The implementation of School-Based Assessment System in Malaysia: A study of teacher perceptions

Nor Hasnida Che Md Ghazali[1]

[1]Faculty of Education and Human Development, Universiti Pendidikan Sultan Idris, Tanjong Malim, 35900, Perak

Correspondence: Nor Hasnida Che Md Ghazali (email: hasnida@fppm.upsi.edu.my)

Abstract

The school-based assessment system is a holistic assessment system conducted in school by subject teachers to assess the students' cognitive (intellect), affective (emotional and spiritual) and psychomotor (physical) aspects in line with Malaysia's National Philosophy of Education and the Standards-based School Curriculum. This study evaluated the implementation of SBA in Malaysian schools in the course of preparing a summative report on the effectiveness of the system. The model proposed examined the interrelationship between the evaluation dimensions (input, process and product). It was hypothesized that process would be positively associated with product, and input with both process and product. Although SBA is still in its infancy it is becoming increasingly important that it be evaluated considering its impact on students' achievement in an upper middle-income developing country like Malaysia. The study took nearly two years. A self-administered questionnaire was designed based on the Daniel Stufflebeam CIPP (context-input-process-product) evaluation format. The primary data were derived from a total of 776 primary and secondary school teachers who have been sampled using a stratified random sampling of schools. The results revealed a reasonable fit with the SBA evaluation model with an interrelationship between the three dimensions of evaluation (input, process and product). Theoretical, methodological and practical implications suggested the importance of the findings to different audiences.

Keywords: CIPP model, evaluation dimensions, holistic assessment system, school-based assessment, SEM procedures, teacher perception

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 (Fullan, 2011). It is also aimed in providing students with competencies and higher order skills which suits the 21st century challenges (Branden, 2012). 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 (Stiggins, 2005), not helping in identifying the students' learning needs or providing information in improving instruction (Trumbull and Lash, 2013) or regulate incapability in assessing skills such as problem-solving skills, orally expressing thoughts or school behavior (Begum and Farooqui, 2008). Recently, the trend of assessment system in some countries is changing with the introduction of formative assessment. 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 (Mercurio, 2008) and in Finland and Sweden, its implementation has begun in the early 1970s (Darling-Hammond and McCloskey, 2008). 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 (Sanders and Sullin, 2006). 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 (Majid, 2011) or product evaluation to identify the benefits of SBA (Mansor *et al.,* 2013). Other researches look at both evaluation dimensions; process and product as in Annie (2011) which determines students' perception of and reaction to feedback or relating process and product (Christina, 2009) as well as the input and product (Koh and Velayutham, 2009). To date, studies that are related to all the four evaluation dimensions of SBA are non-existent.

Literature review

Frameworks for evaluating SBA

According to the CIPP model developed by Daniel Stufflebeam, any system or project could be evaluated in four dimensions – context, input, process and product (Stufflebeam, 1971). A program can be evaluated in a single dimension or in a few dimensions altogether (Stufflebeam and Shienkfield, 1985). CIPP model has been used to evaluate various educational programs and projects from various disciplines (Stufflebeam, 2003a). Furthermore, each dimension in the evaluation should serve a particular decision as evaluation involves decision-making process (Stufflebeam, 1971a). 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 (Stufflebeam, 1971a). In other words, it emphasizes on assessing needs, assets and problems within a defined environment (Stufflebeam, 2002). It is like asking, 'What should we do to evaluate this program?', 'Which objectives should be obtained?' (Isaac and Michael, 1982), 'Have the important needs addressed accordingly?' (Stufflebeam, 2003) or 'What is the context of the program and how will it match with the target population?' (Clinton, 2001). Methods of collecting data for this purpose could be surveys, interviews, document reviews (Dalkey and Helmer, 1963) or through focus group, website, journals or standardized testing (Clinton, 2001). 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 (Stufflebeam, 1971a). 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?' (Stufflebeam, 2003), and 'What input have been received from the ministry?' or 'What professional development did the providers receive?' (Clinton, 2001). The data for input evaluation could be accumulated using surveys, interviews, websites or journals with the majority of data come from the existing documentation (Clinton, 2001).

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 (Zhang *et al.,* 2011). Important questions to be asked for example, 'Are we doing it correctly?' and 'How adequately are these strategies or procedures working?' (Isaac and Michael, 1982). 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 (Stufflebeam, 1971a). Interview is recognized as the best means of collecting data for process evaluation (Clinton, 2001). 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?' (Isaac and Michael, 1982) or 'Did the effort succeed?' (Stufflebeam, 2003). 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 (Isaac and Michael, 1982). Recent studies which is based on the principles of Stufflebeam's CIPP Model are research in the contexts of the English curriculum evaluation (Hakan and Seval, 2011), evaluating the quality of undergraduate hospitality, tourism and leisure program (Horng *et al.,* 2009) or the nano-technology curriculum evaluation by Tseng *et al.* (2010).

The assessment system in the Malaysian educational context

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 the Malaysian Examination Syndicate (Lembaga Peperiksaan Malaysia, 2011a). 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 (*AfL*) and assessment of learning (*AoL*) 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 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:
i)   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').
ii) 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')
iii) 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').

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 the evaluation dimensions (input, process and product). It is hypothesized that process is positively associated with product, and input is positively associated with both process and product.

Methodology

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 (Standard Error of the Mean) in addressing the research objectives as suggested by Hair *et al.* (2010). 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 (Hair *et al.,* 2006). Furthermore, SEM is also capable in correcting measurement error by providing estimation of error variance which could not be conducted by traditional multivariate procedures (Byrne, 2010).

Instrumentation

The instrument used in this study is a questionnaire designed by the researcher. Items are developed based on the operational development for each construct. Constructs are developed from past literature reviews especially the characteristics listed by Stufflebeam in the CIPP Model concerning input, process and product evaluation. In addition, items from the instruments from previous research are matched whenever appropriated since there is no instrument on SBA evaluation which includes all the dimensions as suggested by Stufflebeam. The researcher focuses more on the instruments from Asian countries like Hong Kong, Singapore, Australia and Malaysia to ensure that the items suit the Asian context. For example, items in Faizah *et al.* (2011) such as 'Not having enough time' or 'Time-spent on non-academic matters related to SBA' are considered to be included as items in the questionnaire. Furthermore, items generated by the ministry were also considered. Then, the instrument went through pilot testing to check for validity, reliability and practicality of the instrument. First, the instrument was validated by the professor who is expert in measurement and evaluation. Then, it went through translation process. After that, all the items were checked for their internal consistency reliability, followed by an independent exploratory factor analysis (EFA) using the principal component analysis (PCA) with a rotation called Direct Oblimin was conducted on the questionnaire. Finally, out of 71 items, 68 items were retained.

Results and discussion

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. The CFA results reporting for the measurement model are shown in Table 1.

Table 1. The CFA results reporting for the measurement model

2nd order factor	1st order factor	Standardized Factor Loading			
Input dimension	Appr	0.925			
	Suit	0.824			
1st order factor	Item	Standardized Factor loading (>0.5)	CR Alpha (>0.7)	CR (>0.6)	AVE (>0.5)
Appr	a17	0.523	0.600	0.6354	0.4780
	a18	0.826			
Suit	a19	0.924	0.752	0.8125	0.6042
	a20	0.839			
	a21	0.505			

2nd order factor	1st order factor	Standardized Factor Loading			
Process1	Attitude	0.998			
	Understanding	0.962			
	Skills	0.614			
1st order factor	Item	Standardized Factor loading (>0.5)	CR Alpha (>0.7)	CR (>0.6)	AVE (>0.5)
Attitude	a1	0.768	0.908	0.907	0.498
	a2	0.735			
	a3	0.763			
	a4	0.787			
	a5	0.582			
	a6	0.753			
	a7	0.813			

1st order factor	Item	Standardized Factor loading (>0.5)	CR Alpha (>0.7)	CR (>0.6)	AVE (>0.5)
	a8	0.501			
	a9	0.630			
	a10	0.615			
Understanding	a11	0.870	0.679	0.673	0.429
	a12	0.601			
Skills	b24ii	0.810	0.813	0.902	0.698
	b24iii	0.849			

2nd order factor	1st order factor	Standardized Factor Loading			
Process2	Moderation	0.623			
	Monitoring	0.581			
1st order factor	Item	Standardized Factor loading (>0.5)	CR Alpha (>0.7)	CR (>0.6)	AVE (>0.5)
Moderation	b27i	0.827	0.915	0.831	0.788
	b27ii	0.942			
	b27iii	0.891			
Monitoring	b28i	0.815	0.758	0.763	0.521
	b28ii	0.734			
	b28iii	0.601			

2nd order factor	1st order factor	Standardized Factor Loading			
Process3	Role of SBA	0.733			
	Importance of SBA	0.774			
1st order factor	Item	Standardized Factor loading (>0.5)	CR Alpha (>0.7)	CR (>0.6)	AVE (>0.5)
Role of SBA	d30i	0.584	0.849	0.861	0.615
	d30ii	0.875			
	d30iii	0.914			
	d30iv	0.720			
Importance of SBA	d31i	0.815	0.945	0.943	0.675
	d31ii	0.883			
	d31iii	0.909			
	d31iv	0.901			
	d31v	0.787			
	d31vi	0.829			
	d31vii	0.622			
	d31viii	0.789			

1st order factor	Item	Standardized Factor loading (>0.5)	CR Alpha (>0.7)	CR (>0.6)	AVE (>0.5)
Challenges	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 | | | |

2nd order factor	1st order factor	Standardized Factor Loading			
Product	Attitude	0.924			
	Knowledge	0.784			
	Motivation	0.958			
1st order factor	**Item**	**Standardized Factor loading (>0.5)**	**CR Alpha (>0.7)**	**CR (>0.6)**	**AVE (>0.5)**
Attitude	e32i)	0.855	0.861	0.8674	0.6872
	e32ii)	0.894			
	e32iii)	0.729			
Knowledge	e33i)	0.863	0.895	0.8974	0.8142
	e33ii)	0.940			
Motivation	e34i)	0.829	0.885	0.8885	0.7273
	e34ii)	0.796			
	e34iii)	0.928			

However, when 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. Out of the six measurement models (input dimension, process1, process2, process3, challenges and product dimension) formed in the study, only 3 models are satisfied with the multivariate normality distribution assumption suggested by Zainuddin (2012). In order to improve the multivariate normality of the measurement models, few extreme outliers are deleted. Out of 776 samples, 9 cases are deleted and then few items with low standardized estimate value are also deleted. Finally, 38 items are retained. Finally, all the six measurement models have shown an excellent fit to the data, valid and reliable. Characteristics for each of the six final measurement model tested are presented in Table 2.

Table 2. 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

Measurement Model	Fit indices value	Number of items	Multivariate kurtosis
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 modeling

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 (Hair *et al.,* 2006). SEM technique which is theory-driven, is used to determine the interrelationships between evaluation dimensions in a full structural model. Structural model (see Fig. 1) is assembled from the measurement models based on theoretical interrelationships among the constructs of grounded in empirical research (Byrne, 2010).
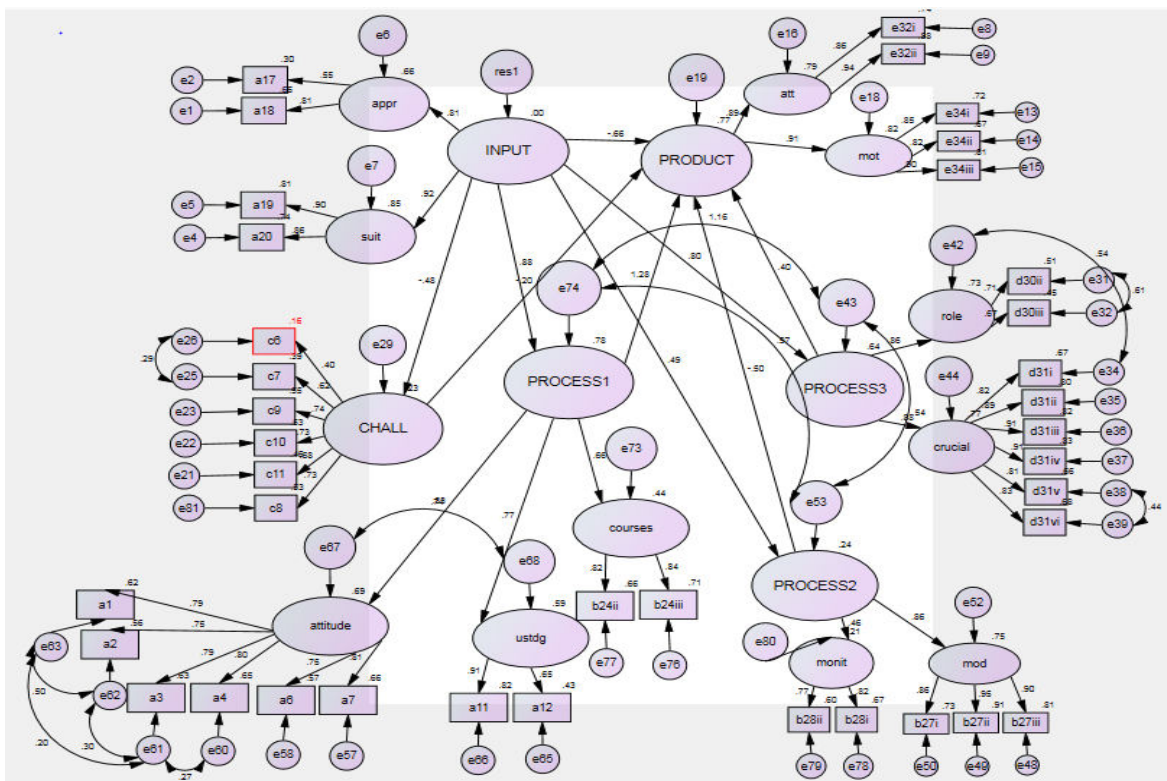
Figure 1. *The hypothesized model – initial structural model*

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 show $X^2/df = 4.213$, 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 Procees1 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 (Byrne, 2010). 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 (2010), 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.
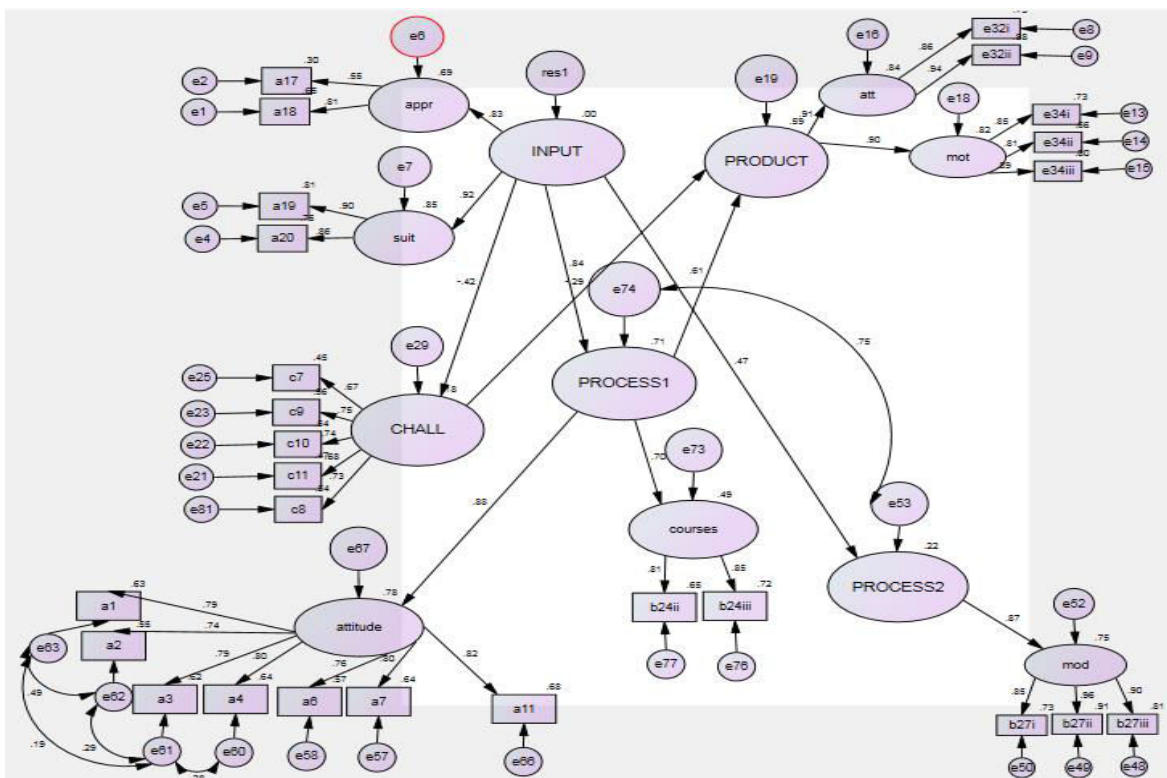

Figure 2. *The hypothesized model – final structural model*

Conclusion

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, the CIPP model is suitable to be used as a means of explaining an evaluation process of SBA. Elements of the 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 (2002) 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 (Suzana & Jamil, 2012). 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 (2010) 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 (Bernama, 2014). 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.

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.

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 (Byrne, 2010). 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 (Norzila, 2013). 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 (2008) believe that the three components are the major components in implementing formative assessment effectively.

References

Aidarwati MB, Abdul Ghani A (2013) A Comparative Analysis of Primary and Secondary School Teachers in the implementation of SBA. *Malaysian Journal of Research* **1**(1), 28-36.

Annie T (2011) Exploring students' perception of and reaction to feedback in school-based assessment. *Malaysian Journal of ELT Research* **7**(2), 104-147.

Anton DG, Jordan PN (2002) School evaluation for quality improvement, An ANTRIEP report. UNESCO: International Institute for Educational Planning, Kuala Lumpur, Malaysia, 2-4 July 2002. Available from: http://unesdoc.unesco.org/images /0013/001398/139804e.pdf.

ARG (2002) *Assessment for Learning: 10 Principles*. Assessment Reform Group, University of Cambridge.

Begum M, Farooqui S (2008) School Based Assessment: Will it Really Change the Education Scenario in Bangladesh?. *International Education Studies* **1**(2), 45-55.

Bernama (2014) Sinar Harian. Available from: http://www.sinarharian.com.my/nasional /penambahbaikan -pbs-diterima-golongan-pendidik-1.273118.

Black P (1995) Assessment and feedback in science education. *Studies in Educational Evaluation* **21**, 257-279.

Branden KVD (2012) Sustainable education: Basic principles and strategic recommendations. *School Effectiveness and School Improvement* **23**(3), 285-304.

Byrne BM (2010) *Structural Equation Modelling with AMOS: Basic concepts, application and programming*. Routledge, New York.

Chin WW (1998) Issues and opinion on structural equation modeling. *MIS Quarterly* **22**(1), 7-16.

Christina LWW (2009) The beneficial wash-back of the SBA component on the speaking performance of students. *Hong Kong Teachers' Center Journal* **8**, 18-29.

Clinton J (2001) *The Power of Evaluation Readiness in Program Evaluation*. School of Education, Auckland UNITEC.

Dalkey NC, Helmer O (1963) An experimental application of the Delphi method to the use of experts. *Management Science* **9**(3), 458-467.

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.

Frye AW, Hemmer PA (2012) Medical Teacher. Available from: http://www .med.unc.edu/aoe/aoe-events/videos-and-materials-from-curricular-sessions/materials-from-curricular-sessions-2014-2015/ program -evaluation-guide.

Fullan M (2011) Choosing the wrong drivers for whole system reform. Available from: http://www.michaelfullan.ca/media /13501655630.pdf.

Gredler ME (1996) *Program evaluation*. Prentice Hall, Upper Saddle River, New Jersey.

Hair JF, Black WC, Babin BJ, Anderson RE, Tatham RL (2006) *Multivariate Data Analysis*. 6th edn. Prentice-Hall International, Inc., New Jersey.

Hair JF, Black WC, Babin BJ, Anderson RE (2010) *Multivariate Data Analysis*. 7th edn. Prentice Hall, Englewood Cliffs, NJ.

Hakan K, Seval F (2011) CIPP Evaluation Model Scale: Development, reliability and validity. *Procedia and Social and Behavioral Sciences* **15,** 592-599.

Helsen K, Van den Bussche E, Vlaeyen JWS, Goubert L (2013) Confirmatory factor analysis of the Dutch Intolerance of Uncertainty Scale: comparison of the full and short version. *Journal of Behavior Therapy and Experimental Psychiatry* **44**, 21–29.

Heritage M (2007) Formative Assessment Model. Assessment and Accountability Center (AACC/ CRESST), University of Carlifornia, Los Angeles. Available from: http://www.gwaea.org/ iowa corecurriculum /docs/ Assessment ForLearning_ LitReviewFinal.pdf.

Heritage M (2008) Learning Progressions: Supporting Instruction and Formative Assessment. Available from: www.ccsso.org/Documents /2008 /Learning_ Progressions_ Supporting_ 2008.pdf.

Horng J, Teng C, Baum T (2009) Evaluating the quality of undergraduate hospitality, tourism and leisure programmes. *JHLSTE* **8**(1), 37-54.

IBE (2011) *World Data on Education*. Available from: http://www.ibe.unesco. org/fileadmin /user_upload/Publications/WDE/2010/pdf-versions/Malaysia .pdf.

IowaCORE (2011) *Literature Review Assessment For Learning*. Available from: http://www.gwaea.org /iowacorecurriculum /docs/Assessment ForLearning_ LitReviewFinal.pdf.

Isaac S, Michael WB (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. EdITS Publishers, California.

JPM (2010) Rancangan Malaysia KeSepuluh. Available from: http://www.pmo.gov. my/dokumenattached /RMK/RMK10_Mds.pdf.

Kahan B (2008) Excerpts from Review of Evaluation Frameworks. Available from: http://idmbest practices.ca/pdf/evaluation-frameworks-review.pdf.

Koh K, Velayutham RL (2009) Improving teachers' assessment literacy in Singapore schools. Available from: http://www.nie.edu.sg/niecma /attachments /topic /14e8a745 eaKV /NIE research brief _09_002.pdf.

Leahy S, Wiliam D (2009) *Annual meeting of the American Educational Research Association held at San Diego, CA*. Available from: www.dylanwiliam .org/Dylan_Wiliams.../KKLTT%20paper%20v2.

Lembaga Peperiksaan Malaysia (2011) *Panduan dan Peraturan PBS. Putrajaya*. Available from: http://www.scribd.com/faridooi/d/81532201-Panduan-Dan-Peraturan-Pbs-2011.

Majid F (2011) SBA in Malaysian Schools: The concern of the English teachers. *Journal of US-China Education Review* **8**(10), 1-15.

Mansor AN, Ong HL, Rasul MS, Raof RA (2013). The Benefits of School-Based Assessment. *Asian Social Science* **9**(8), 101-106.

Mercurio A (2008) Re-imagining school-based assessment at the upper secondary level, SACE Board of South Australia. Available from: http://www. iaea2008 .cambridgeassessment.org.uk/ca/digital Assets /180434-Mercurio.pdf.

Ministry of Education (2011) Penambahbaikan UPSR dan PMR. Available from: http://www.slideshare.net /Fadzliaton/penilaian-berasaskan-sekolah-pbs.

Newmann F, King B Youngs P (2000) Professional development that addresses school capacity: Lessons from Urban Elementary Schools. *American Journal of Education* **108**, 259-299.

Norzila MY (2013) *Cambridge Horizons*. Available from: http://www.cambridge assessment .org.uk/images/139719-sba-seminar-papers.pdf.

Owston R (2008) *Models and Methods for Evaluation. Handbook of Research on Educational Communications and Technology*. 3rd edn. Routledge, New York, NY.

Popham WJ (2008) *Transformative Assessment*. ASCD, Alexandria, Virginia.

Rust C, ODonnovan 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.

Sadler DR (1989) Formative assessment and the design of instructional systems. *Instructional Science* **18**, 119-144.

Sanders J, Sullins C (2006) *Evaluating School Programs: An Educator's Guide* (3rd ed.). Corwin Press, Thousand Oaks, CA.

Sidek R, M Saleh S (2010) *UTM Institutional Repository*. Available from: http://eprints.utm. My /10502/1/KESEDIAAN_GURU_SEKOLAH_MENENGAH_DI_DAERAH_KOTA_BAHARU.pdf.

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 DL (1971a) The relevance of the CIPP Evaluation Model for Educational Accountability, Annual Meeting of the American Association of School Administrators. Available from: http://eric .ed.gov/PDFS/ED062385.pdf.

Stufflebeam DL (1971b) *Educational evaluation and decision-making*. F.E. Peacock Publisher, Inc., Itasca, Illinois.

Stufflebeam DL (2001) *Evaluation Models: New Directions for Evaluation*, 7-98. John Wiley and Sons.

Stufflebeam DL (2003) The CIPP Model for Evaluation. In: Kellaghan T, Stufflebeam DL. *International Handbook of Educational Evaluation*, 31-62. Kluwer Academic Publishers, Dordrecht.

Stufflebeam DL, Shinkfield AJ (1985) *Systematic evaluation: A self-instructional guide to theory and practice*. Kluwer Academic Publishers, Boston.

Suzana AM, Jamil A (2012) Penggunaan teknik pentaksiran formatif dalam subjek BM Darjah satu: Kajian Kes. *Malay Language Education Journal* **2**(1), 17-30.

Tamir P (1998) Assessment and evaluation in science education: Opportunities to learn and outcomes. In Fraser BJ, Tobin KG (eds) *International handbook of science education*. Kluwer Academic, Dordrecht, The Netherlands.

Trumbull E, Lash A (2013) *Understanding Formative Assessment: Insights from Learning Theory and Measurement Theory*. West Ed, San Francisco.

Tseng K, Diez CR, Lou S, Tsai H (2010) Using the Context, Input, Process and Product model to assess an engineering curriculum. *World Transactions on Engineering and Technology Education* **8**(3), 256-261.

UMS (2011) *National Education Philosophy*. Available from: http://www. frst.unimas.my /academic-programme-mainmenu-37/national-educational-philosophy-mainmenu-122.html.

Wei L (2010) Formative assessment: Opportunities and Challenges. *Journal of Language Teaching and Research* **1**(6), 838-841.

Wiliam D (2001) What is wrong with our educational assessments and what can be done about it. *Education Review* **15**(1).

Wiliam D (2009) *Assessment for learning: why, what and how?* (An inaugural professorial lecture). Institute of Education, London.

Wiliam D, Lee C, Harrison C, Black P (2004) Teachers developing assessment for learning: Impact on student achievement. *Assessment in Education: Principles, Policy and Practice* **11**(1), 49-65.

Zhang G, Zeller N, Griffith R, Metcalf D, Williams J, Shea C, Misulis K (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.

Zinovieff MA (2008) *UN Task Force*. Available from: http://esa.un.org/techcoop /fellowships /SFO Meeting/ParticipantArea /BackgroundDocuments /6_REVIEWm %20 report %20FINAL%20.pdf.