereas 'challenge' is negatively associated with students 'attitude and motivation'. In other words, according to teachers, when teachers have a positive attitude and gain skills on assessment, students are more likely to improve their attitude and motivation towards learning. These findings therefore confirm previous research suggesting that process is actually associated with product. Research by ARG [41] suggests that whenever some processes are implemented, such as giving explanations on the purpose of test and constructive feedback or developing students' self-

assessment skills and criteria for learning, it could improve students' motivation. Furthermore, a case study conducted in Malaysia found out that various formative assessment techniques in Malay subject such as questioning technique, the use of scratch cards, loud pronunciation technique, matching technique, singing techniques and discussions have influenced students' attitude and knowledge on assessment [42]. Challenge is found to be negatively associated with students' attitude and motivation towards learning. One might suggest that it might not be worth considering as SBA is just newly-introduced. Wei [43] with his action research study has found that SBA has proved to increase students' intrinsic motivation, strengthen and sustain their motivation, help them to understand their strength and weakness better despite challenges that they are facing including time-consuming, lack of training on assessment, low reliability of assessment tasks and large class sizes. However, when changes are made to the data storage process in the Malaysian assessment system, whereby teachers no longer need to key-in lots of data into the online system, SBA is found to be well accepted by the majority of educators [44]. Hence, it shows that reducing challenges does matter to certain extent. Understanding the context of the school is also important. Computer and internet access might not be a problem to urban schools but to rural schools, they usually do.

There is no statistically significant relationship between 'moderation process' implemented in school and students 'attitudes and motivation'. This finding seems paradoxical with the previous findings on process and product. The possible explanation for this might be that teachers do not fully understand the purpose of moderation process in improving students' attitude on SBA and their motivation towards learning. This shows that there are different perceptions on this issue between teachers in Western countries and the teachers in Malaysia. However, in Malaysia, not all teachers are involved during internal and external moderation process. This might worsen the condition as they are not given opportunities to apply the knowledge they have gained during training session. Taken together, it is not too inordinate to claim that this finding reveals an issue of worth to be researched in the future. The research is better involved around the questions of how far have the teachers practice the idea of giving students opportunities to improve or, to what extent do teachers try to improve their teaching practices after going through moderation process in schools. 'Qualification and equipment' are associated significantly with all the three

components in process evaluation ('attitudes and skills', 'challenge' and 'moderation process'). This is supported by Heritage [34] who asserts that specific knowledge and skills are needed if teachers were to implement formative assessments effectively. In addition, this finding seems to corroborate previous studies in the western and Asian context. Over two years of school (2006 and 2007), Singapore has conducted a longitudinal quasi-experimental intervention study on authentic assessment [16]. When teachers were provided with an ongoing and sustained professional development on authentic assessment task design and rubric development for two years, there was an improvement observed on teachers' assessment task and students' work quality. Similarly, an intervention was conducted under a program called 'Embedding Formative Assessment' to school teachers in Cannington, London [45]. It involved training teachers with various strategies and techniques on formative assessment and also providing them with materials for it. After seven months, it was found that teachers have made a positive progress and the head teachers were looking forward to continue the program in the following year.

There is no statistically significant relationship between 'qualification and equipment' and students' 'attitude and motivation'. This finding contradicts the above assessment models. Newmann *et al.* [46] proposed in their model on capacity-building strategies that technical resources such as high-quality curriculum, books, assessments instruments, laboratory equipment, computer or adequate work space could influence instruction which in turn can improve students' achievement. However, there is still a strong indirect effect between them. It might be appropriate to review the items again or add more items to input evaluation in future studies.

In conclusion, this study provides some support to the effectiveness of SBA implementation in schools. It also supports the models developed earlier such as the CIPP model, formative assessment model and SCAP model. Moreover, this study also suggests that there is an opportunity to improve the system as long as the decision-makers are willing to take those challenges positively. However, on the question of how far has this newly-introduced assessment system achieved the standard, it has yet to be continuously evaluated. Nevertheless, this study has several limitations. First, the samples were taken only from teachers, not involving the other stakeholders, so the development and validation of instruments might be limited. Furthermore, the data merely come from the perceptions of teachers without observing

their real practices. Second, some items included in the survey were deleted during CFA and SEM procedure in order to gain the best fit of the model. Finally, the sample of this study was collected at government schools in one of the states in the north-east of Malaysia. Although the education system in Malaysia might not be much difference between each state as the system is centralized, cultural difference might limit the generalization of findings to other states.

Implications: Theoretical, methodological and practical implications are discussed suggesting the importance of study to different audience. Reviewing the factors in evaluation dimensions and its dynamic interplay examined in the study shows that the CIPP model is supported. The main implications of the strong associations between input and process suggest that resources and procedures supplied to schools should support an effective implementation of SBA. The same goes with the strong relationship between process and product. However, school administrators and the ministry need to scrutinize on the incapability of monitoring and moderation processes in ensuring a strong relationship with students' attitude and motivation. This is important because these two processes are the main components in quality assurance process set out by the ministry and trainings have also been conducted to all the teachers on these matters. The reason for this might be due to the fact that not all teachers are directly involved in the process and therefore, it limits the development and shared understanding between teachers on the interpretation of standards and many more. Although there is an absence of empirical findings on the direct effect of input on students' attitude and motivation', the existence of a strong indirect effect between them shows a good implication to the decision-makers. The failure in improving input in addressing assessment system comprehensively might affect most of the processes implemented and this might be a major reason for the disappointing result in achieving the objectives of SBA later on.

The results further demonstrate that SEM procedures supported the conceptual frameworks set out in this study. The use of SEM as a technique to simultaneously examine a series of relationships between variables is promising. In addition, the capability of SEM in estimating error variance parameters could improve the accuracy of results [39]. It is suggested that teachers' training should be enhanced. Even, the present Cascade model used by the ministry has to be reviewed as it is believed that the

model allows the dilution of information and this could affect its effectiveness [47]. And, training should focused more on hands-on rather than merely from lectures. Development of teachers' attitude, skills and knowledge in formative assessment is by no means a straight forward process and Heritage [48] believe that the three components are the major components in implementing formative assessment effectively. Training providers could then restructure their training programs to make it parallel with teachers' belief system together with their knowledge and skills. It will not be a waste to devote a significant of energy and financial resources in teachers' training if we know our targeted area of concerns. The valid instrument on SBA implementation developed in this study could also be used as a pre and post to evaluate teachers after they have attended trainings. Communication between various parties such as officers from the ministry, schools management, head teachers and teachers is the utmost importance. Policy instructed by the government has to be adhered by most parties. Head teachers have to clearly communicate the importance of assessment policy and stretch a positive school climate not only to the teachers at school but also to the parents and larger communities as SBA is a new innovation for them.

REFERENCES

- Fullan, M., 2011. Choosing the wrong drivers for whole system reform. Available at: <http://www.michaelfullan.ca/media/13501655630.pdf> (Accessed: 3 December 2013).
- Branden, K.V.D., 2012. 'Sustainable education: basic principles and strategic recommendations', *School Effectiveness and School Improvement*, 23(3): 285-304.
- 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.
- Wiliam, D., 2001. 'What is wrong with our educational assessments and what can be done about it', *Education Review*, 15(1).
- Begum, M. and S. Farooqui, 2008. 'School Based Assessment: Will it Really Change the Education Scenario in Bangladesh?', *International Education Studies*, 1(2): 45-55.
- Ministry of Education, 2011. Penambahbaikan UPSR dan PMR. Available at: <http://www.slideshare.net/Fadzliaton/penilaian-berasaskan-sekolah-pbs> (Accessed: 9 November 2011).
- Mercurio, A., 2008. Re-imagining school-based assessment at the upper secondary level, SACE Board of South Australia. Available at: <http://www.iaea2008.cambridgeassessment.org.uk/ca/digitalAssets/180434-Mercurio.pdf> (Accessed: 5 November 2011).
- Darling-Hammond, L. and L. McCloskey, 2008. 'Assessment for Learning Around the World: What Would it Mean to be Internationally Competitive?', *Phi Delta Kappan*, 90(4): 263-272.
- Sanders, J. and C. Sullins, 2006. *Evaluating School Programs: An Educator's Guide* (3rd ed.). Thousand Oaks, CA: Corwin Press.
- Sidek, R. and S.M Saleh, 2010. UTM Institutional Repository. Available at: http://eprints.utm.my/10502/1/kesediaan_Guru_Sekolah_menengah_Di_daerah_kota_baharu.pdf (Accessed: 8 November 2013).
- Majid, F., 2011. 'SBA in Malaysian Schools: The concern of the English teachers', *Journal of US-China Education Review*, 8(10): 1-15.
- Mansor, A.N., H.L. Ong, M.S. Rasul and R.A. Raof, 2013. 'The Benefits of School-Based Assessment', *Asian Social Science*, 9(8): 101-106.
- Annie, T., 2011. 'Exploring students' perception of and reaction to feedback in school-based assessment', *Malaysian Journal of ELT Research*, 7(2): 104-147.
- Wiliam, D., C. Lee, C. Harrison and P. Black, 2004. 'Teachers developing assessment for learning: impact on student achievement', *Assessment in Education: Principles, Policy and Practice*, 11(1): 49-65.
- Christina, L.W.W., 2009. 'The beneficial wash-back of the SBA component on the speaking performance of students', *Hong Kong Teachers' Center Journal*, 8: 18-29.
- Koh, K. and R.L. Velayutham, 2009. Improving teachers' assessment literacy in Singapore schools. Available at: http://www.nie.edu.sg/niecma/attachments/topic/14e8a745eaKV/NIE_research_brief_09_002.pdf (Accessed: 12 November 2013).
- UMS, 2011. National Education Philosophy. Available at: <http://www.frst.unimas.my/academic-programme-mainmenu-37/national-educational-philosophy-mainmenu-122.html> (Accessed: 9 June 2013).
- IBE, 2011. World Data on Education. Available at: http://www.ibe.unesco.org/fileadmin/user_upload/Publications/WDE/2010/pdf-versions/Malaysia.pdf (Accessed: 22 January 2015).

19. JPM, 2010. Rancangan Malaysia KeSepuluh. Available at: http://www.pmo.gov.my/dokumenattached/RMK/RMK10_Mds.pdf (Accessed: 5 November 2014).
20. Stufflebeam, D.L., 1971a. 'The relevance of the CIPP Evaluation Model for Educational Accountability', Annual Meeting of the American Association of School Administrators. Available at: <http://eric.ed.gov/PDFS/ED062385.pdf> (Accessed: 3 September 2012).
21. Stufflebeam, D.L. and A.J. Shinkfield, 1985. Systematic evaluation: A self-instructional guide to theory and practice. Boston: Kluwer Academic Publishers.
22. Stufflebeam, D.L., 2003. 'The CIPP Model for Evaluation' in Kellaghan, T and Stufflebeam, D. L., International Handbook of Educational Evaluation, 31-62, Dordrecht: Kluwer Academic Publishers.
23. Stufflebeam, D.L., 1971b. Educational evaluation and decision-making. Itasca, Illinois: F.E. Peacock Publisher, Inc.
24. Stufflebeam, D.L., 2001. Evaluation Models: New Directions for Evaluation, 7-98, John Wiley and Sons.
25. Isaac, S. and W.B. Michael, 1982. 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.
26. Clinton, J., 2001. The Power of Evaluation Readiness in Program Evaluation, Auckland UNITEC: School of Education.
27. Dalkey, N.C. and O., Helmer, 1963. 'An experimental application of the Delphi method to the use of experts', Management Science, 9(3): 458-467.
28. Zhang, G., N. Zeller, R. Griffith, D. Metcalf, J. Williams, C. Shea and K. Misulis, 2011. 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, 15(4): 57-84.
29. Anton, D.G. and P.N. Jordan, 2002. 'School evaluation for quality improvement' An ANTRIEP report. UNESCO: International Institute for Educational Planning, Kuala Lumpur, Malaysia, 2-4 July 2002. Available at: <http://unesdoc.unesco.org/images/0013/001398/139804e.pdf> (Accessed: 5 November 2011).
30. Lembaga Peperiksaan Malaysia, 2011. Panduan dan Peraturan PBS. Putrajaya. Available at: <http://www.scribd.com/faridooi/d/81532201-Panduan-Dan-Peraturan-Pbs-2011>. (Accessed: 22 November 2011).
31. IowaCORE, 2011. Literature Review Assessment For Learning. Available at: http://www.gwaea.org/iowacorecurriculum/docs/AssessmentForLearning_LitReviewFinal.pdf (Accessed: 25 November 2014).
32. Wiliam, D., 2009. Assessment for learning: why, what and how? (An inaugural professorial lecture), London: Institute of Education.
33. Rust, C., B. Donovan and M. Price, 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.
34. Heritage, M., 2007. Formative Assessment Model. Assessment and Accountability Center (AACC/ CRESST), University of California, Los Angeles. Available at: http://www.gwaea.org/iowacorecurriculum/docs/AssessmentForLearning_LitReviewFinal.pdf (Accessed: 9 July 2013).
35. Sadler, D.R., 1989. Formative assessment and the design of instructional systems', Instructional Science, 18: 119-144.
36. Popham, W.J., 2008. Transformative Assessment. Alexandria, Virginia: ASCD.
37. Hair, J.F., W.C. Black, B.J. Babin and R.E. Anderson, 2010. Multivariate Data Analysis. 7th edn. Englewood Cliffs, NJ: Prentice Hall.
38. Hair, J.F., W.C. Black, B.J. Babin, R.E. Anderson and R.L. Tatham, 2006. Multivariate Data Analysis. 6th edn. New Jersey: Prentice-Hall International, Inc.
39. Byrne, B.M., 2010. Structural Equation Modelling with AMOS: Basic concepts, application and programming. New York: Routledge.
40. Chin, W.W., 1998. 'Issues and opinion on structural equation modelling', MIS Quarterly, 22(1): 7-16.
41. ARG, 2002. Assessment for Learning: 10 Principles. University of Cambridge: Assessment Reform Group.
42. Suzana, A.M. and A. Jamil, 2012. 'Penggunaan teknik pentaksiran formatif dalam subjek BM Darjah satu: Kajian Kes', Malay Language Education Journal, 2(1): 17-30.
43. Wei, L., 2010. 'Formative assessment: Opportunities and Challenges', Journal of Language Teaching and Research, 1(6): 838-841.

44. Bernama, 2014. SinarHarian. Available at: <http://www.sinarharian.com.my/nasional/penambahbaikan-pbs-diterima-golongan-pendidik-1.273118> (Accessed 13 June 2013).
45. Leahy, S. and D. Wiliam, 2009. Annual meeting of the American Educational Research Association held at San Diego, CA. Available at: www.dylanwiliam.org/Dylan_Wiliams.../KKLTT%20paper%20v2 (Accessed: 6 November 2013).
46. Newmann, F., B. King and P. Youngs, 2000. 'Professional development that addresses school capacity: Lessons from Urban Elementary Schools', *American Journal of Education*, 108: 259-299.
47. Norzila, M.Y., 2013. Cambridge Horizons. Available at: <http://www.cambridgeassessment.org.uk/images/139719-sba-seminar-papers.pdf> (Accessed: 14 November 2013).
48. Heritage, M., 2008. 'Learning Progressions: Supporting Instruction and Formative Assessment', Available at: www.ccsso.org/Documents/2008/Learning_Progressions_Supporting_2008.pdf (Accessed: 9 July 2013).