UNIVERSITY OF SOUTHAMPTON

FACULTY OF SOCIAL, HUMAN AND MATHEMATICAL SCIENCES

Southampton Education School

An Evaluation of The Implementation of The School-Based Assessment System in Malaysia

by

Nor Hasnida Che Md Ghazali

Thesis for the degree of Doctor of Philosophy

August 2015

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF SOCIAL, HUMAN AND MATHEMATICAL SCIENCES

Southampton Education School

Thesis for the degree of Doctor of Philosophy

AN EVALUATION OF THE IMPLEMENTATION OF THE SCHOOL-BASED ASSESSMENT SYSTEM IN MALAYSIA

Nor Hasnida Che Md Ghazali

The implementation of the school-based assessment (SBA) system is an effort in improving human capital development in a holistic manner and also to lessen the negative influences of exam-oriented education systems on students. The need to evaluate the new system is of critical importance as the SBA system is still in a relatively early stage of development. The research is conducted to evaluate the implementation of the SBA system in Malaysian schools by using Stufflebeam's CIPP (context-input-process-product) Model. According to Stufflebeam, any programme could be evaluated from four dimensions of context, input, process and product. Applying the CIPP Model and supported by relevant learning theories - behaviourism, Piaget's learning theory, constructivism, multiple intelligence and brain research and the assessment models - formative model, the logic model and the SCAP (Social Constructivist Assessment Process) Model, this research examines the evaluation instrument in order to choose valid, just and quality items. The research also investigates the interrelationship of all the evaluation dimensions in the context of SBA implementation. Although much has been done to investigate the relationship between dimensions in this context, none has related all the dimensions together. Additionally, this research incorporated different types of school (urban-rural) and school category (secondary-primary) as variables, which possibly moderated the relationship between the evaluation dimensions. A stratified random sampling technique was applied to collect data from 776 teachers in primary and secondary schools in Kelantan, one of the states in the north-east of Peninsular Malaysia. All the dimensions of evaluation were measured using a questionnaire developed by the researcher. A structural equation modelling software called AMOS (Analysis of Moment Structures) was applied to test all the hypotheses of the study. Results of the study showed i) input dimension has met the desired result, process dimension could still be strengthened and production dimension showed that participants' needs were partially met; ii) evidence of measurement models for input, process and product. Results failed to reject the hypothesised relationships between input and process dimensions and also between process and product dimensions; iii) input had a significant relationship with process, some process components had a significant relationship with product; and there was an indirect relationship between input and product; and iv) school category was found to moderate the relationships between dimensions. Theoretical, methodological and practical implications are discussed. In short,

this study provides support for the effectiveness of SBA implementation in schools.

Table of Contents

List of tables

List of figures

List of accompanying materials

DECLARATION OF AUTHORSHIP

I, Nor Hasnida Che Md Ghazali declare that this thesis titled, 'An evaluation of the implementation of the School-based Assessment System in Malaysia' and the work presented in it are my own and has been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. None of this work has been published before submission

Signed: ……..

Date:……

Acknowledgements

In the name of God, the Most Gracious and the Most Merciful.

All praises to God for the strengths and His blessing in completing this thesis. Special appreciation goes to my supervisor, Professor Daniel Muijs for his supervision and constant support. His invaluable help of constructive comments and suggestions throughout the thesis have contributed to the success of this research.

Appreciation also goes to all the staffs in the School of Education, University of Southampton and Sultan Idris Education University, Malaysia for the support and help towards my postgraduate affairs.

Last but not least, my deepest gratitude goes to my beloved family for their endless love, prayers and encouragement. To those who indirectly contributed in this research, your kindness means a lot to me. Thank you very much.

Definitions and Abbreviations

AaL	Assessment as Learning
AfL	Assessment For Learning
AoL	Assessment Of Learning
CA	Centre Assessment
CE	Central Examination
CFA	Confirmatory Factor Analysis
EFA	Exploratory Factor Analysis
JPN	State Education Department
MES	Malaysian Examination Syndicate
MOE	Ministry of Education
NEAS	National Education Assessment System
NPE	Natinal Philosophy of Education
PPD	District Education Department
Psi	Psychometric Assessment
PASCA	Physical Activities, Sports and Co-Curricular Assessment
PCA	Principal Component Analysis
PMR	Lower Secondary Examination
QA	Quality Assurance
SA	School Assessment
SBA	School Based Assessment
SPPBS	Management System of SBA
SPM	Malaysian Certificate of Education
STAM	Malaysian Higher Islamic Religious Certificate
STPM	Malaysian Higher School Certificate Examination
UPSR	Primary School Evaluation Test

Chapter 1: Introduction

This study evaluates the implementation of the school-based assessment (SBA) system in primary and secondary schools in Malaysia. The SBA has been introduced to Malaysia's education system since 2011. The study starts by highlighting the growing attention and interest of the shifting process of an assessment system, the conceptual framework of the study and then follows with research questions and hypotheses of the study. Chapter two discusses the education system in Malaysia and focuses more on the new assessment system called the SBA system. Chapter three provides a literature review of the assessment system, school improvement, related models and theories, program evaluation and the dimensions of evaluation. Chapter four then provides the methodology of the study. Chapter five presents the analyses and results of the study in answering the research questions proposed. Finally, Chapter six illustrates the main findings of the study, discussion of the findings, implications of the results, limitations of the study and makes suggestions for future work.

1.1 Background to the Study

An education system is a fundamental aspect in building a developed nation. The structure of the current Malaysian education system was inherited from the western colonial powers and many of its characteristics reflect other Southeast Asian Ministers of Education Organization countries like Brunei, Indonesia, Philippines and Singapore (SEAMEO Secretariat, 2001). The Malaysian education system consists of pre-tertiary education (preschool, primary and secondary education) followed by tertiary or higher education. Starting at the age of five or six years old, most children go to preschool which serves as a medium in providing basic education, such as basic communication skills and English language, foster love for the country and moral values and to develop critical thinking skills amongst children (Ramlee, 2009). Primary education takes a period of six years (7 to 12 years old) with the admission age of seven years old. During the six year period, the objectives are to master the '3Ms' which are 'reading, writing and arithmetic' in the first three years, followed by the reinforcement of the '3Ms'. Pupils also need to acquire general knowledge, pre-vocational education and personality, attitude and social values

development (Ramlee, 2009). Secondary education then consists of five years of learning (13 to 17 years old), encompassing three years of lower secondary and two years of upper secondary level. Then follows a two-year period in post-secondary education either joining matriculation, technical and vocational, short term courses or Form six.

The objectives of the education system as manifested by Malaysia's National Philosophy of Education (NPE) (UMS, 2011, p.1) formulated in 1988 stated that:

> Education in Malaysia is an on-going effort towards further developing the potential of individuals in a holistic and integrated manner so as to produce individuals who are intellectually, spiritually, emotionally and physically balanced and harmonious, based on a firm belief in and devotion to God. Such an effort is designed to produce Malaysian citizens who are knowledgeable and competent, who possess high moral standards, and who are responsible and capable of achieving a high level of personal well-being, as well as being able to contribute to the betterment of the family, the society and the nation at large (UMS, 2011, p.1).

This philosophy is implemented in all schools all over Malaysia. In order to achieve its objective, the Malaysian education system implements a national curriculum that aims to develop individuals in a holistic and integrated manner to produce a well-balanced community (IBE, 2011). Additionally, the five-year plan in the Tenth Malaysian Plan (2011-2015) provides a guideline in implementing the mission towards achieving a developing and high-income country as indicated by the Vision 2020 (JPM, 2010) and includes the 'Government Transformation Program [sic]' aspiration and the 'Economic New Model' to develop the country in five years ahead. In the fifth chapter of the Plan, which is to 'develop and retain world-class human capital', the stated aim is to improve the education of the people starting from early education, basic education, tertiary education until the working environment by using three core strategies:

- Reform the education system to improve students' performance
- Increase people's skills for employability, and
- Reform the labour market to make Malaysia a high income country

The first core strategy, which is consistent with the interests of this study, is to reform the education system based on the NPE and to focus more on the

involvement in sports and co-curriculum. Furthermore, the education system is expected to inculcate values and ethics in achieving Vision 2020 and also to consider the use of ICT and communication to increase students' creativity, innovation and skill (JPM, 2010). Education systems around the world are also going through reforms in students' performance which is aimed at raising the bar for all students and closing the gap for lower performing groups (Fullan, 2011) and also to provide students with the competencies and higher order skills to prepare them for the challenges of the twenty-first century (Branden, 2012).

In order to improve the teaching and learning process or even to gain better impact in educational improvement, it is important to take into account the interaction between the three main aspects in education which are curriculum, instruction and assessment (Young and Giebelhaus, 2005). Higher-quality instruction could be supported by an integrated system of curriculum and assessment (Darling-Hammond and Pecheone, 2010). Since assessment plays an important part, considerable attention has been placed on it by various parties. For example, lately in the United States of America, formative assessment is seen as a strategy for improvement that links the three aspects mentioned above (Clark, 2011).

1.2 Statement of the Problem

There is a large body of literature which studies the negative impact of the traditional concept of assessment on student learning. A traditional concept of assessment, which focuses most on public examination, has long been practised in the education system (Wiliam, 2001). Traditional assessment practises are not integrated into the teaching and learning process and could not give much help in improving learning or diagnosing the strength and weakness of students. The purpose of assessment has previously focused more on selection and certification for students (Wiliam, 2001) or on determining the outcome of a particular curriculum, programme or policy with little concern for any actionable information on improvement, often meant to compare with the predetermined objective (Caffrey, 2009). According to Fan (2011), students are assessed purely on their academic achievement, including knowledge and skills and are routinely practised in a time-limited situation,

tool-limited and venue-limited test. Then, most probably it will be conducted for marking and grading purposes to give report or selection process.

The traditional concept of assessment is also seen to affect students' emotion and confidence levels and to some extent could negatively influence their inner emotional strength to succeed (Stiggins, 2005). Teachers also tend to focus on those pupils they perceive as better students whom they feel have the higher chance to pass and neglect some other students, which could impact negatively on students' development (Buhagiar and Murphy, 2008). Wiliam (2001) believed that traditional assessment distort school curricula and produces results which are less reliable and valid.

Focusing on public examination as a form of assessing students has a negative rather than a positive impact on society. Public examination brings more deterioration to students and teachers as it risks producing both passive students who tend to absorb information and a passive type of teachers who tends to concentrate only on rote learning (Mercurio, 2008). It is also incapable in assessing skills like problem solving, orally expressing thoughts, school behaviour and personal and social values of students (Begum and Farooqui, 2008). Testing with multiple-choice items in most public examinations is seen to be a form of assessment that is urging students to recall and recognise discrete facts without analysing these facts critically (Darling-Hammond and McCloskey, 2008). It is also unable to evaluate pupils' broader accomplishments other than academic aspects, as with co-curriculum and sports. Public examinations therefore require teachers to focus more on examination questions rather than developing students' potential. Similarly, the Malaysian public examination is a method that orientates the public to focus on the examination as this makes up the basis for promoting students into higher level of education or for them to be awarded various scholarships (Cheah, 2010), a phenomenon which seems to deviate from the real objectives of the education system as manifested by the NPE as mentioned above. Table 1.1 lists the main features of the traditional and new concepts of classroom assessments for mathematics (Fan, 2011, p. 4).

Table 1.1: Comparison of traditional and new concept of mathematics assessment

Mathematics assessment	Traditional concept	New concept
What (content)	Cognitive domain (mainly knowledge and skill) and the results of learning	Cognitive and affective domains (knowledge, skill, ability and disposition) and both the results and process of learning
Where is it conducted	Within classrooms	Within or outside classrooms
When is it conducted	During class for a block of time	During or after class for days, weeks, months or years
How is it conducted	Conventional way (written test)	Conventional and alternative ways
Why is it conducted	Single purpose for grading and reporting learning results	Multiple purpose for improving teaching and learning
Relationship with learning	Assessment of Learning (AoL)	Assessment for Learning (AfL), Assessment of Learning (AoL) and Assessment as Learning (AaL)

Currently, the trend of the assessment system in Malaysia is changing. The Malaysian education system has started to implement SBA, an assessment system which is conducted in school and is planned, administered, scored and reported in a mannered way based on the procedures from the Malaysian Examination Syndicate (Lembaga Peperiksaan Malaysia, 2011a).

As Black (1998) believed that the terms, methods and procedures used in any assessment and testing system in a country depend on the historical and cultural background of its educational system, so it would be helpful to consider the chronology of the SBA implementation in Malaysia. The beginnings of the SBA can be traced back to the launching of the International Colloquium on the 13th to 15th of September 2005 and the Kuala Lumpur International Conference on Assessment (KLICA) on the 16th to 19th of May 2006, the latter of which used 'Humanising Assessment' as its theme and an idea for the SBA system in Malaysia (Ministry of Education, 2006). The

implementation of the National Education Assessment System was approved in September 2008 as a pilot project incorporating 500 schools. The follow-up meeting by the Cabinet of Ministers No. 3/2009 agreed to the Implementation of the SBA (Ministry of Education, 2011). The implementation of SBA for primary schools started in 2011 with the Year 1 students made the first move with the instruction from the MOE followed by a circular letter KP (BPSH-SPDK) 201/005/01/ Jld. dated January 2011.

Since the SBA system is still in a relatively early stage of development, it seems timely to discuss some of the issues raised. Hence, a comprehensive evaluation is necessary to evaluate the system in all the dimensions previously stated (context, input, process and product). This is concordant with Gredlers' idea on programme evaluation where any educational system has to undergo careful and rigorous examination in order to improve or enhance students' educational experiences (Gredler, 1996). Mitchem *et al.* (2003) asserted that evaluation is becoming important to prevent failures in programme implementation.

1.3 The Conceptual Framework

A model or framework is a conceptual picture that shows the interrelationship between various elements involved in any given activities (Razali, 1987) and are normally presented in the form of flowcharts, web diagrams or other forms of schemata (Leshem and Trafford, 2007). Conceptual frameworks provide a theoretical clarification to support the study and a clearer picture on the purpose of the study and the process of achieving it. A conceptual framework is defined as the current version of the researcher's map of the territory being investigated (Miles and Huberman, 1984, p.33).

When developing an evaluation model to gauge the success or otherwise of a learning or assessment model, it is important to recognise there is no single evaluation design which is perfect or complete. The evaluation model does not only represent the overall evaluation framework but should also relate to the research questions or the purpose of the study (Patton, 1990). Some academics might define evaluation differently according to the purpose of evaluation. Hence, the difference in its purpose will determine the way

evaluation is conducted, the models chosen and the standards used to formulate and prosecute evaluations (Brinkerhoff *et al.,* 1983).

Following a comprehensive review of several evaluation frameworks and taking into consideration the definition and the purpose of evaluation in this study, the *context, input, process and product* (CIPP) Model (Stufflebeam, 1971a) is a suitable model to be used as a framework for this study, the reasons for which will be elaborated in section 3.7.6. The use of the CIPP Model is concordance to the operational definition of evaluation used in this study, which follows the definition provided by Stufflebeam and Shinkfield (1985), who stated that evaluation is a process of gathering information in guiding decision-making and quality assurance or to sum up the worth and merit of a programme. The information gained is useful to various groups of people such as the school administrations, the head teachers or the education officers. Evaluation in this study context is not aiming at collecting an interim continuous report along the implementation process and then providing stakeholders with enough information for them to revise and make improvements. Instead, it is a form of evaluation to assess a completed system to determine the system's success or failure. By using this framework, it is hoped that evaluation would help in providing guidelines to decision makers, in producing records and in the creation of concrete explanations on phenomena that happen during the programme's implementation. But, it still could turn out to be a failure if the authorities do not handle the findings in a correct manner (Stufflebeam and Shinkfield, 2007).

The conceptual framework developed in this study originated from an established model, the CIPP Model by Stufflebeam (1971a). The CIPP Model was first used back in 1965 to provide information for the purpose of decision-making and accountability through the process of evaluation of programmes, projects, products and systems. It includes four essential features which are context, input, process and product, with an ultimate aim of securing an improvement in the education field. CIPP was developed by groups of researchers that based their research from the work of the Ohio State University Evaluation Centre and the Department of Evaluation and Research in the Columbus, Ohio Public Schools.

According to Stufflebeam, evaluation involves decision-making. The context, input, process and product evaluation subsequently serve planning, structuring, implementing and recycling decisions respectively. When 'context evaluation' is involved, it gives an opportunity for the decision makers to plan the programme objectives either to confirm the present objectives, to modify the existing objectives or to develop new objectives. So, the selection of programme objectives influences planning decision. In a simple form, it is like asking oneself, "What should we do to evaluate this program [sic]?" or "Which objectives should be obtained?" (Isaac and Michael, 1982, p.10) or "Were important needs addressed?" (Stufflebeam, 2003). Next is 'input evaluation', which allows decision makers to make decisions on the structure of the programme related to strategies, personnel, resources, procedures or a prospective cost assessment in achieving the programme objectives that have been derived from planning decisions. Input evaluation involves asking questions such as, "How should we evaluate the program [sic]?" or "Which strategies or procedures should be tried?" (Isaac and Michael, 1982, p.10) or "Was the effort guided by a defensible plan and budget?" (Stufflebeam, 2003).

Making decisions on the implementation of a programme during 'process evaluation' means that decision-makers have to decide on everything related to the implementation of already selected designs, strategies or action plan; asking questions such as "Are we doing it correctly?" or "How adequately are these strategies or procedures working?" (Isaac and Michael, 1982, p.10) or "Was the service design executed competently and modified as needed?" (Stufflebeam, 2003). Lastly is the 'product evaluation'. It serves as the programme recycling decisions to determine and examine the specific outcomes of the programme, to conduct a retrospective cost assessment or cost effectiveness assessments and includes such questions as "Should the program be continued or not?" or "How effectively are the goals and objectives being accomplished?" (Isaac and Michael, 1982, p.10) or "Did the effort succeed?" (Stufflebeam, 2003). In other words, it compares outcomes of the programme with its objectives.

The relationship between the types of evaluation and decisions is shown in Figure 1.1 and Table 1.2 (Isaac and Michael, 1982). Figure 1.1 shows the dynamic action of evaluations serving the decisions in the CIPP Model. It is

'dynamic' in the sense that information from any stage of the decision making activity could be provided to the previous stage so that modifications on evaluations could be made. In Table 1.2, the 'Ends' row explains the realization of the objectives which is achieved by the planning and recycling decisions whereas the 'Means' row shows the act of achieving ends which is achieved by the structuring and implementing decisions (Isaac and Michael, 1982). Furthermore, decisions are also categorised under an intended act or an actual one. Planning and structuring decisions are grouped as intended ends and means respectively whereas recycling and implementing decisions are grouped as actual ends and means respectively.

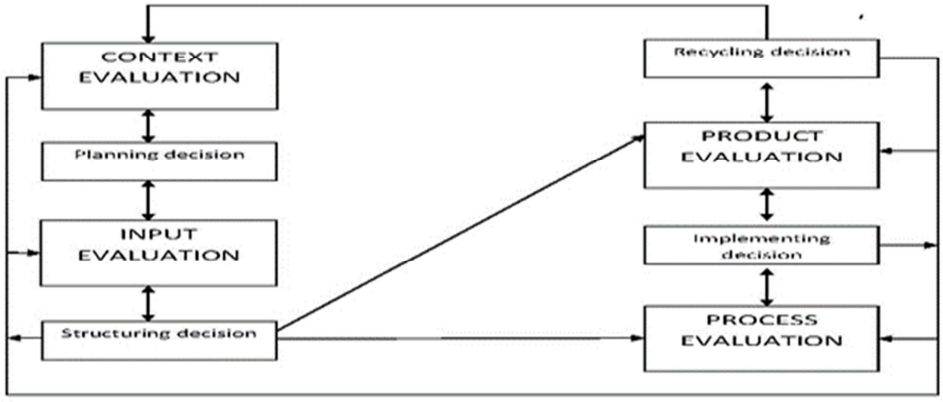


Figure 1.1: Dynamic Action of the CIPP Model

Table 1.2: Types of Evaluation and Decisions

	INTENTIONS	**ACTUALITIES**
ENDS	Planning Decisions supported by Context Evaluation (*What needs to be done?)*	Recycling Decisions supported by Product Evaluation (*Did it succeed?)*
MEANS	Structuring Decisions supported by Input Evaluation (*How should it be done?)*	Implementing Decisions supported by Process Evaluation *(Is it being done?)*

Next, the framework of the CIPP Model on types of evaluation together with their steps in the evaluation process is shown in Table 1.3. The steps involved are delineating, obtaining and providing. Delineating involves outlining questions to be answered and focusing on the information required by

decision makers (Stufflebeam, 1971a). Obtaining involves obtaining relevant information by organizing and analysing information using measurement and statistics and providing is a step which synthesises information to be provided to the decision-makers.

Table 1.3: The Framework of the CIPP Model (Stufflebeam, 1971a)

	EVALUATION TYPES			
STEPS IN EVALUATION PROCESS	**CONTEXT**	**INPUT**	**PROCESS**	**PRODUCT**
Delineate	System variables and values	Problem specifications Design criteria Constraints	Process decision points Milestones Barriers	Effectiveness criteria
Obtain	Performance and judgment data	Identification and analysis strategies	Monitoring of procedures	Primary, secondary and tertiary effects
Provide	Profile of needs, opportunities and problems	Strategies by problem matrix	Progress reports Exception reports	Description and explanation of projects' impact

When using the CIPP Model to evaluate any program, it is best done with all the four components of the model. However, Stufflebeam and Shinkfield (1985) believed that evaluation could be conducted using any one, two or three components. In the context of this doctoral study, evaluation is made using all the four components. Additionally, items or constructs under each dimension are constructed accordingly to suit the CIPP Model interpretations and also based on the findings and instruments from previous studies. Its graphical presentation can be seen in Figure 1.2. This seems to be the conceptual framework for this study.

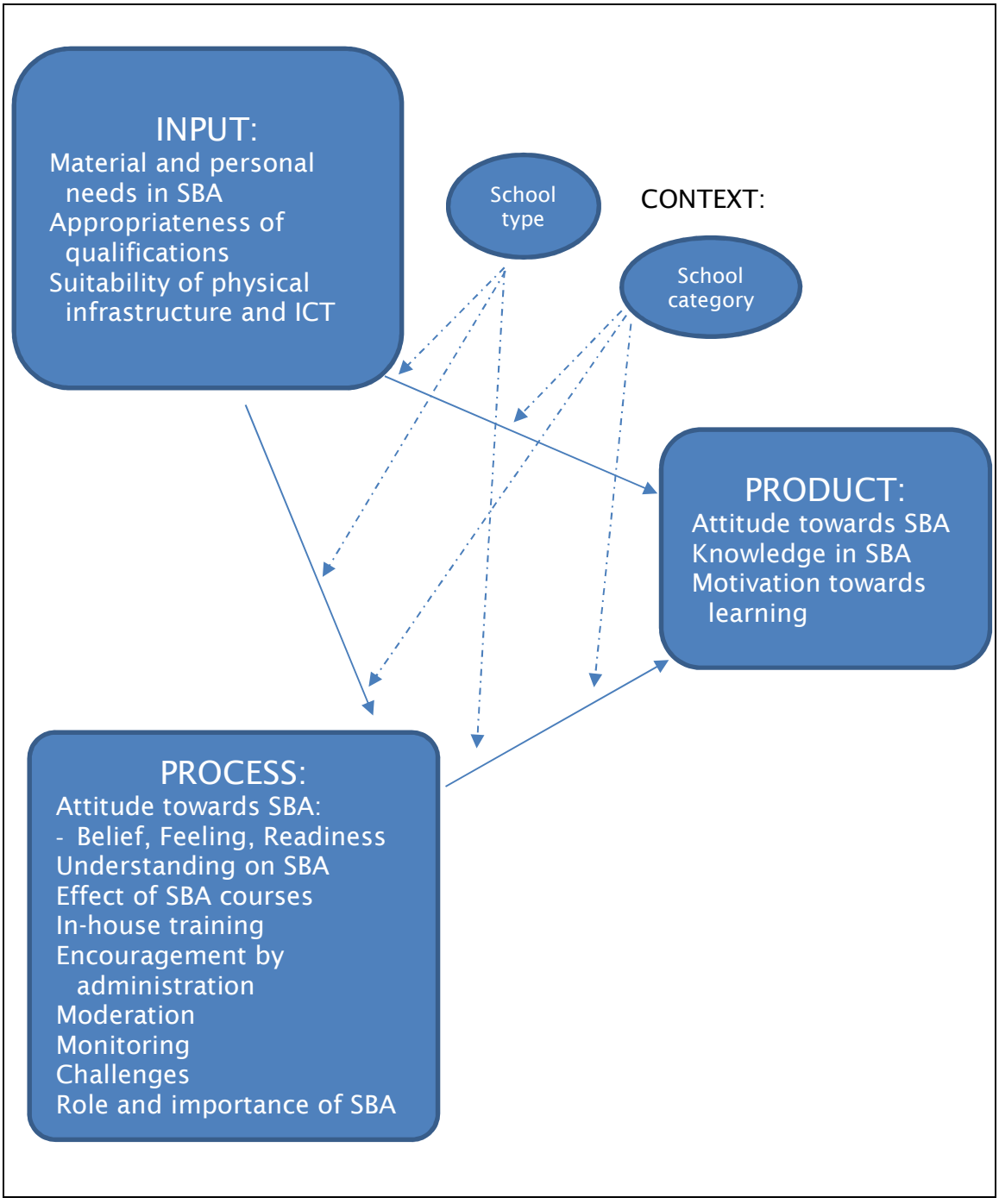


Figure 1.2: The Conceptual Framework for the SBA System Evaluation

Table 1.4: The Context, Input, Process and Product Evaluation

CONTEXT EVALUATION	INPUT EVALUATION	PROCESS EVALUATION	PRODUCT EVALUATION
School type; School category	Material and personal needs in SBA; Appropriateness of personnel's qualifications; Suitability of physical infrastructure and ICT	Teachers' attitude towards SBA: Belief, Feeling, Readiness; Teachers' understanding on SBA; Effect of SBA courses; In-house training; Encouragement by administration; Moderation process; Monitoring process; Challenges; School Improvement: Role of SBA, Importance of SBA.	Students' attitude towards SBA; Students' knowledge in SBA; Motivational source towards learning

Context evaluation focuses on the environment where the changes happen. In this study, as shown in Table 1.4, context involves a number of factors distinguished by school context in Malaysia which involves schools from different type (urban-rural) or different category (primary-secondary). Input evaluation consists of three constructs which are material and personal needs in SBA, appropriateness in personnel's qualifications and suitability of physical infrastructure and ICT. Since process evaluation focuses on providing information during the programme implementation and a complete description of the actual programme activities (Stufflebeam, 1971a), it involves the significant and relevant processes in SBA to achieve the programme objectives. Process evaluation consists of twelve constructs which form the teachers' attitude towards SBA system and includes belief, feeling and readiness, teachers' understanding of SBA system, effect of SBA courses in improving teachers' skill, in-house training, the encouragement by the administration, moderation process, monitoring process, challenges, role of SBA and importance of SBA in the school improvement process. Lastly is the product

evaluation which is looking at the aspects of programme impact concerning students' attitudes towards SBA, students' knowledge in SBA and students' motivational source towards learning.

In conclusion, the question of 'why use a conceptual framework?' considered the importance of the framework to this doctoral research. The figure representing the conceptual framework of this study shows the coherence between the theoretical perspectives of the study, its design and methodology and lastly is the conceptual significance of the evidence (Leshem and Trafford, 2007). In short, it is known as a 'bridge' connecting research paradigm to the investigations on the phenomena or a 'device' that makes sense of the data.

1.4 Making Contribution to Knowledge

SBA, if implemented effectively, is able to promote students' learning (Cheng *et al*., 2011). SBA is designed to improve the teaching and learning process and make an improvement to schools as a whole. Yet, many people are now starting to get worried and to question whether could SBA make a change in an education system, especially when it is newly implemented (Utusan Online, 2014b). Teachers regularly ask, 'Does the change in the assessment system really make a change in our education system?' or 'Can we continue SBA with all these unsolved challenges?'(Aishah, 2014). Parents are asking, 'With no more public examination, can my kids perform?' and students are asking, 'Can we be motivated towards learning if there is no more public examination as in previous days?' This study intends to answer most of the questions and attempt to provide solutions to the varying concern.

Some of the current research develops and validates an evaluation scale using Stufflebeam's CIPP Model but these have been conducted in the contexts of the English curriculum evaluation (Hakan and Seval, 2011), the quality of undergraduate hospitality, tourism and leisure programme evaluation (Horng *et al*., 2009) or the nano-technology curriculum evaluation (Tseng *et al.,* 2010). This doctoral study differs by attempting to determine the validity and reliability of an evaluation scale developed by the researcher based on the principles of Stufflebeam's CIPP Model within the context of the SBA system. CIPP has not been used in this context before in this way; therefore SBA

approaches have not been evaluated using a vigorous and comprehensive framework in the past. So, the scale might be of interest by many parties. If successful, the research could be an efficient instrument to evaluate the implementation of SBA at any preferable dimensions according to different interests.

Most evaluation processes look only at some dimensions which do not give a fully rounded indication of the effectiveness of the system implemented while others were not conducted in a systematic and focused manner. Previous research on the evaluation of the SBA system from the process dimension has focused on: looking at teachers' attitude towards SBA (Sidek and M. Salleh, 2010; Majid, 2011; Adediwura, 2012; Salmiah, 2013), teachers' leadership (Boon and Shaharuddin, 2011), teachers' knowledge and best practises in SBA (Juliana, 2008; Chan and Gurnam, 2012) or challenges faced by the teachers (Lo, 2006; Malakolunthu and Hoon, 2010; Lukman and Uwadiegwu, 2012). Research has also been conducted using the product dimension which identifies the benefits in the SBA (Mansor *et al.,* 2013). Some research looks into both, process and product dimensions as in Annie (2011), which determines students' perception of and reaction to feedback. Some current research relates to process and product dimensions of evaluation (Wiliam *et al*., 2004; Christina, 2009; Cheng *et al.,* 2011) while Koh and Velayutham (2009) determined the correlations between input and product dimension. But little attention has been paid to input and product dimensions alone. Furthermore, to date, studies that combine all the four dimensions of evaluation are non-existent.

This study examines the structural equation modelling technique to assess measurement properties and tests the theoretical relationship between the four stated dimensions of evaluation in the context of the SBA system implementation. Furthermore, research has also been conducted evaluating all the four dimensions of the CIPP Model but does not *relate* the dimensions; for example, Chen (2009) used the CIPP Model in the context of the English training courses evaluation, while other researchers considered the evaluation of character education programme (King, 2008) and the evaluation of an English language teaching programme (Tunc, 2010).

The need to evaluate the SBA system is becoming increasingly important considering the system's impact on students' attitude towards SBA, their knowledge of SBA and their motivational source towards learning. This study also has a potential value for teachers to develop themselves and to improve schools as a whole. There is an urgent need to understand more about the challenges that teachers are facing as there are pressures being imposed towards the Malaysian government on this issue, especially from the teachers, parents and students (Aishah, 2014). Information gained later is also important for schools, education departments or ministry to plan and implement effective strategies related to SBA. The Malaysian government would benefit greatly from this study as it determines the effectiveness of the system at the early stage of the implementation. So, by conducting empirical research on the evaluation process of the programme will benefit a variety of groups in the field of education.

1.5 Purpose of the Study

The main purpose of this study is to evaluate the SBA system implementation in Malaysian schools using the CIPP Model in preparing a summative report of the programme so that results of each dimension of evaluation and the interrelationships between dimensions of evaluation would lead to further improvement later. To achieve the purpose, the study aims to validate the variables to ensure that they are in fact measuring what they intend to measure. The study then considers the hypothesised interrelationships among context, input, process and product dimensions. In addition, this study will determine the moderating effect of the moderating variables from context dimension such as school type and school category. Based on these purposes, the study's research objectives are as follows:

RO1. To identify the teachers' perception of the input, process and product dimensions.
RO2. To empirically test the probability of a three-factor model for input dimension.
RO3. To empirically test the probability of a 12-factor model for process dimension.

RO4. To empirically test the probability of a three-factor model for product dimension.

RO5. To identify if process influences product dimension.

RO6. To identify if input influences process dimension.

RO7. To identify if input influences product dimension.

RO8. To identify the moderating effect of type and category of schools on the relationship between evaluation dimensions in SBA.

1.6 Research Questions

A number of research questions emerge and are further refined in the course of developing the research design. The study is guided by the following research questions;

RQ1: What are the teachers' perceptions on the input dimension of the SBA system implementation?

RQ2: What are the teachers' perceptions on the process dimension of the SBA system implementation?

RQ3: What are the teachers' perceptions on the product dimension of the SBA system implementation?

RQ4. Is the measurement scale for input dimension construct valid?

- RQ4.1: Can the evaluation in input dimension be explained by the following three factors: material, personnel and infrastructure?
- RQ4.2: Does each indicator have a non-zero loading on the hypothesised or targeted factor?
- RQ4.3: Does each indicator have a zero loading on the other non-targeted factor?
- RQ4.4: Are the error terms uncorrelated?

RQ5. Is the measurement scale for process dimension construct valid?

- RQ5.1: Can the evaluation in process dimension be explained by the following twelve factors?
- RQ5.2: Does each indicator have a non-zero loading on the hypothesised or targeted factor?
- RQ5.3: Does each indicator have a zero loading on the other non-targeted factor?
- RQ5.4: Are the error terms uncorrelated?

RQ6. Is the measurement scale for product dimension construct valid?

RQ6.1: Can the evaluation in product dimension be explained by the following three factors?

RQ6.2: Does each indicator have a nonzero loading on the hypothesised or targeted factor?

RQ6.3: Does each indicator have a zero loading on the other non-targeted factor?

RQ6.4: Are the error terms uncorrelated?

RQ7. Does process influence product dimension?

RQ8. Does input influence process dimension?

RQ9. Does input influence product dimension?

RQ10. Does school type (urban-rural) moderate the relationship between evaluation dimensions in SBA implementation?

RQ11. Does school category (primary-secondary) moderate the relationship between evaluation dimensions in SBA implementation?

1.7 Research Hypotheses

In line with the research questions and research objectives, this study tests a number of hypotheses. RQ1, RQ2 and RQ3 are answered using descriptive statistics so they do not need hypotheses. The hypotheses for RQ4, RQ5 and RQ6 are tested using confirmatory factor analysis in order to check the validity of the hypothesised relationships between a construct and its indicators. Next, RQ7, RQ8 and RQ9 are used to determine the relationships among input, process and product dimensions and RQ10 and RQ11 are looking at the moderating effect. The research hypotheses are as follows:

RQ4. Is the measurement scale for input dimension construct valid?

H_1: Evaluation in input dimension is explained by three factors.

H_2: Each indicator has a nonzero loading on the targeted factor.

H_3: Each indicator has a zero loading on the non-targeted factors.

H_4: The error terms are uncorrelated.

RQ5. Is the measurement scale for process dimension construct valid?

H_5: Evaluation in input dimension is explained by eleven factors.

H_6: Each indicator has a nonzero loading on the targeted factor.

H_7: Each indicator has a zero loading on the non-targeted factors.

H_8: The error terms are uncorrelated.

RQ6. Is the measurement scale for product dimension construct valid?

H_9: Evaluation in input dimension is explained by three factors.

H_{10}: Each indicator has a nonzero loading on the targeted factor.

H_{11}: Each indicator has a zero loading on the non-targeted factors.

H_{12}: The error terms are uncorrelated.

RQ7. Does process dimension influence product dimension?

H_{13}: Process dimension is positively associated with product dimension.

RQ8. Does input dimension influence process dimension?

H_{14}: Input dimension is positively associated with process dimension.

RQ9. Does input dimension influence product dimension?

H_{15}: Input dimension is positively associated with product dimension.

RQ10. Does school type moderate the relationship between dimensions in SBA implementation?

H_{16}: School type moderates the relationship between process and product dimension

H_{17}: School type moderates the relationship between input and process dimension

H_{18}: School type moderates the relationship between input and product dimension

RQ11. Does school category moderate the relationship between dimensions in SBA implementation?

H_{19}: School category moderates the relationship between process and product dimension

H_{20}: School category moderates the relationship between input and process dimension

H_{21}: School category moderates the relationship between input and product dimension

1.8 Summary

In this chapter, I have described the basic elements of the study. It started with the background to the study and the statement of the problem. The conceptual framework used in this study is taken from the CIPP Model by Stufflebeam. The chapter has explained the concepts and terms in evaluating the system. Then, it followed with how the study contributes to the body of knowledge, the study

purpose and the research questions. It is crucial when it comes to evaluation to determine, not only the context of the study but also the input, process and product dimensions. As such, this study provides opportunities for learners, educationists and other stakeholders to take into consideration of all factors from each dimension when implementing SBA in schools. Therefore, this study proposed a research framework which offers a platform for evaluating programmes in various dimensions. It is the belief of the researcher that the findings will result in an effective and meaningful way in managing and making decisions on the implementation of SBA. This is important because until now, no models have been developed to relate the dimensions of evaluation in the context of SBA implementation in the Asian countries, especially in Malaysia.

Chapter 2: The Assessment System in the Malaysian Educational System

2.1 Introduction

This chapter describes the assessment system within the Malaysian education system. Previously, there were two types of assessment in Malaysia: SBA and public examination. SBA is practised in both the primary and the secondary education levels but more in an informal and unstructured basis. There are five major public or national examinations: the Primary School Evaluation Test (UPSR), the Lower Secondary Examination (PMR), the Malaysian Certificate of Education (SPM), the Malaysian Higher School Certificate Examination (STPM) and the Malaysian Higher Islamic Religious Certificate (STAM) (Ong, 2010). The first three public examinations are managed by the Malaysian Examination Syndicate (MES) whereas the last two are managed by the Malaysian Examination Council, with both being under the MOE (Anton and Jordan, 2002). The MES and the Malaysian Examination Council are responsible in preparing tests, setting examination rules and regulations, revising syllabi, coordinating registration of students, marking and processing results. Additionally, they are also responsible in formulating assessment and measurement in education policies based on the NPE and the curricular objectives (Lim and Zhao, 2005).

The first public examination, the Primary School Evaluation Test (UPSR) is a formal centralised examination taken by all students at the end of Year 6 at the age of 12 (Anton and Jordan, 2002). It serves as a summative assessment to diagnose students' strengths and weaknesses at the end of the primary school level and does not affect their entry into the secondary school level. The second public examination is the Lower Secondary Examination (PMR). It is a formal centralised examination taken by all students at the end of the lower secondary level at Year 9 at the age of fifteen. All students have to take all the eight subjects offered (Ong, 2010). The difference with this examination is that it includes marks from the SBA scores from students' portfolios for history, geography, integrated living skills and practical science. PMR examination serves as a summative assessment to channel students either into arts streams

or science streams in their upper secondary level of schooling. Students also undergo continual summative assessments including weekly, monthly or term class tests, subjects and regular class assignments set by the respective teachers of the subjects. The third public examination is the Malaysian Certificate of Education (SPM), taken by students at the age of seventeen at the end of their final year in the upper secondary school level (Anton and Jordan, 2002). SPM consists of six core subjects of Malay language, English, Mathematics, Science, History and Islamic Studies or Moral Education. After completing the SPM, students have few options to continue their study either to enter post-education or Form 6 (Year 12) or studying for a diploma at various institutions of higher learning. The first three public examinations are shown in Table 2.1 (Ong, 2010).

At the end of Year 12, there is the Malaysian Higher School Certificate Examination (STPM) or the Malaysian Higher Islamic Religious Certificate (STAM). STPM is a pre-university examination taken by students at the end of a two-year study programme (GTI Media, 2013). A maximum of five subjects could be chosen out of twenty-three subjects offered to prepare students to enter university. STAM is an examination similar to STPM in post-secondary level after a two-year study programme in secondary religious schools with religious education as its requirement to enter the programme (Communication of Learning and UNESCO, 2004). These are to prepare students into higher education or into the employment market.

Table 2.1: Public examination in the Malaysian education system

PUBLIC EXAMINATIONS	SUBJECTS	FORM OF ASSESSMENT	REPORTED RESULTS
UPSR	Malay Language, Malay writing, English Language, Mathematics, Science	Centralised examination: Multiple-choice and open-ended items	In the form of grade A, B, C, D or F
SBA	Science Practical Assessment		Level of skills 1, 2, 3 or 4
PMR	Malay Language,	Centralised	Grade A, B, C,

Centralised examination	English Language, History, Geography, Islamic Education or Moral Education, Mathematics, Science, Integrated Living Skills	examination	D or E
SBA	Geography, History, Integrated Living Skills, Practical Science	Course work	Level of skills, 1, 2, 3, 4 or 5
SPM Core subject Elective subject	Malay Language, English Language, Mathematics, Science, History, Islamic Studies 2 or more subjects	Centralised examination: multiple-choice items, short-answer constructed-response items, essay items	Grade 1A, 2A, 3B, 4B, 5C, 6C, 7D, 8E or 9G
SBA	Malay and English language	Oral exam	
	Moral Education	Project work	
	Science, Physics, Chemistry, Biology	Science practical test	

2.2 The Change in the Assessment System

The New Primary School Curriculum (KBSR), which was introduced in 1983, has gone through a transformation process and is now replaced by the Standards-based Primary School Curriculum (KSSR). Since 2011, KSSR has been implemented in the Malaysian education system. The government urges all schools to implement KSSR for the Year 1 students as stated in the circular letter KP (BPSH-SPDK) 201/005/01/Jld.3 (5) dated 14th of October 2010.

The KSSR as a new curriculum involves changes of eight basic aspects such as content structure, pedagogy, time allocation, assessment method, curriculum

materials, organisation, form and management (Kementerian Pelajaran Malaysia, 2011b). All of these are based on the NPE and constitute the main focus in the fifth chapter in the Tenth Malaysian Plan (2011-2015). KSSR has been introduced to ensure that students are well-equipped with relevant knowledge, skills and values of the twenty-first century. The major difference between KBSR and KSSR is shown in Table 2.2 (PPD Hulu Selangor, 2011).

Table 2.2: Difference between KBSR and KSSR

KBSR	**KSSR**
Curriculum design:	
Based on 3 attributes: Communication; human and environment; and the individual development	Based on 6 attributes: Communication; Spiritual, attitude and values; Humanities; Personal competencies; Science and technology; and Physical and aesthetics development
Curriculum materials:	
Syllabus	Curriculum Standard Document
Curriculum Design:	
Linear	Modular
Curriculum Organization:	
Level 1 (Year 1, 2 and 3): Core, compulsory and additional subjects Level 2 (Year 4, 5 and 6): Core, compulsory and additional subjects	Level 1 (Year 1, 2 and 3): Basic core modules, theme core modules and elective modules Level 2 (Year 4, 5 and 6): Core and Elective subjects
Thinking Skills:	
Critical and Creative Thinking Skills	Innovation and Creative Element, Entrepreneurship, ICT and Communication with explicit
Focus:	
3Ms (reading, writing and calculating)	4Ms (reading, writing, calculating and reasoning)

KSSR is enacted based on a Standard Curriculum consisting of content standards and learning standards. Standard Curriculum is developed by the

Curriculum Development Section from the MOE (Lembaga Peperiksaan Malaysia, 2011b). Content standards are specific statements on what a student should know and be able to do within a specific period of schooling, including aspects such as knowledge, skills and values. Learning standards are sets of criteria or indicators of education quality and achievements which have been set and can be measured for each content standards (Kementerian Pelajaran Malaysia, 2011a). Standard Curriculum is applied to ensure that each and every student achieves the standard that has been set, being certain with knowledge, skills and values that should be measured, to determine the improvement strategy as in AfL, to minimise student drop-out and to implement the new SBA assessment system. The examples of both standards are shown in Table 2.3.

Table 2.3: Content Standard and Learning Standard of KSSR

STANDARD CURRICULUM	
Content Standards	**Learning Standards**
Bahasa Malaysia: To listen, understand and give a correct response to the instruction or question.	Bahasa Malaysia: To listen, understand and give a correct response orally to various types of questions.
English Language: Able to express personal responses to literary texts.	English Language: Able to respond to book covers, pictures in books and characters with guidance.
Mathematics: To name and determine the value of number.	Mathematics: To name the numbers up to 100: a. To count objects in a group b. To name a number to a group of objects

The new assessment system in KSSR is the National Education Assessment System (NEAS). By definition, NEAS is an educational assessment system conducted by the MES together with the schools with the main aim to introduce a set of indicators in assessing students' potential and their readiness to learn rather than testing solely on their achievement (Lembaga Peperiksaan, 2010).

2.3 National Education Assessment System (NEAS)

The objectives of NEAS are;

i) To reduce the focus on public examinations
ii) To strengthen SBA
iii) To improve students' learning
iv) To create a holistic and everlasting assessment
v) To develop a better human capital (Lembaga Peperiksaan Malaysia, 2009)

NEAS consists of Central Examination (CE) and SBA. SBA consists of four components of assessment, divided into the academic components, Central Assessment (CA) and the School Assessment (SA) and the non-academic components, Physical Activity, Sports and Co-Curriculum Assessment (PASCA) and Psychometric Assessment (Psi), as shown in Figure 2.1 (Ong, 2010).

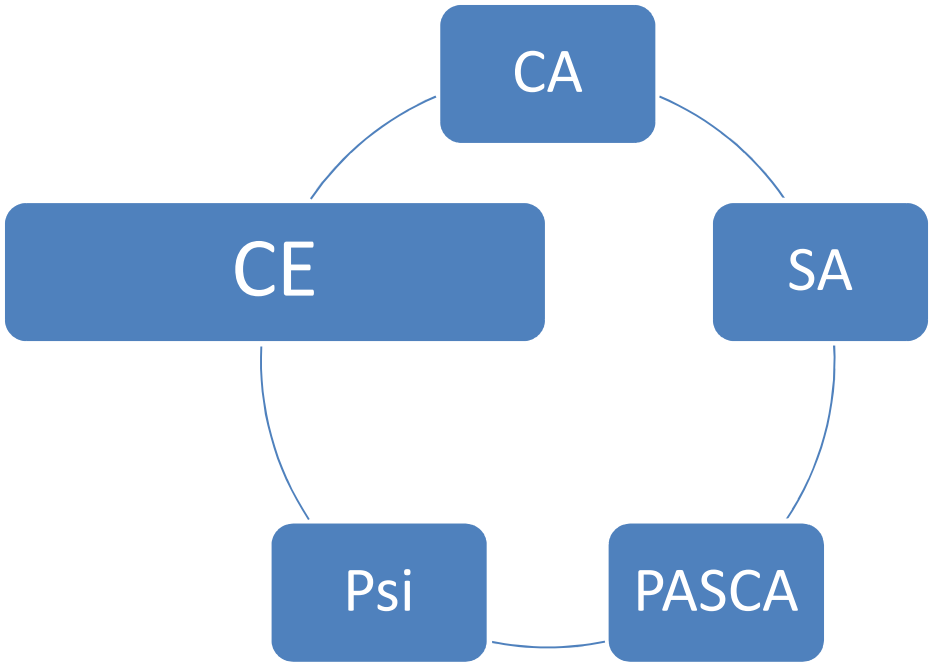


Figure 2.1: National Education Assessment System

2.3.1 Central examination (CE)

CE is an examination to assess students achievement conducted fully by the MES. This is the same as the existing public examination (Ong, 2010). Reports of CE are presented using certificate by the MES together with the SA and the CA (Lembaga Peperiksaan Malaysia, 2011a). There will still be a CE for Year 6 students for six core subjects of Malay Language, English Language, Science, Mathematics, Chinese Language and Tamil Language but only a small percentage contributes to the final score of students' performance in the primary school level (Norzila, 2013).

2.3.2 School-based Assessment (SBA)

In the Malaysian context, SBA, the focus for this study, is a form of assessment that is conducted in school, and is planned, administered, scored and reported in a mannered way based on the procedures from the MES (Lembaga Peperiksaan Malaysia, 2011a). The SBA is conducted by the subject teachers continuously during the teaching and learning process or at the end of a stage or task. It is a holistic assessment assessing cognitive (intellect), affective (emotional and spiritual) and psychomotor (physical) aspects in line with NPE and the Standards-based School Curriculum. It assesses the process and product in formative and summative way by practicing AfL and AoL concepts. The main objectives of SBA are:

i) To get the overall picture of an individual's potential
ii) To monitor individual's development and help to increase their potential
iii) To make meaningful reporting on individual learning

(Lembaga Peperiksaan, 2010)

In particular, it is to enhance the meaningfulness of assessment by focusing more on students' learning development rather than grade, to reduce over-reliance on grade, to empower schools and teachers with quality assessment especially from SA and finally, to make sure that students' performance are comparable to the world standards in terms of their knowledge, skills and competence (Che Noraini *et al.,* 2013).

The implementation of SBA is based on the 1996 Education Act (550 Act): Section 67 – Students' Assessment, Section 68 – Examination and Section 69 – Prohibitions on Examination Management and 1997 Education Regulations (Assessment and Examination), Division II, Case 3 which states that the MES is responsible for advising, monitoring and analysing the SBA system implementation and the management of all the examinations, and also to run the examination, monitor it and enforce the examinations regulations and guidelines (Ministry of Education, 2011). Furthermore, its implementation is also based on the Performance Standards Document and the SBA Management System Manual. The weighted of all the components of SBA is shown in Table 2.4 except for the psychometric assessment which will be reported separately (Lembaga Peperiksaan Malaysia, 2009).

Table 2.4: SBA and public examination from preschool to upper secondary school

<table>
<tr><th rowspan="2">Pre-school</th><th colspan="2">Primary School Level 1</th><th colspan="5">Primary School Level 2</th><th colspan="3">Lower Secondary School</th><th colspan="2">Upper Secondary School</th></tr>
<tr><td>Pre</td><td>Y1</td><td>Y2</td><td>Y3</td><td>Y4</td><td>Y5</td><td>Y6</td><td>Y7</td><td>Y8</td><td>Y9</td><td>Y10</td><td>Y11</td></tr>
<tr><td colspan="3" rowspan="2">SA-70%
PASCA-30%</td><td colspan="3" rowspan="2">SA-50%
PASCA-20%</td><td>SA-40%
PASCA-20%</td><td colspan="2">SA-35%
PASCA-15%</td><td>SA-35%
PASCA-15%</td><td colspan="2">SA-20%
PASCA-10%</td></tr>
<tr><td>CA-30%</td><td>CA-20%
CE-20%</td><td colspan="2">CA-50%</td><td>CA-20%
CE-30%</td><td colspan="2">CA-30%
CE-40%</td></tr>
<tr><td>Psi</td><td>Psi</td><td>Psi</td><td>Psi</td><td>Psi</td><td>Psi</td><td>Psi</td><td>Psi</td><td>Psi</td><td>Psi</td><td>Psi</td><td>Psi</td></tr>
</table>

Y1 = Year 1
SA = school assessment
CA = central assessment
PASCA = physical activity, sports and co-curriculum assessment
Psi = psychometric assessment
CE = central examination (public examination)

The difference between SBA and the previous assessment is that SBA is more holistic in assessing students and focuses on many aspects such as physical, emotional, spiritual and humanistic aspects, whereas the existing assessment is focusing on academic achievement only (Ministry of Education, 2011). Furthermore, SBA is practising AfL and AoL but the previous assessment is more on AoL. AfL and AoL will be detailed in Chapter 3. By using Standards-

referenced Assessment in SBA, students' performance are not compared to each other in their classroom like those practiced previously, but rather they are assessed on their growth in learning based on the standard statements.

Until the date the researcher went to the schools to collect data in February and March 2013, the schools were practising SA only and not the other three components of SBA. The teachers informed the researcher that the other three components will be implemented in 2014. And the most senior group of students practising SBA at that time was the group in Year 3 for the primary school and Year 8 for the secondary school. Hence, the empirical work for this study was carried out during the third year of the SBA system implementation, which was in 2013.

2.3.3 Central Assessment (CA)

CA is a summative assessment used to assess academic achievement of students especially their knowledge, skills and values (Lembaga Peperiksaan, 2014). In 2014, PT3 (Year 9 assessment), an example of a CA was conducted. The MES prepares the items, test specification schedule, instrument format, instrument installation guide, item scoring guide, administration regulation and reporting format. Schools are responsible for preparing instruments, administer them, marking answer scripts and produce reports. The installation of instruments is based on items, test specification schedule, instrument format, grading schedule and instrument installation guide. Schools then prepare the schedule and administer a one-off test in the range of time and regulations. Next, the marking of completed scripts undertaken by the subject teachers, which is then followed by moderation process. Scoring and grading process are based on grading guide and schedule. Grades are then be verified by the head teacher before they are sent to MES. Lastly, schools produce reports based on the reporting format.

In general, CA 'tasks' could be long term tasks, which are carried out within three to four weeks or short term tasks which are carried out within two weeks (MOE, 2013). Tasks could be done in groups with each student gets an individual score. It must be implemented as *'one-off'* and reassessment is not allowed (Lembaga Peperiksaan, 2012). CA reports will be added to the SA at

the end of the year (Lembaga Peperiksaan Malaysia, 2011a). Other than serving the basic characteristics of SBA, CA also presents an integrated form of assessment as it assesses various constructs and contexts. It is reliable as the assignments refer to the actual condition and the scoring rubrics depend on the specific assignments (Lembaga Peperiksaan Malaysia, 2011a). The instruments are grouped into three main types – oral test, written test and using various instruments such as projects, laboratory works, demonstration, case study, field study or making presentation (Lembaga Peperiksaan, 2014). One of the difference between summative assessment in SA and CA is that, for SA, the summative reportings from monthly tests, mid-year or final year examination show the students' achievement in the current year. However, for PT3 (Year 9 assessment) for example, it is to assess students' achievement with what they have learnt over three years' time - Years 7, 8 and 9.

2.3.4 School Assessment (SA)

SA is an assessment that is planned, built, governed, checked, recorded and reported by the teacher in the school based on the Performance Standards. It consists of two forms of assessment: 'formative' and 'summative', which act as an AfL and AoL (Lembaga Peperiksaan Malaysia, 2011a). Formative assessment is implemented along with the teaching and learning processes such as in worksheets, observations, quizzes, checklists, reports and homework. Summative assessment is an assessment conducted at the end of a teaching unit, month, semester or year (Ministry Of Education, 2011).

The most significant characteristic of SA is that it has to refer to the Document of Performance Standards, which is constructed based on Standard-based Curriculum. The Document of Performance Standards comes from the Standard-referenced Assessment, the idea being of a fair assessment towards students as they are assessed according to their capacity, ability, talent, skills and potentials referring to standard statement without comparing to others (Lembaga Peperiksaan Malaysia, 2011a). Figure 2.2 shows the relationship between Curriculum Standard Document and Performance Standard Document (Lembaga Peperiksaan, 2012).

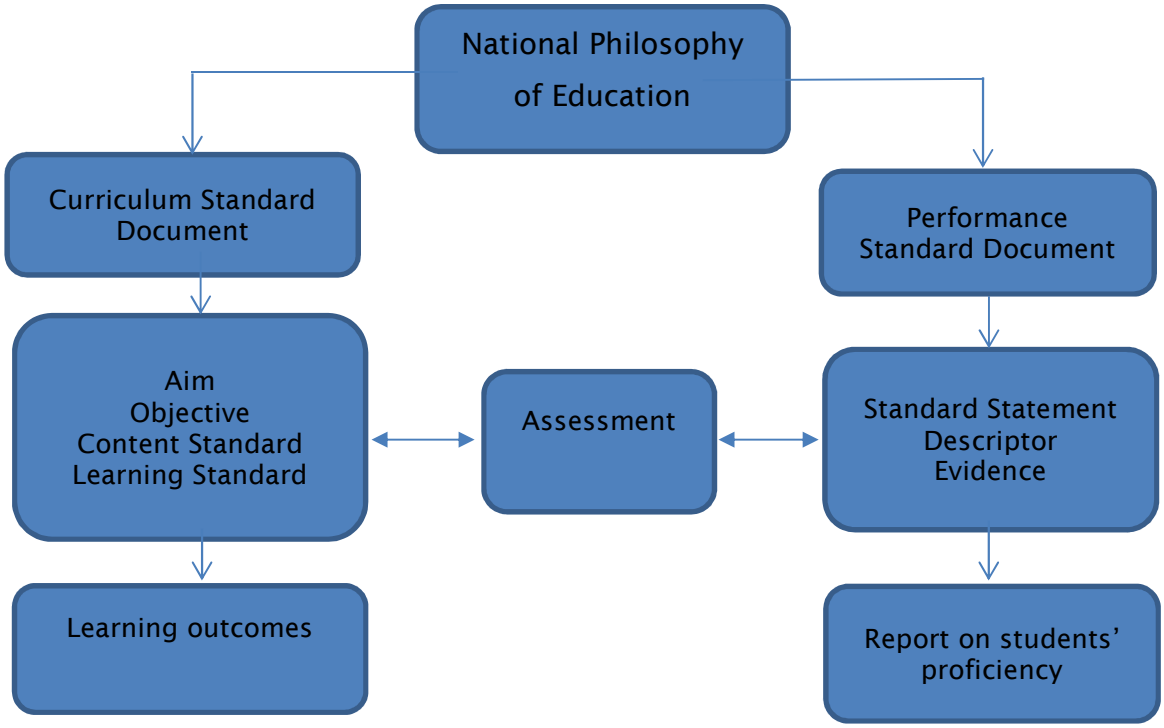


Figure 2.2: The Relationship between Curriculum Document and PSD

The Performance Standard Document (PSD) includes bands, statements of standards, descriptors and evidence with guidelines following curriculum to assess and report students' achievement at different stages of learning (MOE, 2013) as shown in Tables 2.5, 2.6, 2.7 and 2.8. The PSD is a statement on an individual's achievement in a certain field in a given range of time according to a particular benchmark (Lembaga Peperiksaan Malaysia, 2011a). The PSD serves a very important concept for teachers, students, parents and stakeholders. The PSD allows teachers and students to know the aims to achieve, what is being assessed and the expected quality in assessment that students have to possess. Furthermore, the PSD also allows parents to realise what has been achieved by their children and allows stakeholders to know whether the aim of the curriculum has been achieved and the quality of students that have gone through our education system, whilst on a national level acting as Malaysia's standard benchmark to compare with the other countries' education standards (Lembaga Peperiksaan Malaysia, 2011a). Band 1, 2 and 3 is the 'basic', Band 4 is 'excellent', Band 5 is 'brilliant' and Band 6 is 'distinction' as shown in Table 2.5. Table 2.6 shows an example of bands and the statements of standard of English Language Year 1. In developing the PSD,

certain tips are given by the District Education Office where teachers should ask themselves questions such as 'what are the real objectives that we try to achieve?', 'what are the students' level of work that we expect?', 'what standards are we using?' or 'what are the suitable feedbacks that will be given to the students for improvement based on students' product which has been assessed?' (PPD Barat Daya, 2010).

Turning to the teaching processes using SA in a classroom, it is worth considering what teachers do instead of simply teaching students and then waiting for public examination, as was previously the case. All teachers are expected to use the Curriculum Standard Document for all the elements that they are supposed to deliver during the teaching and learning process and to use the PSD to assess students (Lembaga Peperiksaan, 2012). Teachers choose topics and plan the curriculum referring to the Curriculum Standard Document and deliver the curriculum to students using various teaching strategies such as using lecture, discussion, questioning or using computer, video or textbooks depending on the teachers' creativity. Teachers are then required to assess students' proficiency levels using various strategies and instruments based on the PSD. They are given autonomy in implementing formative or summative assessments in their classrooms. Based on the evidence given by students, students' performance is assessed. If the students do not master the predetermined standard, teachers have to guide them until they achieve. If they are able to answer questions from Band 1, they will be guided to answer those from Band 2 and henceforth until Band 6, if possible. Teachers then record and report students' performance and proceed to the next topic.

Table 2.5: Framework of the Performance Standard

BAND	**STANDARD**
6	Know, Understand and Apply Knowledge Obtained Using Exemplary Manner
5	Know, Understand and Apply Knowledge Obtained Using Admirable Manner
4	Know, Understand and Apply Knowledge Obtained Using Systematic Manner
3	Know, Understand and Apply Knowledge Obtained
2	Know and Understand
1	Know

Table 2.6: Framework of the Performance Standard of English Language Year 1

BAND	STANDARD OF ENGLISH LANGUAGE YEAR 1
6	Appreciate literary works by performing and presenting ideas using exemplary manners
5	Demonstrate well the ability to apply knowledge of listening, speaking, reading and writing for various purposes using admirable manners
4	Apply knowledge obtained through listening, speaking, reading and writing in various situations using systematic manners
3	Know, Understand and Apply knowledge obtained through listening, speaking, reading and writing
2	Know and Understand words, phrases and sentences heard, spoken, read or written
1	Know basic skills in listening, speaking, reading and writing

Table 2.7: Framework of the PSD

Band	Statements of Standard	Descriptor	Evidence
1 2 3 4 5 6	Generic or general information on learning level (Refer Objectives of the Subject)	WHAT students KNOW and ABLE TO DO (Refer Curriculum Standards of a particular subject for a particular year)	HOW students present what they KNOW and ABLE TO DO

Table 2.8: PSD of Mathematics Year 1 (Band 1 only)

BAND	STATEMENTS OF STANDARD	DESCRIPTOR	EVIDENCE
1	Know the basic mathematics knowledge	B1D1 Quantity and measurement of objects in an intuitive comparison	B1D1E1 To show or mention quantity and measurement of objects in a comparison form: a) More or less b) Long or short c) Heavy or light d) Same amount or not
		B1D2 Terminology and vocabulary	B1D2E1 To name: a) time in a day b) days in a week c) months in a year d) time in term of hours

			B1D2E2 Recognise the vocabulary for addition and subtraction operation B1D2E3 Recognise the vocabulary for non-standard measurement unit
		B1D3 Whole number, 3-dimension (3D) and 2-dimension (2D)	B1D3E1 To name the whole number till 100
			B1D3E2 To name 3D and 2D shape

2.3.5 Physical Activities, Sports and Co-Curricular Assessment (PASCA)

PASCA is implemented at school level to assess physical endurance and body mass index and also students' participation, involvement and contribution in sports, co-curriculum and extra-curriculum activities (Lembaga Peperiksaan, 2012). This assessment is intended to make the assessment more holistic and to complement the 'One Student One Sport' concept and the concept on students' development as advocated in the NPE. PASCA documentation gives a whole picture of the level of students' development in knowledge proficiency and core values appreciation fostered from education processes like innovation, competency, responsibility, good moral values, patriotism or entrepreneurship (Lembaga Peperiksaan Malaysia, 2011a). The instruments and assessment guides are prepared by the MES.

2.3.6 Psychometric Assessment (Psi)

The Psi is implemented at school or centre level to measure psychology traits including students' aptitude and personality to help students to improve their learning (Lembaga Peperiksaan Malaysia, 2011a). The Psi instrument was developed by psychometric experts and psychologists and is administered manually or computerised by the teachers or the students themselves. The examination board prepares a comprehensive and user-friendly examination

manual (Lembaga Peperiksaan Malaysia, 2011a) and it is not based on the curriculum or teaching and learning.

There are three different types of psychometric assessments: the general aptitude test, the specific aptitude test and the personality inventory (Ministry Of Education, 2011). The general aptitude test measures thinking skills, problem solving skills and interest and inclination, the specific aptitude test measures nine constructs which are music, creativity, interpersonal, intrapersonal, kinaesthetic, naturalist, verbal, visual and mathematics logic whereas personality inventory measures ten constructs which are communication, group work, leadership, spiritual stability, emotional functioning, culture adaptability, accountability, entrepreneurship and ethics (Lembaga Peperiksaan Malaysia, 2011a). Table 2.9 shows all the four components with their characteristics (MOE, 2012).

Table 2.9: Four components of SBA

Component	Implementation Schedule	Instrument	Administration	Scoring	Report
SA	All through the year	Prepared by subject teacher based on PSD	School SBA committee Subject teacher	Subject teacher assesses based on PSD	Descriptive report from SPPBS
CA	Schedule from MES	Prepared by MES	School SBA committee Subject teacher	Subject teacher assesses based on rubric	Descriptive report from SPPBS
PASCA	All through the year	Prepared by MES	School SBA committee Teacher	Teacher assesses based on criteria	Descriptive report from PASCA
Psi	As required	Prepared by MES	School SBA committee Class teacher Counselor	MES	Descriptive report from MES

2.4 SBA Training Plan

The latest training plan for SBA is shown in Table 2.10 (Ministry Of Education, 2011). The training refers to the Cascade Training Models, a formal training programme conducted for in-service teacher training to 'cascade' information from national level, to state level and followed by the school level (Chan *et al.,* 2006). The training begins with national group of trainers who in turn train another group of trainers at the state level who subsequently train another group from the district level. Normally, training at the national level takes a week, at the state level takes three days and at the school level, it is just one day. Trainings involve talks, slide presentation and providing booklets to the teachers with no hands-on training involved. The advantage of this model is that more teachers could be trained with less time taken. Nevertheless, the disadvantage is that it is conducted off site so when teachers go back to their own schools they cannot practise it accordingly. Chan and Gurnam (2012) also believe that the models are not effective because they are not hands-on, so teachers cannot practise it during the training itself and the exposure for the training varies depending on the levels of administration - national, state or school level. In the year 2010, training in KSSR and SBA was conducted to 40, 000 Year 1 teachers since in the year 2011 SBA has been implemented to Year 1 students. The same goes to 40 000 Year 2 teachers which have it in 2011 since SBA for Year 2 students have been implemented in 2012 (PPD Barat Daya, 2010).

Another example of the training is a course conducted for Year 1, 2, 3, 7 and 8 teachers, called the PSD upgrading course (PPD Maran, 2013). By the end of 2012, hands-on training on SBA was conducted in SBA model schools in all districts. Two schools, one primary and one secondary are chosen from each district. There is also a programme called 'SBA Outreach Program – PIBG Public Engagement' specifically conducted to teachers and parents consortium members. This programme provides explaination and clarification on all issues related to SBA to the teachers and parents throughout the country. Parents are made aware of their ability to get information on their kids' performance from the report.

Table 2.10: SBA training plan

UPSR IMPROVEMENTS ACTIVITY		PMR IMPROVEMENTS ACTIVITY	
Year	Activity	Year	Activity
2010	KSSR and SBA training for Y1 teachers		
2011	KSSR and SBA training for Y2 teachers	2011	KSSM and SBA training for Y7 teachers
2012	KSSR and SBA training for Y3 teachers	2012	KSSM and SBA training for Y8 teachers
2013	KSSR and SBA training for Y4 teachers	2013	KSSM and SBA training for Y9 teachers
2014	KSSR and SBA training for Y5 teachers	2014	Implementing PMR with SBA
2015	KSSR and SBA training for Y6 teachers		
2016	Implementing UPSR with SBA		

2.5 Quality Assurance

The main indicators for QA in assessment are validity, reliability, feasibility and fairness (Lembaga Peperiksaan, 2012). In order to achieve QA indicators, the four aspects of QA which are moderating, monitoring, tracking and mentoring should be implemented. The MOE has instructed all schools to implement QA to ensure that the quality of the SBA implementation follows the stated guidelines (PPD Maran, 2013). Moderating and monitoring process are detailed in Chapter 3, the literature review chapter. So, in this section, only tracking and mentoring processes are discussed.

Tracking is a process to evaluate the suitability, strengths and weaknesses of the assessment instruments used during SA and CA (Lembaga Peperiksaan, 2012). Before the tracking process is conducted, the officers set the date, time and place to conduct tracking and will then, as part of the process, discuss any feasibility issues and the suitability of the assessment instruments used. The officers have to make sure that the development of instruments follows the principles instructed by the MES and if not, they propose solutions and ensure their suggestions are implemented. Finally, reports are submitted either to the head teachers for the internal tracking or the MES for the external tracking.

Next is the mentoring process. It is a process in assisting, guiding and improving teachers' knowledge and understanding in implementing SBA following the procedures and guidelines set forth by the MES (Lembaga Peperiksaan, 2012). The internal or external mentors conduct scheduled mentoring in a meeting or during a professional development session. Or, it could also be any informal programme before, during or after SBA implementation. In assisting and guiding teachers, the mentors can guide them through problems such as guiding them with the performance standard and the application of the management system of SBA, helping them with the instrument preparation, making sure that the assessment implementation is hand in hand with their teaching and learning process, getting feedback on the teaching improvement and solving problems and challenges they are facing. Those responsible for all the four aspects of QA are listed in Table 2.11. Table 2.12 shows the QA work schedules in the SBA implementation.

Table 2.11: Officers responsible for QA

QA Aspects	Internal officers	External officers
Moderating	Head teacher, excellent teacher, head of panels, main instructor, teachers selected by school SBA committee	Officers from MES, JPN or PPD
Monitoring	Head teachers, Assistant head teachers, excellent teachers, head of panels, main instructors, head of school assessor and teachers selected by school SBA committee	Officers from MES, JPN or PPD
Tracking	Assistant head teachers, excellent teachers, head of panels, main instructors, head of school assessor and teachers selected by school SBA committee	Officers from MES, JPN or PPD
Mentoring	Head of panels, main instructors, excellent teachers, head teachers and some selected teachers	Officers from MES, JPN or PPD

MES = Malaysian Examination Syndicate
JPN = State Education Department
PPD = District Education Office

Table 2.12: Work schedules of QA

QA Aspects	Duration	Actions taken
Moderating	SA: Throughout SBA implementation following timetable from the SBA committee CA: After students finished the assignments	MES/ JPN/ PPD/ Schools
Monitoring	Throughout SBA implementation following timetable from the SBA committee	MES/ JPN / PPD/ Schools
Tracking	During or after one cycle of the SBA implementation (for SA and CA)	MES/ JPN/ PPD/ Schools
Mentoring	According to the needs	MES/ JPN/ PPD/ Schools

2.6 Management System of SBA (SPPBS)

SPPBS is an application system to record and store academic and non-academic data (Lembaga Peperiksaan, 2012). All the stored data can be automated quickly and efficiently to produce reports either at the school level, PPD, JPN or the ministry level.

2.7 Latest SBA Implementation

Following a circular letter from the ministry which was KP.LP.003.07.14(81) dated 31st of March 2014, an improvement was made on the implementation of SBA starting on the 1st of April 2014 in all schools which use national curriculum (Lembaga Peperiksaan, 2014). The changes made are as follows:

a) PSD is replaced with Student Learning Development Guide which is more flexible as it only uses bands and descriptors. Teachers can record students' stage of mastery according to their own time and method.
b) There are no more student development files, showcase files and transit records. Student development is reported informally according to the needs, teachers' creativity and availability.
c) The SPPBS online system is also abolished.
d) The Lower Secondary Examination (public examination) for Year 9 students will be replaced with Year Nine Assessment (summative SBA).

e) SBA and CE components are shown in Table 2.13.

Table 2.13: SBA and CE components for primary and secondary level

	Primary School						Secondary School				
	1	2	3	4	5	6	7	8	9	10	11
CE											
CA											
SA											
Psi											
PASCA											

2.8 Summary

This chapter set out to provide the reader with a clearer understanding on the assessment system practised in Malaysia and has focused especially on the SBA system along with its previous and current implementation process. In the next chapter, the literature on SBA is discussed in greater detail.

Chapter 3: Literature Review

3.1 Introduction

This chapter aims to provide a contextual background from the relevant literature on the SBA system and programme evaluation. My study focuses on the research findings from the last two decades, since changes have occurred over this time period, especially in the education context. It is important, when evaluating SBA, to present an overall picture of the subject of SBA and its related elements (learning theories, assessment models, programme evaluation and the associated evaluation dimensions).

This chapter consists of five main sections. The first section reviews the previous literature regarding the definitions and concepts of different types of assessment and SBA and follows with the definition and concept of school improvement. The second section discusses the previous literature on the learning theories together with their implications for teaching and assessment. It follows with a discussion on assessment models that show the relationship between elements involved in the assessment process and their relevance to this study. The fourth section presents the literature on programme evaluations and the various evaluation models that form the basis for the framework model used in this study. The final section discusses the main part of the evaluation dimensions of context, input, process and product with their major constructs which are relevance to this study. All of these sections are the components that are deemed important to the subject of this study which is the evaluation of the implementation of the SBA system in schools.

3.2 Assessment

Assessment as a vital component in education is an integral part of the educational process and a critical part of instruction. Rather than being limited to tests, rubrics or exam marks as may easily be assumed, the development of assessment and testing in education has evolved over time in many countries. This evolution could be driven by various reasons. For instance, it is possible that there was a feeling that the focus of the assessments needed to change

from serving the needs of a particular group of people to a wider group of population or from having various tests controlled by particular group of people to having central tests controlled by politicians (Black, 1998). Alternatively, this evolution could also happen due to the development of new methods in education. It could also be a combination of reasons. Referring to historical records, it is believed that assessment and testing have started in China over 2000 years ago (Black, 1998). In China, a formal written test was conducted for selection into the civil service profession which was the most powerful profession at that time. The test was abolished in 1995 when it turned out to be failing to produce competent officials.

In the 19th century, public examinations in the form of written tests became popular in England and the USA (Black, 1998). In England, the strong reliance on public examinations extended to the mid of the twentieth century but by the end of the century, in 1996, assessment and testing went through a reform with the introduction of a five parallel system called *five systems of testing and assessment*. They included i) standardised test for selection purposes; ii) national assessment at the age of 7, 11 and 14, iii) test from GCSE and GCE bodies; iv) test for vocational and occupational selections and; v) day-to-day assessment of students' learning implemented by their teachers in schools.

As for the USA, the standardised and centralised tests for selection purposes were implemented in the 1920s and 1930s to assess students from various backgrounds due to huge immigration into the country (Black, 1998). Other than serving selective purposes, assessment and testing also served other purposes such as for awarding scholarships, entering jobs, selection for higher education institutions, making decisions for school funding or assessing schools efficiency. Yet these purposes were not directly related to teaching. This has been the weakness of assessment and testing since all of them seem to serve summative assessment purposes only. In 1915, multiple-choice testing was invented but was criticised because it could not really improve students' learning as a whole as it tests only on certain aspects of learning, focuses more on factual recall with less emphasis on the conceptual thinking of students and not directly related to curriculum and instruction (Black, 1998). In conclusion, the differences in the form and purpose of assessment and

testing within the countries are determined by the history and culture of the educational and social systems of that particular country.
Various definitions have been put forward by researchers on assessment. Generally, assessment in education is a process of "deciding, collecting and making judgments about evidence relevant to the goals of the learning being assessed" and it could be distinguishable by its purpose (Harlen, 2004). Assessment is a process of assigning values to a measurement or behaviour on a formal or a non-formal activity (Mokhtar, 2009, p.375). For example, teachers assess their students whether they are 'good' or 'not' in subject like Mathematics or Science based on their marks and also their previous performance records. Mokhtar's definition of assessment looks compatible with Rinaldi's definition as to decide what elements to give value to (Rinaldi, 2006, p.70). Another definition that focuses directly on assessment as a process practised by teachers is the definition by Drummond, cited in Swaffield (2011, p.434), where assessment is seen as a "process of teachers looking at pupils' learning, striving to understand it, and using the knowledge in the interests of the pupils".

Assessment can be defined as a "process whereby evidence is obtained through the outcome of specific questioning such as tests and surveys, and used to determine results based on the findings of such methods. These may also define the status or value of an event, thing or person's abilities based on performance or importance" (Rayment, 2006, p.2). In the late twentieth century, assessment has been defined as "those activities undertaken by teachers, and by their students in assessing themselves, which provide information to be used as feedback to modify the teaching and learning activities in which they are engaged" (Black and Wiliam, 1998b, p.1). The key features of this definition is that assessment happens during the teaching and learning process, used by both teachers and students and provides feedback to both parties to make adjustments on their learning. Since then, this definition has been implemented by most teachers to adapt their teachings in order to meet the needs of the students. Assessment activities conducted by the teachers can vary from observation and discussion in the classroom (Black and Wiliam, 1998a), marking exercise books, judging athletic ability, assessing performance in drama or assessing behavior in class (Rayment, 2006). Moreover, feedback information which is learner-centred and not graded

(Angelo and Cross, 1993) can be obtained not only from the teachers, but also from their peers and the students themselves and thus acts as an important source in restructuring students' understanding or skills (Nicol and Macfarlene, 2004). An example of an adjustment made by a student on improving public speaking presentation is by practising talking in front of a mirror as a result of feedback that he or she receives (Popham, 2008). It is not only students that benefit from the information, but parents also are provided with updates about their children's progress; teachers are provided with feedback on their teaching instruction and stakeholders have data on which to develop policies related to education (Skopje, 2007). This shows how an assessment could be used as a mediator in increasing the standards of learning across the broader spectrum of society.

It looks like the definition presented by Rayment (2006) and Black and Wiliam (1998b) are different but they share some commonalities. Black and Wiliam's definition focuses more on formative assessments and less on summative work (Popham, 2008) whereas Rayment's is more on summative type. This can be seen clearly from the Popham's simple definition on formative assessment as a "planned process in which teachers or students use assessment-based evidence to adjust what they are currently doing" (2008, p.6). However, these different forms of assessment need not be seen as contradictory to each other. Explanations on formative and summative assessment will be discussed in a later section of this chapter.

In conclusion, in this study I will be using the term 'assessment' as defined by Black and Wiliam who perceived assessment as a process of gathering information used by teachers on students during the teaching and learning process. Its purpose is to gain feedback from the teachers, their peers and the students themselves to restructure the students' knowledge, skills and capabilities. And, the aim of assessment is to raise the standard of learning or support learning. Although SBA consists of formative and summative assessment elements, the focus of this study is more on formative assessment. That is why the definition from Black and Wiliam is chosen because it focuses more on formative assessment whereas definitions by Mokhtar, Rinaldi, Drummond and Rayment seem to focus more on summative assessment.

Next is to look at terms that are related to assessment such as measurement and evaluation. Assessment is sometimes confused with terms like measurement or evaluation. In its simplest form, measurement is a process of "assigning numbers to something in a quantitative form" (Mokhtar, 2009, p.22) or a process to gain information by using an instrument and it normally ends up with a quantitative form of information like the raw marks, raw percentage marks, percentile and aggregate (Boon and Shaharuddin, 2011). Measurement is involved when, for example, the English Language skills of a student are measured using an English Language Test, and she or he might get a mark of 30 out of 50. On the other hand, evaluation is "the process of judging, valuing and ranking pupils" (Muijs and Reynolds, 2011, p. 266). For example, evaluation is involved when a student is categorised as 'good' or 'not good' based on his or her test marks and also the previous performance record.

Some authors do not distinguish between assessment and evaluation and terms such as assessment, evaluation or appraisal are used interchangeably (Rust, 2002). In comparing assessment and evaluation in an educational setting, *assessment* is seen as a 'journey' whereas *evaluation* is a 'snapshot'. Although both measure whether or not learning objectives are being met, still there are major differences between them. The key differences of both types of measurement can be perceived using three dimensions of i) the content, ii) the focus of measurement and iii) its usage in the teaching and learning process (Angelo and Cross, 1993). Assessment is seen as an ongoing process to improve learning, focusing on how learning is going and is used in identifying areas of improvement. On the contrary, evaluation is a process to gauge education quality at any final task, focusing on what has been learned in order to gain overall grade or score. So, in the teaching and learning process, assessment matters more than evaluation because assessment being formative gives opportunities for teachers to restructure areas where students may be weaker, whereas the latter acts in a summative capacity that is more beneficial in judging the quality and characteristics of students. In conclusion, during the teaching and learning process, when a test is given and data is collected from the students, that is called *assessment*; when they are assigned grade or mark from the test, it is *evaluation* (Muijs and Reynolds, 2011).

3.2.1 Type of Assessment

There are various ways to classify assessment. For example, Cronbach (1990) approaches assessment whether it focuses on the objectivity of the outcome, the psychometric assessment or the process of it, the impressionistic assessment. It could also be criterion versus norm-referenced assessment (Rust, 2002). In terms of purpose and timing, assessment can be divided into four types: formative, summative, evaluative and educative. It was asserted by NCEST in Valencia (2002) that there is no such assessment that can serve all assessment purposes.

For the purpose of this study, categories produced by the MOE in Malaysia that are clearly relevant to SBA are used (Mokhtar, 2009) which are formative assessment and summative assessment. This classification system is based on the purpose and timing of the assessment but although it should be borne in mind that, formative assessment and summative assessment are not mutually exclusive (Rayment, 2006), both are not contradictory practises, but rather they are complementary approaches (NIC, 2007). To be more exact, formative and summative assessment should not be labelled as two different forms or types but more on how each is being used; this is because each assessment could involve the same tasks but serve different purposes (Mansell *et al*., 2009). The idea of formative and summative assessment started as early as in 1971 when it was introduced by Benjamin Bloom, Thomas Hastings and George Madaus (Stiggins, 2005) followed extensively by Dylan Wiliam and Paul Black from United Kingdom around 1998 (Learning Point Associates, 2009). Moreover, a helpful distinction between internal and external assessment is to think of formative assessment as being an 'internal assessment' as it is an assessment managed by the school and is internally marked. 'External assessment', by contrast, is an assessment held through a national standardised examination and is externally marked (Hopkins, 2006).

a) Formative Assessment

Being one aspect of teaching and at the heart of classroom culture, formative assessment is very important because it contributes to an effective teaching. By definition, formative assessment is a "process used by teachers and students during instruction that provides feedback to adjust ongoing teaching and learning to improve students' achievement of intended instructional outcomes" (Popham, 2008). The process involved between teachers and students is to identify the students' learning goals, intentions and criteria for achievement (Young and Giebelhaus, 2005); in short, it is a continuous and systematic process used to gather evidence about learning (Heritage, 2007a). Next, feedback information gained from the assessment in an educational setting from the conversation between a teacher and the students that is built on "what is known and what is to be learned" (Young and Giebelhaus, 2005) can be used by the teacher to realign the teaching strategies, materials and approaches following learning objectives. Furthermore, for the students, it can serve as a turning point to change their understanding or skills that they have previously misinterpreted and also their motivational beliefs and self-esteem. Feedback could also serve as a medium to promote deeper learning and understanding and to enhance thinking skills amongst students (Hargreaves, 2005). In fact, feedback not only can be generated from teachers' views on performance in class or from students' assignments but also from peers or from the students themselves (Young and Giebelhaus, 2005). Formative assessment not only provides feedback for students, it also serves as a "framework for sharing educational objectives and for charting their progress" (Nicol and Macfarlene, 2004, p.1). Some teachers feel that feedback could increase their workload and they would be getting negative responses from the administration and parents but vice versa happens.

Good feedback must be constructive and can aid the students to think they are being appreciated, such as a comment posed by a student saying that "Now I know that she is interested in what I think, not in whether I have got the right answer" (Black, 2007). Being constructive means that the feedback should be in a descriptive phrase, clear in terms of goals and standards, well-timed as agreed between teachers and students, respectful supportive, motivating, action and solution oriented, strictly confidential in order to maintain trust, specific to observable behaviour and informative. It should avoid generalisations or assumed intentions and interpretations (Hamid and

Mahmood, 2010). Feedback should also avoid being given in the form of grades, comparisons between students' scores or not being specific because those form of feedback has no effect on students' learning (IowaCORE, 2011). Similarly, constructive feedback is also feedback that can make a student think and this could not be achieved by using grades, scores or simple comments like 'good job' (Leahy *et al.,* 2005). Feedback in the form of comments should be clear and exact so that students know what the teachers require of them and are able to improve their work and if possible, comments must be related to the rubrics of the syllabus (Leahy *et al.,* 2005). An example of a constructive comment by a teacher, *"*Five of these 20 answers are incorrect. Find them and fix them*"* (Leahy *et al*., 2005, p. 4). Research has shown that comments alone given by teachers is more effective than giving only marks or marks and comments in order to increase the belief system of students for them to improve (Black, 2007). Formative assessment can be categorised into five different types as shown in Table 3.1 (Wiliam, 2011). So, from here it shows that kind of feedback that students gain from teachers could influence students learning.

Table 3.1: Category of formative assessment

Feedback interventions	Explanation
Weak feedback	Give mark or grade to students
Feedback only	Give mark or grade and feedback on correct answers
Weak formative assessment	Give information on correct answers and some explanation
Moderate formative assessment	Give information on correct answers, some explanation and specific suggestions for improvement
Strong formative assessment	Give information on correct answers, some explanation and specific activities for improvement

Carefully-focused feedback can help students to realise the gap that exists between the current performance and the desired learning goal (Young and Giebenhaus, 2005). Nicol and Macfarlene-Dick (2006) have listed seven principles of good feedback practise which could facilitate self-regulated learning in students, in particular. Self-regulated learning is a process of regulating thinking, motivation and behaviour aspects of students during learning in which they have a greater degree of control of their own learning.

The first one, a good feedback helps students to realise the gap as has been mentioned by Young and Giebenhaus above. Next, it helps students to understand their learning goals, improve their self-assessment, inform them with high quality information about learning and develop their motivational beliefs and self-esteem. It also helps in improving dialogue activities between teachers and students. Lastly, feedback also provides a good source of information for teachers to improve their teaching because by reflecting on students' performance data allowing them to develop students' self-regulation. However, there are also problems with feedback practises (Rust, 2002). For example, not all feedback is constructive. And, some students realize that feedback does not really help them in their learning, and they have problems understanding the feedback itself and even worse, it does not have any effect on students' learning.

Formative assessment is meant to be conducted during the teaching and learning process so it is not an add-on to teaching but rather integrated into the process (Rayment, 2006). But, Heritage (2007a) believed that in reality, assessment processes are still perceived in terms of competition with the teaching process and something external to the teaching practises. Rather than being an integral part of the process, assessment and teaching are seen as two distinct processes and assessment is not even providing information in improving teachers' instructions.

Formative assessment strategies can be oral or written feedback in a classroom environment on a daily basis (Rayment, 2006); they may take the form of a weekly quiz or a monthly test (Mohamad Najib, 2011); check lists, observations and home work (Lembaga Peperiksaan Malaysia, 2011); or even a clinical interview (Ginsburg, 2009). All strategies in formative assessment can be grouped into three main strategies which are 'on-the-fly assessment', 'planned-for interaction' and 'curriculum-embedded assessment' (Heritage, 2007a). 'On-the-fly assessment' is a spontaneous and unplanned assessment made by the teacher, much like a 'pop-up lesson' to clear up students' misconceptions, perhaps as a response to students' actions during teaching. By contrast, 'planned-for interaction' incorporates strategic questions posed to students during an evidence gathering process. Last is the 'curriculum-embedded assessment'. It could be the assessments in the curriculum made by the

teachers; or curriculum developers, as in notebooks; or ongoing activities such as student presentations. In addition, discussions should be developed during formative assessment interaction whereby teachers collect various ideas from students, summarise them and then challenge the students with further questions.

In general, the purpose of formative assessment is to help students, teachers or stakeholders to identify areas for improvement, areas in which students are having difficulties (CERI, 2008). This process of improvement in turn can promote a solid culture of education system evaluation and promote the goals of lifelong learning. Moreover, major learning improvements could be achieved in the classroom environment with the use of well-constructed formative assessment. In what way can this happen? Teachers are supposed to believe that each and every student possesses his or her own state of knowledge which includes knowledge of facts, procedures and the relation between the two. By using a variety of techniques in assessing their students' knowledge followed by a suitable instruction upon them will be the right way to improve students' learning (Hunt and Pellegrino, 2002). So, the identification, analysis and response method made by the teachers towards each student's problem seems to be the important criteria in implementing formative assessment. This is consistent with the decision made by the National Council of Teachers of Mathematics in urging teachers to use formative assessments to determine the state of preconceptions and problem-solving techniques that the students have before learning starts (Hunt and Pellegrino, 2002).

Moreover, Boud (2000) comments that self-assessment skills developed continuously during the formative assessment process might play a very important role in preparing students for lifelong learning, self-motivation and self-esteem (Wang, 2008). Formative assessment should also be a medium for teachers to impart their evaluative skills to the students to ensure that students benefitted from the feedback practises during formative assessment (Nicol and Macfarlene, 2004). Low-achieving students and students with high disabilities can also benefit from well-constructed formative assessment to improve their learning they become more confident while expressing themselves (Wang, 2008). Formative assessment thereby provides an opportunity to develop students' metacognitive skills, skills on 'learning how to learn' (Hinton *et al.,* 2008).

However, an assessment should not be considered as formative if it serves only as a graded test in the end-of-course assessment (Dodge, 2009), or as part of a grading system, a random observation or a good instruction. Of debate also is the issue of formative assessment as not being a graded test; again it depends on the feedback style and relevancy that the students achieved. Black and Wiliam (1998a) give an example of a graded assignment of an end-of-module assessment that might be a form of formative assessment when students are given good quality feedback by the teachers and they can improve on it. On the other hand, ungraded assignments might not be a form of formative assessment if the feedback that is given by the teachers is only a one-short statement like 'good', 'well done' or 'not bad'. Without good quality feedback, students are unlikely improve. Academics think that frequent testing conducted in the classroom is a formative assessment which is not correct, even if learning targets are set and a measure of students' progress is updated (Black, 2011). According to Popham (2008), an interim or bench-mark test is not a formative assessment although it is a *periodic* assessment that allows teachers to do a better instructional job but still there is no empirical research evidence that shows it can improve students' achievement. Perie *et al.* (2007) also support the idea of distinguishing an interim assessment from a formative assessment because an interim assessment might be implemented in a classroom but its result can be reported at a broader level and the timing of the assessments are controlled by the school or district and not totally by the teacher.

Summative assessments, by contrast, can also function formatively as in the end-of-topic test conducted after the first lesson of the topic (Black and Jones, 2006). How can this happen? Preparation made by the students through discussion activities, mind map technique, analysing past test responses, students asking peers for explanations, teachers explaining answers to past examinations or teachers focusing their teachings on the difficult ones are examples of how preparation for summative test could contribute to formative functions of assessment. Students might alternatively generate and answer their own questions before they go for final examination (Gardner, 2006). But, this situation does not apply to summative assessment like high-stakes tests which are set and marked externally. Similarly, formative assessments can also

function summatively; it is a matter of interpretations. The evidence above suggest well-constructed formative assessments can provide the basis for future summative assessments.

Formative assessment is often used interchangeably with AfL. AfL is seen as a 'newer' concept compared to formative assessment since AfL has been introduced in the late 1980s and early 1990s whereas formative assessment had played its role a few decades earlier (Gardner, 2006). Generally, the main purpose of AfL activity is to promote pupils' learning (Wiliam, 2006). In addition, AfL is considered as a formative assessment if its evidence, which is the feedback gained by teachers and students alike, serves as an adaptation of the information to meet students' learning needs. Gardner (2006) believed that AfL as a process of seeking and interpreting evidence focuses on how students learn and yet helps them to improve their learning. AfL becomes a central part of the classroom practises and also a part of an effective planning. However, in this study, formative assessment and AfL will be used interchangeably.

b) Summative Assessment

Summative assessment is an assessment which measures students' level of attainment in the form of monitored examination by subject teachers or board of external examiners at the end of a stage or task (Rayment, 2006; CERI, 2008) as in a semester test or an end-of-year examination. It has been suggested that assessments such as short-cycle assessments, common assessments, benchmark tests, end-of-course examinations, quarterly or monthly formative standardized tests could also be used if the number of summative assessment given to students are increased (Stiggins, 2005). Additionally, portfolios, written essays, extended-response items or multiple-choice tests could be used (Learning Points Associate, 2009). While further methods might comprise individual projects, group projects, vivas, performance and self or peer-assessments (Rust, 2002). Summative could also involve feedback but is in the form of grade or comments that is given to students after they have completed their test, paper or homework (Fynewever, 2008).

Summative assessments act as a form of measuring the result of learning but do not measure students' performance while they are doing any kind of work during the process of learning (Fynewever, 2008). Summative also acts as a judgment of students' progress referring to national standards (Rayment, 2006), promote students for entry into further education, getting certification or entering occupation (Wiliam, 1998), act as an internal school tracking (Harlen, 2005), inform parents, students' next teachers or students on students' achievement or regular grading for schools' record keeping (Harlen, 2004). Evaluating the effectiveness of an educational programme like evaluating curriculum or evaluating consequence on students' future could also be accomplished using summative assessment (Wang, 2008). Together with other information, summative assessment could be used to monitor the performance of teachers and schools (Harlen, 2005). For the ministers, it provides information on education quality and also a means of comparing schools with national education systems (CERI, 2008) or monitoring schools' performance and accountability (Harlen, 2004).

Summative assessment can be divided into two categories depending on the manner in which it is interpreted: normative and criterion assessment. Normative assessment assesses students' progress in relation to other students in a certain group and it involves ranking pupils whereas criterion assessment assesses students to meet pre-specified standards and assigning students with a pass or a fail or competent or not (Mohammad Najib, 2011 and Rayment, 2006). In conclusion, the difference between formative and summative assessment are listed in Table 3.2 (Troy, 2011, p.6 and Mansel *et al.*, p.9).

Table 3.2: Formative and summative assessment

Dimension	Formative Assessment	Summative Assessment
Purpose	Instructional; aims to develop knowledge and understanding	Managerial; aims to assess knowledge and understanding at a given point in time
Implementation	Driven by moment to moment decision; teacher-generated; individualized	Regulated by protocols developed in or out of the classroom; teacher-generated or externally-generated

Timing	During instruction with high frequency	After instruction with low frequency
Scope	Narrow with one or very few learning objectives at a time	Broad with comprehensive set of objectives
Audience	Classroom (students, teachers and parents)	Classroom, administration or public
Feedback	Student-teacher, descriptive forms	System-audiences; evaluative forms

It seems clear that the combination of formative assessment and summative assessment is important in promoting students' learning. This argument is strengthened further by Popham's statement which states students' scores on most of today's accountability tests could not be improved by using formative assessment. As most accountability tests are incapable of measuring the powerful learning produced by formative assessment due to their instructional insensitivity (Popham, 2008). So, this suggests that formative assessment is more important to students' learning.

3.3 School-Based Assessment

This section describes the definition and concept of SBA and the role of SBA in the teaching and learning process. SBA has received a great amount of attention within the educational research field. Different terms are used for SBA within the literature, with little or no explanation given for the reasons why one term has been preferred over another. Among the articles that constitute this literature review, the most frequently used term is SBA. Another term used in educational research that would normally be considered to be SBA is 'classroom assessment' or 'teacher assessment'. However, some researchers consider teacher assessment as only a part of SBA (Thimmappa and Sharma, 2003). Classroom assessment is sometimes referred to as 'internal assessment' (Fan, 2011) while, as discussed above, SBA is also known as 'internal assessment' or 'coursework and continuous assessment' (Grima, 2003).

Classroom assessment is, as the name suggests, an assessment that is conducted in the classroom by the teachers (Black and Wiliam, 2004). The material could come from the teachers themselves or some external agents. Various strategies could be used to implement classroom assessment such as by using traditional written examinations, oral examinations, open book examinations, resources used from the internet, tests or quizzes. Classroom assment can also be conducted using communication with students, informal assessments, content area inventories, come out with products or presentations, classroom works, rubrics and scoring guides, open-response questions, computer-assisted testing, writing journals and learning logs or checklists (Suzieleez *et al.,* 2009). This form of assessment allows interaction between the students and the teachers or amongst the students themselves (Black and Wiliam, 2009). In terms of scoring, it depends whether it serves formative or summative purposes. Classroom assessment could serve various purposes like formative, summative or evaluative purposes (Black and Wiliam, 2004). Popham (2008) suggested that classroom assessment is not necessarily formative or vice versa as is confused by many people but classroom formative assessment remains the best means of formative assessment. All of the characteristics mentioned above are similar to the SBA implementation except for the issue of the locus of assessment. Since classroom assessment is meant to be conducted only in classroom, only strategies which could be applied in classroom setting are considered. For SBA, it could involve assessment of activities outside classroom such as assessments on sports or co-curriculum activities (Lembaga Peperiksaan Malaysia, 2011a).

There is an increasing interest in the SBA system implementation. SBA has been conducted in countries like Australia, New Zealand, Hong Kong, Finland, United Kingdom, USA, Canada, Africa, Sweden, Scandinavia and Singapore (Assessment Support Material, 2001; Faizah, 2011). In Australia, the emergence of SBA started in the late 1960s or early 1970s (Mercurio, 2008) whereas in the United Kingdom, it has also been practised for some time especially with the influenced of the Assessment Reform Group led by Professor Paul Black and Professor Dylan Wiliam. Similarly, Finland and Sweeden first implemented SBA in the early 1970s (Darling-Hammond and McCloskey, 2008).

3.3.1 Definitions and concepts

Various definitions of SBA have been put forward by researchers. Generally, SBA is an assessment managed by the school based on an assessment outline whereby the teacher plans for an assessment of a unit which includes assessment types and weighting, its task and the outcomes for the course listed on the syllabus assessment table (WACE, 2008). Some early experts on assessment in Hong Kong introduced SBA as 'Teacher Assessment Scheme' in 1978 which was an assessment handled by the subject teachers on practical skills over the whole course, particularly chemistry and it contributed to twenty percent of the total marks of the subject (Cheung and Yip, 2003). This is the same definition as applied to Hong Kong Diploma of Secondary Education Examination when SBA is treated as a form of assessment that is administered in schools by the subject teachers according to the stipulated procedures and criteria (HKEAA, 2012a). Marks gained by the students contribute to the final result of their diploma. The students are assessed by various forms of activities such as reading activities, written reports, oral presentations, group discussions, individual presentations, an independent enquiry study, laboratory work, project work or portfolios depending on their subjects. SBA can also take the form of an assessment conducted wholly by the students' own teacher in their own classroom involving planning the assessment programme, identifying suitable assessment tasks and making final judgments (Hamp-Lyons, 2006; Davison, 2007). SBA could also be defined as an assessment conducted in schools and is designed, planned, administered, scored and reported along with the standards and criteria by the subject teachers as instructed by the relevant external authority (Ong, 2010). It involves assessment of academic and non-academic field with recognition and autonomy given to the teachers to conduct it in a formative or summative form but it is fully school-based (Lembaga Peperiksaan Malaysia, 2011a).

Generally, SBA can either be formative or summative assessment but does not extend to external assessment (Mercurio, 2008). A formative SBA is an assessment which is school-based to promote students' learning and is conducted in line with the teaching and learning process using various methods of gathering information such as worksheet, observation, quiz, checklist, assessment report, homework or test (Lembaga Peperiksaan

Malaysia, 2011a). By contrast, a summative SBA is an assessment which is school-based providing a record of a student's overall achievement at the end of each learning unit at a certain time such as at the end of the month, semester or year using various methods of gathering information such as monthly or semestered testing (Harlen, 2004; Lembaga Peperiksaan Malaysia, 2011a). In order to see the difference, the definition of an external assessment is indicated. External assessment is an assessment task set by an authority external to the school and each student's performance is assessed by a person appointed by the external authority and not by their subject teachers (Mercurio, 2008).

With SBA, it is hoped to change the culture of pending final exam, unannounced 'pop quiz' and the threat of low or failing report card or sorting and ranking students according to their achievement (Stiggins, 2005) or 'teach to test' phenomenon (Lim and Zhao, 2005) to a culture with less focus on examination but more on students' achievement in learning context (Mokhtar, 2009) and all students learn up to the specified standards (Stiggins, 2005). Linn and Miller (2005) and Wang (2008) also commented on the recent global educational assessment reforms whereby cultures are changing from the dominant examination culture to more flexible assessments, which the authors refer to as called 'quality-oriented education' with various assessment techniques to develop students' creativity and practical abilities.

The SBA component should be developed and administered as an integrated part of the curriculum, not as a 'separate' paper by continuously using the assessment information during teaching and learning process as a medium for improvement (Cheng *et al.,* 2011; HKEAA, 2013). Another characteristic of SBA is to promote students' active involvement in learning involving development in self-assessment skills and peer-assessment skills together with teacher feedback (Davison, 2007; Cheng *et al.,* 2011), shared knowledge about students' current learning and appropriate actions taken in order to improve learning (Learning Point Associates, 2009). Self-assessment skills can be understood as skills developed when one understands the learning goal and are then able to assess steps taken in achieving them whereas peer-assessment skills are developed when students are trying to teach their peers, being assessed by peers or have mutual assessment in their 'own natural'

language (Black, 2007; Wang, 2008). Self-assessment requires that the students are capable of assessing their own performance tasks, checking their own learning outcomes or showing their achievements after completing their tasks (Iberson, 2012). Furthermore, Black and Wiliam (1998a) believe that attention to these two sets of skills can lead to appropriate actions in closing the gap in student learning because students will tend to think and learn more with those skills. Active involvement in learning could also be promoted by asking students to keep and review their own progress regularly as most of the skills in SBA will be tested again in the external examinations (HKEAA, 2013).

The essence of the SBA system effectiveness lies in the relationship between the teachers and the students, so both have their roles in SBA. That teachers understand their roles in the classroom is a very important factor in giving meaning to the full implementation of SBA, in order to avoid a mismatch between the stated aims of SBA and the realities in the classroom. Firstly, in order to implement formative assessment effectively, teachers have to have the following four basic forms of knowledge: domain knowledge, pedagogical content knowledge, knowledge of students' previous learning and knowledge of assessment (Heritage, 2007a). These have been explained in the section on teachers' understanding of SBA. Then, HKEAA (2013) stated that teachers must have strong assessment skills on assessing students' work, following the assessment criteria. Secondly, in implementing formative assessment, teachers are responsible for monitoring students' performance by the setting of targets, recognising the gap between the students' current performance and the desired standard and then be willing to provide various strategies to adjust learning to close this gap (Hargreaves, 2005). But, researchers believe that this role should be shared with the students as well if the students are to gain the best learning outcome (Heritage, 2007a). Thirdly, the most important aim in formative assessment is to improve students' performance by means of feedback. Heritage (2010c) believes that in conducting effective feedback, teachers need to focus on the students' task by giving them suitable suggestions, hints or cues and not to give comments only on students' performance. Fourthly, teachers are also expected to explain all the relevant information on SBA to students, to submit SBA marks on time and to provide feedback to the authorities (HKEAA, 2013).

Student roles are also of major importance. Firstly, students as integral partners in the assessment process need to know where they are heading and be willing to follow the direction with the constructive feedback that they gain (Learning Point Associates, 2009). Moreover, they should also be clear of the level they are at and any gap between their current performance and the desired standard performance (Hargreaves, 2005). Secondly, students should be willing to develop skills of self-assessment and self-reflection in learning by evaluating and giving feedback on their own work or their peers' work. It could also be an action of making decision on the best feedback that they expect to gain from their work or reflecting their own achievement (Nicol and Macfarlene-Dick, 2006). Thirdly, student engagement is important in developing scoring rubrics and quality and work up to the standards in a formal way (Learning Point Associates, 2009). Fourthly, students should try to develop their self-assessor skills, metacognitive strategies and their perceptions on assessment as a self-improvement process. Lastly, students are required to undertake their assessment tasks honestly, responsibly and timely and also willing to keep a proper record of their SBA-related work for later inspection (HKEAA, 2013).

3.3.2 Role of SBA

The role of each of the parties needs to be clearly laid out. Since SBA is being discussed in the two different forms of formative and summative assessment, the role of the teachers and the students is also discussed in a similar manner. In the eyes of an educator, SBA is seen as a medium to acquire generic competencies such as analysing problems and problem-solving (Thimmappa and Sharma, 2003) and to encourage learning and maintain standards (Joughin and Macdonald, 2003). Moreover, the feedback gained enables teachers to improve their teaching effectiveness (NIER, 1999) which not only provides them with information on the teaching effectiveness, but in general it gives teachers information on decision-making (Suzieleez *et al.,* 2009). SBA, especially formative assessment, could reduce the overall achievement gap between the low achievers and other students by helping low achievers to improve learning (Young and Giebelhaus, 2005). With SBA, teachers can assess students' areas of performance which cannot be assessed in public examination, as well as enhancing students' lifelong learning and self-evaluation skills (Davison, 2007).

SBA could also improve assessment practises of teachers and also their professionalism (Esther, 2012). SBA does not only serve formative purposes but should serve summative purposes as well. Educationists should view SBA as a medium in measuring the result of learning (Fynwever, 2008) and in making judgment of students' progress referring to national standard (Rayment, 2006). A well-designed SBA system can also be a means of informing parents or students on the students' performance (Harlen, 2004). Both assessment purposes could also improve assessment practises of teachers (Esther, 2012). In short, SBA acts as a medium to develop and support high-quality learning of students and also to summarise students' learning at some point in time (Harlen, 2004).

In the eyes of a student, SBA is seen as a medium that provides them and their parents with the essential information about their progress, thus encouraging and assisting them to take actions to improve (Harlen, 2004). Not only that, when SBA is implemented, it seems to remove many negative issues on 'one-shot' external examination that has long been haunting the students and covers some learning objectives that is not covered with written test (Cheung and Yip, 2003). Students can play an active role through self-assessment or peer assessment during the assessment process (Davison, 2007). This will in turn develop students' meta-cognitive strategies which enable students to have a strong control of their own learning process, all of which can help students to develop and apply those skills to their daily life problem-solving (CERI, 2008). Other than helping students to understand their strengths and weaknesses, SBA also increases students' confidence level and motivation to learn (HKEAA, 2012a). By means of summative purposes, students should understand that summative assessment is important when it comes to promoting them into further education, getting certificate or applying for jobs later in life (Wiliam, 1998; Harlen, 2004).

In the eyes of authorities, such as the ministry and the school administrators, SBA is seen as a system that is able to supply students with the acquired skills which are needed in the twenty-first century skills (Cheah, 2010). Those skills include critical thinking, problem-solving skills, computer and technology skills, communications and self-direction skills. And, all the while, the issue of being too focused on public examination has been burdening many parties.

With the introduction of SBA, it might reduce pressure in public examinations (Esther, 2012). SBA could also increase literacy and numeracy skills of students (Department of Education and Skills, 2011). In terms of SBA's provisions of summative assessment, it acts as an internal school tracking system (Harlen, 2005), provides information on education quality and enables comparisons be made between schools and the national education system (CERI, 2008), evaluates the effectiveness of an educational programme such as curriculum evaluation and evaluating consequences (Wang, 2008). Together with other information, summative assessment can be used to monitor performance of teachers and schools (Harlen, 2005), to gain a certification or accreditation of learning by external bodies or as a selection for employment or higher education (Harlen, 2004). Summative assessment could also act as a means to monitor achievement trends or students' proficiency levels or as a source of decision-making in resource allocation and goal-setting (Young and Giebelhaus, 2005). So, it is clear from the above arguments that SBA is beneficial because it introduces both types of assessment; the formative and summative assessment.

3.4 School Improvement

Education is going through an improvement in many countries as it is central to improving quality of life. Major educational reform such as individual school improvement has been recognised since the 1970s (Chapman, 2011). School improvement should serve the betterment for everyone in the country, creating an effective learning environment and enhancing outcomes for all students. And, most countries are hoping that reform which has been set out and legislated is able to give a significant impact on raising levels of student learning.

3.4.1 Definition of School Improvement

School improvement or school reform serves as one level of a tri-level whole system reform in which school and community act together, with the district and state levels forming the remaining components in this multi-way interrelationship of educational reform (Fullan, 2010c). Whole-system reform is the implementation of several core policies and strategies in every classroom,

every school and every district in the state, province or country for the betterment of everyone in the system (Fullan and Levin, 2009) with a full commitment from top political leaders (Fullan, 2010c). Fullan (2009a) asserted that although there exists a multi-way interrelationship, the levels do not necessarily have to reach a certain standard before certain level could have an interconnection and mutual influence with other levels. But, each level must continuously focus on its internal development. School, being the most important unit of change in any educational reform process is the focus of this study.

Before moving on to school improvement, it would be helpful to consider what educational change has to offer. Educational change is a multidimensional approach consisting of several aspects, be it at classroom or teacher level, school level or district level (Fullan, 2001). Educational change can best be understood as a process and not an event (Hopkins *et al*., 1994). Moreover, it is a process with a series of events to be implemented in schools to achieve a set of educational goals or engender a particular educational philosophy. Fullan (2001) proposed three main aspects that contribute to educational change: new or revised teaching materials or technologies; new teaching approaches, strategies or styles; and beliefs or values towards new policies or programs. Out of these three aspects, changing the belief system seems to be the most difficult one to implement because it involves changing people's values. Changes in teaching approaches could also present challenges to some teachers as they may be required to develop new skills or instructional activities. On this basis, any change in an education system, especially at a classroom level, requires a change in teaching materials, teaching strategies and attitude of teachers which includes cognitive, affective and behavioural aspects. These aspects listed by Fullan are limited in presenting an overall picture of the educational changes initiatives.

Hopkins *et al.* (1994) had previously suggested another two features that contribute to the changes: structural changes and changes in the organisation of schools, such as in time tabling or the perceived imposition of new knowledge to teachers, such as, for example, on the acquisition of skills in order to adapt to changes in information technology. Other than involving *theories of education* as mentioned above ('what to change?'), Fullan believed

that educational change also involves *theories of change* ('how to change?') to explain ways to implement those changes. And, both theories seem to interact with one another in making changes in education so both are important and should be treated equally. The latter theories will be discussed in the later section of this chapter when discussing school improvement.

By definition, school improvement is "a distinct approach to educational change that enhances student outcomes as well as strengthening the school's capacity for managing change" (Hopkins *et al*., 1994, p.3). This suggests that in order for schools to improve, strategies and efforts which modify the internal conditions of schools, including school capacity, should lead to a change in school culture which will then influence the teaching and learning processes that lead to increases in student's achievements. The 'internal conditions' of schools are the teaching-learning activities and all procedures, role allocations and resources used to support activities in schools. These internal conditions do not necessarily have to be consistent with the national reform agenda at all times. Yet it cannot simply be said that school improvement initiatives equal educational change because there are cases where the changes implemented do not impact upon student outcomes or are concerned about school cultures (Hopkins *et al.,* 1994) whereas the aim in school improvement is to raise student achievement and enhanced school capacity in managing change.

3.4.2 The Concept of School Improvement

Harris (2003b, p.70) asserted that the aim of school improvement is to transform school cultures. It is also observed that school culture is the most important feature contributing to the success of the school improvement process (Hopkins *et al.,* 1994). Given the importance of school culture, this section considers how school improvement initiatives could lead to cultural changes in schools and enhanced student achievements.

School culture is the shared beliefs and values within a school; is the things that people agree are true and things that people agree are right (Hobby, 2004). Generally, people in an organisation would agree with all the things that they assume to happen in certain ways. For example, it could be something

like teaching methods or communication skills of a teacher or the way teachers respond to challenges within a school environment. And they would also agree with all the values that they expect to happen in certain ways such as the balance commitment between work and home by teachers or respect made towards a head teacher by staffs and students. So, the culture of a school is a social force involving all the school participants.
Levin (2001) stated that the culture of school does not include only the shared beliefs and values, but also the shared understandings, behaviour and attitudes of school participants which are the school staff, students, parents and members of the larger communities. Other examples of school culture are students' expectations on the roles of adults in a school or teachers' opinions on acceptable educational practises. No two schools are alike; each school has its unique culture and this makes reform even more complicated. School culture is based on tradition, habit, expectations, opinions and images within the school itself (Levin, 2001). Other than the general culture that shapes the culture of the school, school culture could also be affected by the school history and the background or the specific populations within the school. Changing a school culture is difficult because it involves changing a blend of tradition, habit, expectations and images of the schools and their operations.

Although school culture has a very important part in improving the quality of student learning, it is important to state that culture is interdependent with the school structure. School structure can refer to the school size, school overcrowding or schools-within-schools (Ready *et al*., 2004). These two factors, school structure and school culture could influence one another in making changes in schools. Since they have a mutual relationship, equal attention is needed for both if schools are to improve although changing school culture is believed to be much more challenging than changing school structure (Fullan, 2006; Hopkins *et al.,* 1994).

What, then, are the strategies that can lead to a change in school culture? It could be capacity-building strategies. School capacity is a collective competency consisting of components such as individual development (teachers' knowledge, skills and dispositions), organisational development (school-wide professional learning communities), organisational integration (programme coherence, technical resources) and lastly is the principal

leadership (Newmann *et al.,* 2000). From the Newmann model, it shows that there are five dimensions that make up the potential of school in improving the instructional quality in terms of its curriculum, instruction and assessment, which can lead to an improvement in students' achievements. Strategies or actions taken in building capacity as suggested by the model are as follows:

1. Teachers who are supposed to be the catalyst for classroom change should be well-prepared with the suitable knowledge, skills and attitudes towards the curriculum, instruction and assessment. This could lead to an improvement in students' achievement. Such teacher resources could be gained through professional development or staff development (Hopkins *et al.,* 1994). Schools should give opportunities for teachers to work together, share ideas, observe one another and give feedback so that learning of teachers could be developed and the quality of pupil learning could be enhanced. Furthermore, some strategies, like collaborations, mentoring and meaningful debate on teaching practises which are conducted between teachers could develop their attitudes and yet could improve schools as a whole (Harris, 2003b). Hence, research has shown that teacher performance in the classroom and self-esteem has a direct effect on school culture and vice versa.
2. Next is a professional learning community of teachers to support the individual development of teachers. Professional learning community concerns the relationship among teachers and also between teachers and the school principal (Fullan, 2005). This relationship quality, if improved could develop better knowledge creation and also stimulate changes amongst teachers (Harris, 2010) and an improved students' achievement. Examples of the improvement initiatives that can enhance learning communities of teachers include teachers' participation in decision-making, their engagement in collaborative work or accepting the joint responsibility between them (Harris, 2003a), collaboratively setting clear goals for student learning, conducting assessment and consequently developing action plans (Fullan, 2000). But still, Fullan (2007) believes that its implementation is not yet up to the standard.
3. Developing teachers' capacity together with professional learning communities may be insufficient to change the culture of school, particularly if innovations appear unconnected or changes introduced in

such a way that teachers feel they are being bombarded. So, another factor has to come in, which is the coherence of programme, meaning that any programme at school has to be conducted in a coordinated manner, focuses on clear learning goals and has to be sustainable if students and staff are to be empowered. Schools also have to be selective with the innovations that are imposed on them (Fullan, 2000).

4. Next is the technical resource. This includes high quality curricula, books, instructional materials, assessment instruments, laboratory equipment, computers, technology, workspaces, time or access to expertise (Fullan, 2000).

5. The final initiative is the principal leadership component. This initiative is very important because it gives a great influence in fostering all the four main aspects of school capacity that have previously discussed. The former United Kingdom Education Secretary, Michael Gove stated that head teachers are crucial to school success (Burns, 2012). A head at school plays a crucial role in managing school capacity by developing individual teachers with knowledge and skills followed by developing team- work between them and then make sure that all the programmes are implemented in a coordinated manner together with the technical resources to support innovations (Fullan, 2000).

As Harris (2003b, p.68) stated that "school leadership is more than the effort of a single individual and not static", leaders should also be able to create a suitable learning organization for staff and student. Schools as a learning organisation is a place where "employ processes of environmental scanning, develop shared goals, establish collaborative teaching and learning environments, encourage initiatives and risk taking, regularly review all aspects related to and influencing the work of the school, recognize and reinforce good work, and provide opportunities for continuing professional development" (Silins *et al.,* 2002, p.24). And, if this happens, classroom practices could be improved leading to a subsequent improvement in student learning. School leaders should be able to transfer the leadership role to teachers to develop teacher leadership. Teachers who have a direct link with the teaching and learning process can also play a very important role in leadership. In conclusion, effective leaders tend to develop and implement their vision in various forms of school capacity building strategies which in

turn could harness the learning of students, teachers, parents and even governors.

In conclusion, this concept of how schools contribute to student achievement implies that any strategies or action taken should attempt to address all the five aspects of school capacity. But, Fullan *et al.* (2001) believed that the five dimensions of the Newmann Model for school capacity was inadequate for improving schools as a whole as they do not include parents and community relationship factors and are also lacking strategies related to assessment literacy, such as the interpretation of students' data to improve teacher practices.

The interconnected and interrelated dimensions are formed into a framework in initiating reforms in schools and producing superior learning, known as 'The Innovation Framework' (Partners in Learning, 2009). This framework is adapted from the work of Knapp *et al.* (2003), which provides a very good alternative and response to the Newman Model and Fullan's critique as an innovation in the main areas of school reform. It consists of an interconnection of four main dimensions of i) teaching, learning and assessment; ii) ongoing professional development; iii) leadership and a culture of innovation and; iv) learning environments, as shown in Figure 3.1.

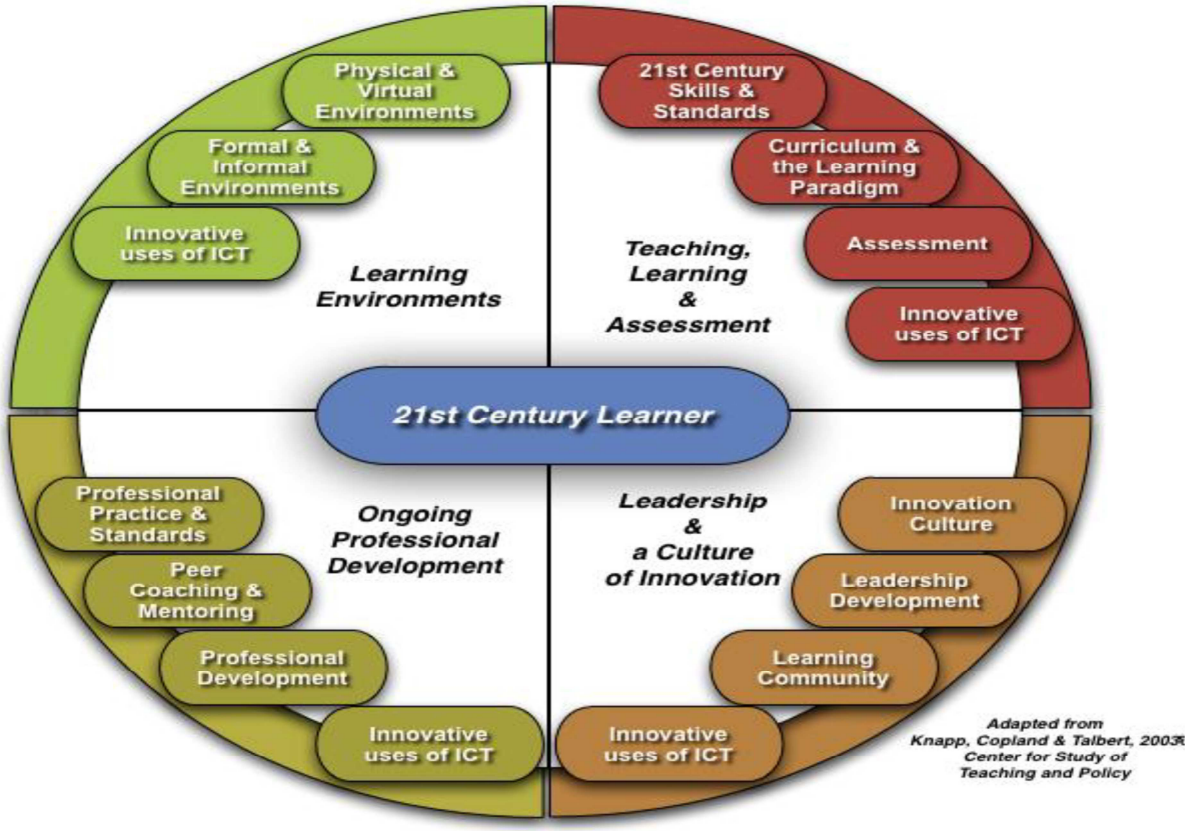


Figure 3.1: The Innovation Framework of Major Areas of School Reform

The teaching, learning and assessment dimension includes the design of curriculum, the use of suitable assessment processes and teaching and learning strategies, the integration of ICT into learning and active engagement of students in preparing appropriate skills for the students. Next, schools are required to develop an ongoing professional development for teachers. This kind of professional community establishment, with its focus on learning, aims to improve the skills of teachers, determine the appropriate training and development needed and find ways to measure the effects of the development. By means of leadership and a culture of innovation, schools have to develop a dynamic and empowered school culture, having relationships with people from various levels, having instructional leaders, changing school structures and also having a continuous school evaluation. Lastly, reform should also represent schools that are able to design a learning environment which includes a physical and virtual environment or a formal and informal environment that is suitable for teaching and learning practises along with easy access to technology.

Having considered this model, the focus is on the discussion on the strategies adopted in the Malaysian education system in improving student outcomes. The strategies are listed below (Economic Planning Unit, 2010);

1) Developing top talent in teaching. This includes:
    - upgrading the quality and professionalism of existing teachers especially in classroom instruction
    - improving the quality of new teacher recruits
    - making teaching the profession of choice
    - removing guaranteed placement of trainees
    - delivering continuous professional development
2) Investing in great leaders for every school which includes strategies such as:
    - introducing New Deal approach
    - increasing support and coaching
3) Ensuring that every child can succeed. This is the only strategy that focuses directly on students rather than teachers and thus differentiates it from the above model and framework initiatives. The strategies are:
    - providing a stronger head-start for more children
    - expanding pre-school enrolment and improving pre-school education quality
    - lowering entry age for schooling
    - improving literacy and numeracy skills through Literacy and Numeracy (LINUS) program
    - upholding Malay Language and strengthening English proficiency
4) Holding schools accountable for changes in student outcomes by:
    - implementing the School Improvement Programme (SIP)
    - implementing the High Performing Schools (HPS) Programme (see below)
5) Transforming the effectiveness of delivery by:
    - delivering 'Service Lines' to support school improvement
    - implementing a new curriculum
    - developing pre-schools by public-private partnerships
    - developing basic education by public-private partnerships

The SIP is a sub-programme within the broader 'Government Transformation Program [sic]' which aims to improve student outcomes (Asli, 2012). It was launched in April 2010 (Economic Planning Unit, 2010) and has been implemented since July 2010 with the aim of helping and guiding all primary and secondary schools, especially those are at Band levels 6 and 7 (low performing schools), to improve their performance as stated in the circular letter KPM.NKRA.100-15/1/1 Jln 3(7) dated 23 February 2011. According to Pemandu (2010), there are seven bands of schools with Band 1 indicating the best performing schools and Band 7 the lowest. By February 2011, the implementation of SIP has been undertaken in collaboration with all the parties in the MOE, state education department, district education office, teacher education institute and all the schools nationwide. There are 10,000 schools involved altogether. Two initiatives have been introduced which are the School Improvement Partner (SI Partner) and the School Improvement Specialist Coaches (SISC). The SI Partner's role is to help and guide the principals and the head teachers from Bands 6 and 7 schools whereas the SISC is designed to help the teachers in upgrading the subjects' grade achievement. There is also a centralised data collection system and self-help tool called the School Improvement Toolkit used by the head teachers to track school performance (Pemandu, 2010). Another programme is the High Performing Schools (HPS) Programme. The HPS are schools with a unique identity having a strong work culture and strong international linkages and networks. The students are considered excellent in all aspects of education. The schools have a very high school standard and able to provide coaching and mentoring support to other schools.

Considering the action and strategies taken by the Malaysian education system, there are some similarities and differences to the Newmann Model and the Innovation Framework proposed by Knapp *et al*. All the models focus on improving curriculum, assessment, teaching and learning strategies and the integration of ICT into the teaching and learning process along with the professional development of teachers, leadership development and continuous school evaluation. The differences are that the Malaysian education system focuses more on improvement of the pre-school education and skills for children at the early-age. Furthermore, the strategies in Malaysia lack in activities considering the coherence of all programs conducted in schools,

technical resources supplied to schools, the development of school culture, school structure and the learning environment in supporting learning. Overall, the strategies show that the government is taking appropriate actions in strengthening school capacity and supporting all students without neglecting underperformers. There is, however, still a long way to go for the Malaysian education system in improving schools but as an effort to initiate the process of reform, it is a good start.

For this study purpose, I will focus on assessment as one component in the curriculum and the role of assessment in improving schools which is detailed in section 3.10.9. Its role is important as the impact of classroom is the most important feature in school improvement.

3.5 Assessment Models

At the beginning of the chapter, discussions were made on assessment in great detail. Some researchers try to integrate all the elements involved in assessment activities based on research findings from the literature review into models of assessment. To explain, a 'model' is a conceptual picture on any activities that shows the relationship between the elements involved in those activities (Razali, 1987). Since assessment, especially SBA, is the central focus of this study, assessment models will be discussed in this section. There are many assessment models and most of them have been tested in schools in many places. A model can be used as a tool to analyse and improve teaching practices (Nicol and Macfarlene, 2004). So, assessment models can be used as tools by teachers, administrators or any stakeholders to analyse and improve the implementation of SBA in schools and also provide information for the researchers to improve the evaluation process of any programme. Thus, this study will use assessment models as a guidance to determine the elements involved when evaluating the SBA system implementation or in other words, it is to determine elements to posit hypotheses of this study. Three models are discussed here:

a) The Formative Assessment Model (IowaCORE, 2011)
b) The Logic Model (Wiliam, 2009)
c) The SCAP Model (Rust *et al.,* 2005)

a) Formative Assessment Model

The Formative Assessment Model was developed by Margaret Heritage of the 'National Center [sic] for Research on Evaluation, Standards and Student Testing' (CRESST) (IowaCORE, 2011). The dynamic relationship between the elements contained in this model is shown in Figure 3.2. The elements are actually developed from the major attributes identified by the FASTSCASS of the Council of Chief State Officers together with reviews from the national and international researchers in formative assessment. The five major attributes that they have found are learning progressions, learning goals and success criteria, feedback, self- and peer-assessment and classroom culture.

The formative assessment model consists of four core elements which are identifying the gap, feedback, student involvement and learning progressions (Heritage, 2007b) in a form of a cycle as an ongoing process. The element on identifying the gap in student learning can be viewed as identifying the gap between the current state of learning and the desired educational goal. This gap represents the 'zone of proximal development', as noted above by Lev Vygotsky. If teachers practise formative assessment effectively, they should be able to identify the gap for each student and scaffold or support them to close the gap by adapting their teaching. Support can also come from students' collaborative processes with their more expert peers.

Feedback in formative assessment can be viewed as descriptive and criterion-based information from teachers to students on their current learning status, their misunderstanding or on how to improve. Or, as defined by Sadler (1989, p.120), feedback is an "information about how successfully something has been or is being done". Feedback could also be interpreted as information about the gap discussed above. But, the most important thing is that feedback should inform and benefit both groups; the students and the teachers. Feedback is supposed to serve formative assessment at various levels of learning. The main aim of feedback is to improve students' learning and also to improve their affective aspects like motivation and self-efficacy which have a direct effect on learning. Feedback is explained in detail in the formative assessment section earlier in this chapter. As discussed above, if students were to benefit from formative assessment in their learning, they would have to be actively involved in their own assessment. Such active involvement in

assessment is acquired when they develop skills in self-assessment and peer-assessment and also through collaboration with their teachers as they develop their self-regulatory strategies or metacognitive processes. All of these skills are important so that they have a shared understanding with their teachers on their current learning status and to ensure that an appropriate action can be taken to achieve the desired goal along the learning progression.

Learning progressions are series of progressions of what should be learned by the students towards the ultimate goal. Teachers can use these as a benchmark to determine the current status of their students' learning and the expected progression. Learning progressions also functions to support instructional planning and formative assessment (Heritage, 2008). Learning progressions support formative assessment as it is the basis for all the main criteria of formative assessment such as the process of gaining evidence of learning, giving feedback to students and an active involvement of students during formative assessments. But, to develop learning progressions is not an easy task for the teachers.

So, from this model it can be concluded that formative assessment is a continuous cycle process that starts with the establishment of learning goals of students. Together with the criteria for success and the identified gap, teachers plan learning or modify their instructions accordingly. Then it follows with students' self-assessment, peer-assessment and collaboration with teachers to scaffold learning. At this stage, student performance concerning their knowledge, concepts or skills should improve. Then, the process of assessment continues until the ultimate goal in learning is achieved.

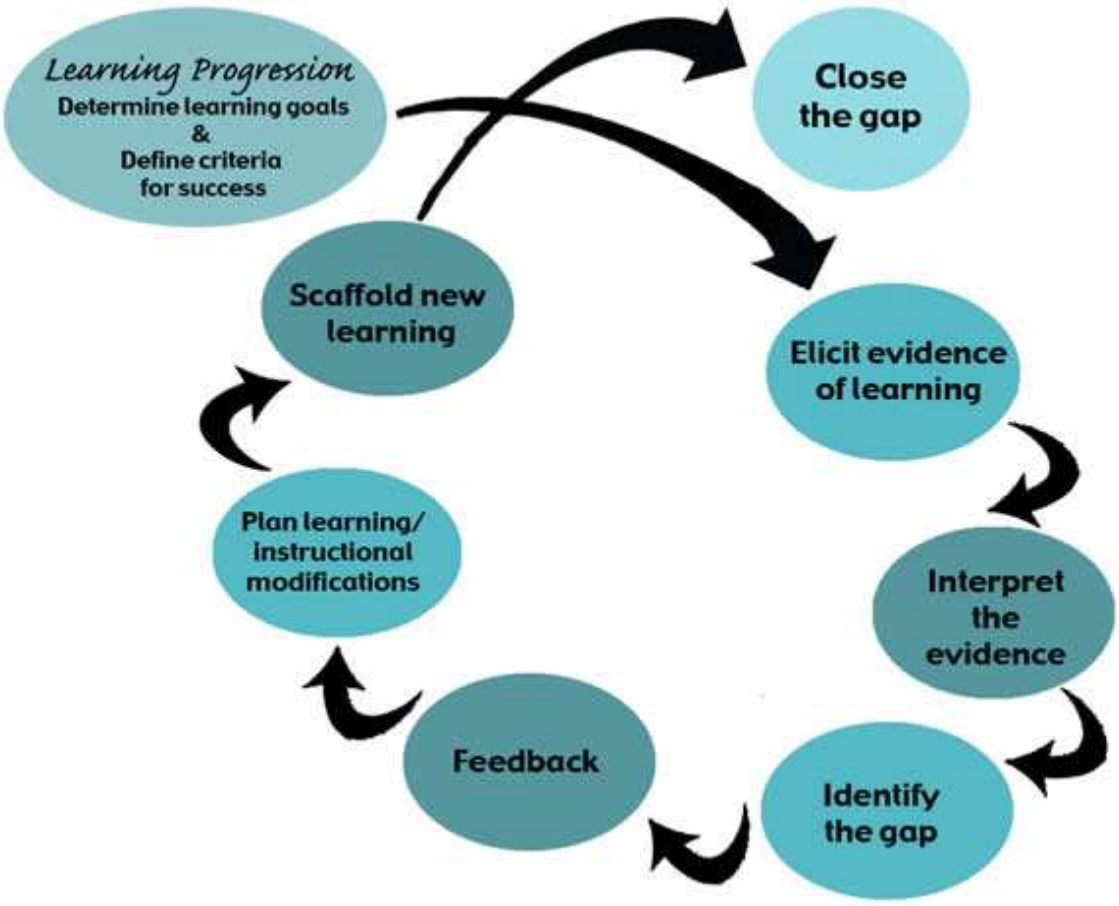


Figure 3.2: Formative Assessment Model

b) Logic model

The logic model of educational intervention is proposed by Siobhan Leahy and her colleagues from the Educational Testing Service (ETS) (Wiliam, 2009). This model has been used as a package for teachers under the 'Keeping Learning on Track' programme (KLT). Being a 'commercial product' resulted from research on formative assessment led by Dylan Wiliam and the ETS team of researchers, KLT program serves a two-year professional development programme for teachers (NWEA, 2013). This programme has been implemented since 2004 in the US school districts in California, Delaware, Maryland, New Jersey, North Carolina, Ohio, Pennsylvania, South Carolina, Texas, New York, Hawaii, Colorado, Tennessee, Vermont and also in Australia. Furthermore, it has been tested in all subjects in various classrooms, schools and districts. The programme includes workshops to build teacher experts followed by building local capacity that allows teacher experts to practise it and develop an ongoing teacher learning community (TLC) in their schools.

TLCs are groups of teachers who continually inquire into their practice and, as a result, discover, create and negotiate new meanings that improve their practises (NCTE, 2010).

There are three components altogether in the logic model for the KLT programme as shown in Figure 3.3. Two components focus on intervention and one component focuses on the outcome of the intervention (Wiliam, 2009). The far left component is the KLT components or the components of interventions. It includes KLT Foundations, TLC Leader Foundations, 'KLT Program Modules' and support materials. It comprises materials like sixteen modules, workbooks and guidebook and also facilitated events to train teachers in formative assessment, in particular. As suggested by Wiliam (2008), other activities that could be conducted amongst teachers are video watching, a book study or doing shared activities such as marking samples of student work. This then causes changes in the middle component. The component in the middle of the figure consists of 'theory of action' of the intervention which includes several teacher outcomes and student outcomes. There are five main subcomponents, which are the role of teachers to continuously elicit evidence of student learning, to identify and share learning expectations, to structure opportunities for students to be responsible on their own learning and motivate students and also to provide feedback. This in turn causes changes in the last component, the student outcomes, leading to improvements in pupils' learning, including their behaviour and achievement. Actually, the theory of action describes both intention, which is how the intervention works and a belief of how intervention works and if it really works.

In order to improve student learning, which is the main goal of this model, training on feedback is provided to the teachers. This gives the opportunity for the teachers to improve their instructions. This in turn can improve students' activities when students respond to their teachers' feedback plus support from their peers and they become responsible for their own learning. This model also presents not only the theory but also the expected reality, called 'the evidence base'. This can be seen from the numbered arrows in Figure 3.3. So, this model or product seems to answer 'how' and 'what' should be done if someone were to implement a better educational intervention on formative assessment in aiming for an improved students' learning. Wiliam (2009)

believes that as long as teachers focus on the five main key strategies of formative assessment as shown in Table 3.3, the intervention will create a positive student outcome. Each strategy might have several assessment techniques and teachers could modify them as long as they are valid and consistent with research on formative assessment (Leahy and Wiliam, 2009). The strategies are as follows:

i) Sharing Learning Expectations: Clarifying, sharing and understanding learning intentions.
ii) Questions: Engineering effective classrooms discussions, activities and tasks.
iii) Feedback: Feedback that moves learning forward.
iv) Peer-Assessment: Activating students as instructional resources for one another.
v) Self-Assessment: Activating students as the owners of their own learning.

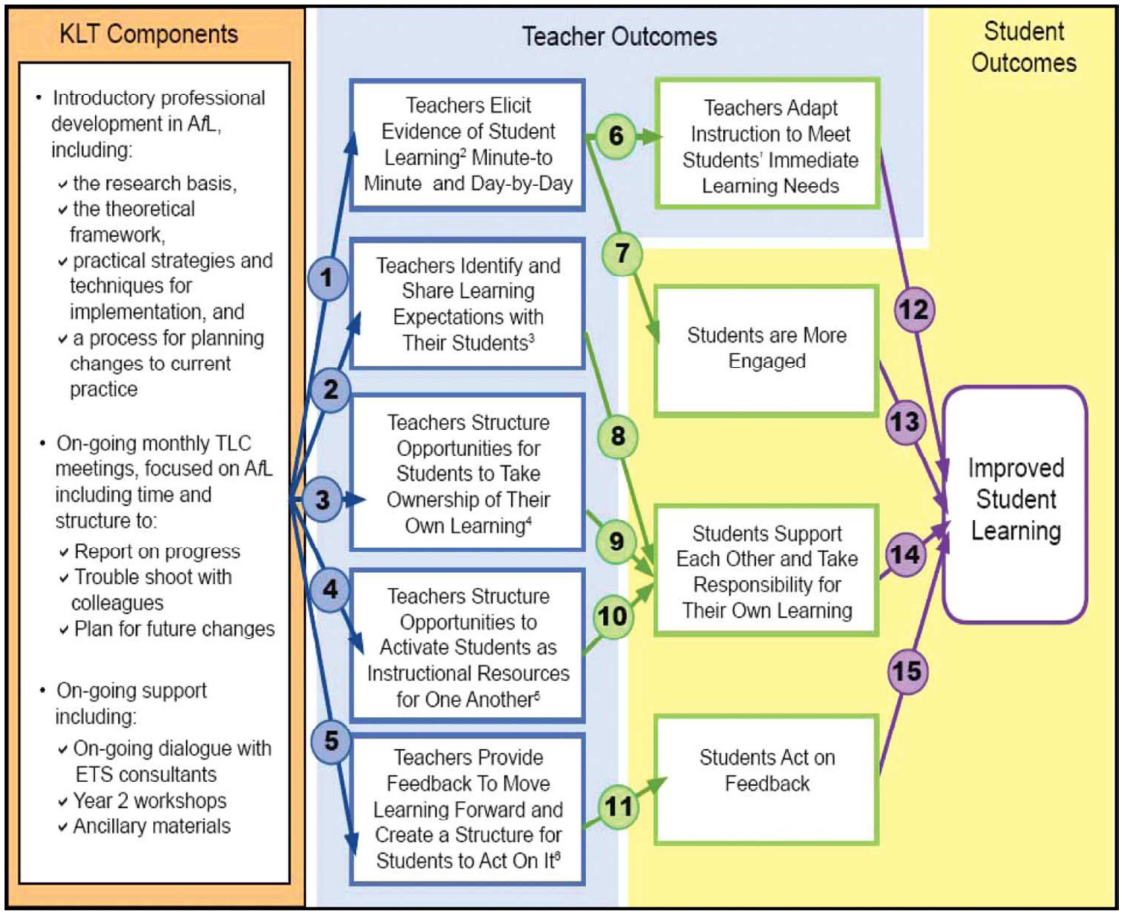


Figure 3.3: A Logic Model

Table 3.3: The key strategies of formative assessment

	Where the learner is going	Where the learner is right now	How to get there
Teacher	1 Clarify and share learning intentions and criteria for success	2 Questions	3 Feedback
Peer	Understand and share learning intentions and criteria for success	4 Peer Assessment	
Learner	Understand learning intentions and criteria for success	5 Self Assessment	

c) The Social Constructivist Assessment Process Model (SCAP Model)

The SCAP Model was developed by Chris Rust and his colleagues from the Oxford Centre for Staff and Learning Development (Rust *et al.*, 2005). They created this model as an initiative to lead assessment practises in classroom. Assessment practises still have not managed to be a distinct feature in

assisting learning although everyone is aware of their great influence on the learning process. The SCAP Model was developed as an assessment process based on a social constructivist approach. So, how does social constructivist approach view learning? In this paradigm, knowledge is seen as mutually-constructed or co-constructed with the combination of someone's prior knowledge with new knowledge influences by past and present social context (Brooks, 2002).

The overall model consists of two sub-models with parallel ongoing cycles of four different stages for each model. One model is applied for staff and the other one is for students. These two models are then combined to form a dynamic system of the SCAP Model. A very important requirement for a constructive assessment process is to have well-defined explicit assessment criteria. Next, students and teachers have to be actively-engaged with the criteria for the students to have a meaningful understanding of the assessment to improve their work, such as students marking their own work, peer marking, teachers having peer-discussions on criteria, marking process or moderation process. It is believed that it is better for the teachers to create and make decisions on the assessment criteria rather than be given finished criteria by the course leader; however, for students, it is the other way round. Following active-engagement with the criteria is active-engagement with the feedback by the students.

Constructive feedback lies at the heart of assessment processes, so active engagement in it is very important. Feedback does have an effect on students' learning as long as it is practised in a correct manner and students are actively engaged in it. Teachers have to train their students on how to deal with feedback to get the benefit out of it. But, how? Firstly, responding to teachers' feedback, students are encouraged to have peer discussions by referring to the checklists and come out with suggestions. Secondly is by providing comments without assigning students with any grades or marks. Similarly to this, it could also be a reflective self-assessment made by the students before the teachers return back the students' work. Thirdly is to give a general generic feedback to the whole class as soon as their work is completed and not to provide them with detailed individual feedback. Effective feedback gives way to staff to design assessment and develop explicit assessment criteria. Then, the criteria

would have to be clearly explained to students by the staff and the process continues in the form of a cycle.

The difference with this model compared to the previous models is that it clearly shows the interactive process that happens between teachers and students and should be observable through the interaction of the teachers and students at every stage of the assessment process. Also included is the marking and moderation process central to SBA. In conclusion, this dynamic model shows how assessment could be practised effectively based on the social constructivist approach where "common understanding being shaped and constantly evolving within a community of practice" (Rust *et al.,* 2005, p.236) by the students and the teachers as shown in Figure 3.4 and Figure 3.5.

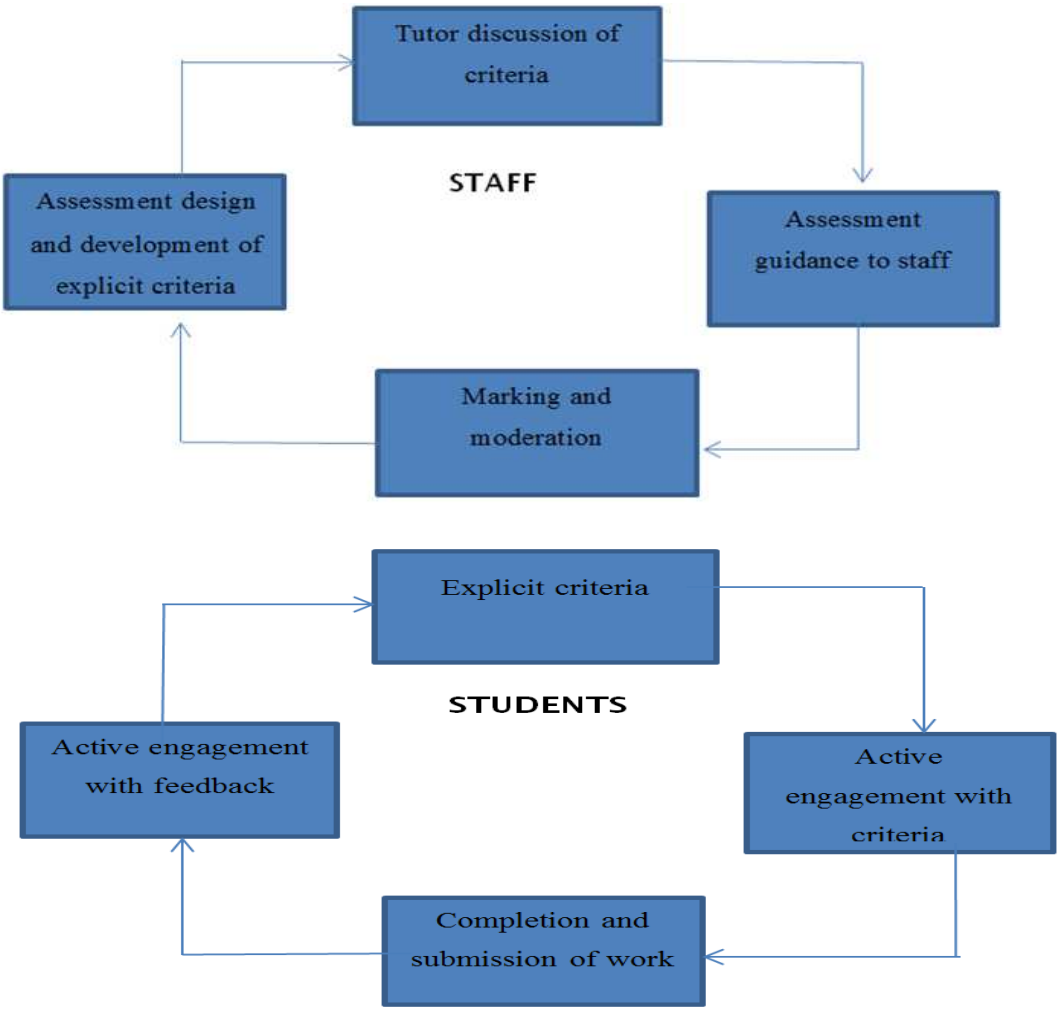


Figure 3.4: Social Constructivist assessment process: two parallel cycles

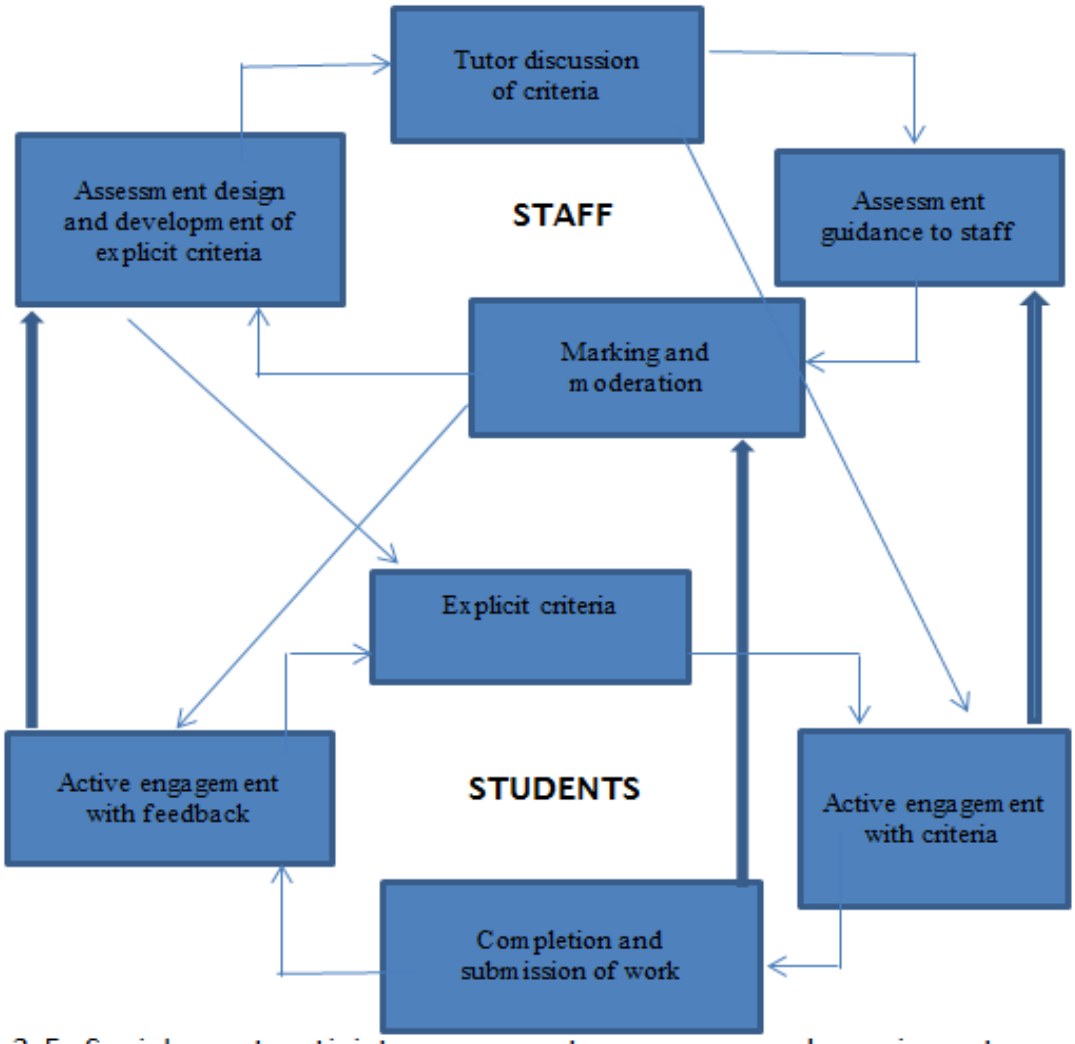


Figure 3.5: Social constructivist assessment process: one dynamic system

There are both advantages and the disadvantages of the three chosen models, however. The formative assessment model shows the dynamic relationship of all the important elements involved in the implementation of the formative assessment practises in a classroom. And it gives a clear picture on the implementation of formative assessment in one cycle so it enables teachers to evaluate their practises as to whether they have met the learning needs. Unfortunately, the model is of great value only if each individual really understands the separate attributes and is interested to see how these attributes work together in supporting an instructional process in a classroom. Next, the second model is the Logic Model. According to Leahy and Wiliam (2009), the advantage of this model of an educational intervention is that it follows a 'tight but loose' formulation with a framework that is rigid with its principles but quite flexible with the needs, resources, constraints and particularities from various school context as long as it agrees with its theory

of action. And, it allows a quick evaluation on the suitability of its context if the intervention were implemented (Wylie, 2008). The disadvantage of this model is that it lacks the details to see the impact on students' achievement. The idea from the logic model is quite different from the formative assessment model because this KLT programme is a professional development package for teachers to implement AfL but both models still focus on improving student learning.

The third model, the SCAP Model is similar to the formative assessment model in terms of presenting the operations involved in an assessment. The formative assessment model includes the integration of formative assessment elements into classroom practises whereas the SCAP Model includes formative assessment practises implemented in schools. The SCAP Model has the advantage that it applies to any assessment practise and there is a dynamic relationship between elements related to staff and students involving various processes related to assessment. However, the problem with this model is the assessment process is based solely on the social constructivist approach. Additionally, this approach alone invites several criticisms upon it, such as students might not be ready to plan the strategy, think and evaluate according to their own experience (Sang, 2006).

These three models are chosen because they focus on the assessment processes that fit the objectives of this study with the end result on student outcomes. Furthermore, these models are general and do not focus on certain subjects like the other models such as a model of formative assessment by Cowie and Bell (1999) which focuses on science education only or the formative and summative assessment model by Dr Herb Fynewever (2008) which focuses on chemical education. Similarly to the above models is a formative assessment cycle by Foster and Poppers (Fairfax and Country Public School, 2012) who created a model based on a case study in mathematics, so it is not suitable to simply be applied to other subjects. This is due to the fact that school- subject and the students' ages are the two factors that can influence teaching, and normally teachers teaching different subjects are engaged in different sets of professional development activities (Leahy and Wiliam, 2009).

All of the three models: the formative assessment model, the Logic Model and the SCAP Model are recent models of the assessment system. But these models still have their limitations. All of the models are developed based on research conducted in the western countries, so they might or might not be suitable to the Asian education system. The first two models only concentrate on formative assessment process and not on summative assessment. In this study, I will be referring most to the formative assessment model and the SCAP Model as the former model shows the integration of formative assessment, a main component of SBA into a well-executed cycle of assessment and instruction in a classroom and the latter model shows the major processes in assessment implementation in schools by teachers and students in such a dynamic way.

3.6 Programme Evaluation

This section on programme evaluation begins with different definitions on evaluation as these definitions reveal varying aspects of general study. It then moves on to consider the different types of evaluation that could be used when evaluating programme based on the role of evaluation itself. This is followed by the definitions on programme evaluation together with its purpose. Then, an overview of some of the assessment models is presented including the details of the CIPP model and its development as it is the conceptual framework of this study. The reason why it is considered as the most suitable model for this study is also given.

3.6.1 Concept of evaluation

Evaluations first became popular in the 1960s with evaluation of national and international programmes in various fields such as agriculture, crime prevention, health improvement or transport policy to determine the effectiveness of the different programmes (Silver, 2004). Evaluation also happens in the field of education when evaluation is conducted on the whole school reform initiatives or on specific curriculum changes. Evaluation is a complex process, but especially so in the field of educational programme, which is dependent on time ranges for evaluation to be conducted fully (Mohamad Najib, 2009). Furthermore, many factors in the teaching and

learning process are continuously being debated so the immaturity in the educational theories makes education a complex area for evaluation and yet at the same time evaluation seems to be almost part and parcel of daily life in education systems.

One of the pioneers in evaluation is Ralph Tyler. Tyler defined evaluation as a school-based process used to evaluate the extent to which curriculum objectives have been reached when he conducted a longitudinal study of schools in the 1930s (Gredler, 1996; Owston, 2008). Then, from the 1960s onwards, various roles of evaluation have been discovered. The theme has changed from being a solely school-based activity into numerous programmes which have been evaluated by various agencies and has also changed its focus from programme improvement to programme comparison (Gredler, 1996). Michael Scriven expanded Tyler's concept of evaluation by redirecting the attention of evaluators to consider not only the intended outcomes of a programme but also the unintended outcomes (Owston, 2008) as both outcomes are important considerations in evaluating programmes (Glatthorn *et al.,* 2012). For example, a mathematics programme may be conducted where the intended outcome is an improvement in the computational skill of students; however its unintended outcome maybe the diminishing interest in mathematics of students. If this is the case, then the programme is evaluated as not successful.

Evaluation is "the process of determining the merit, worth and value of something" (Donaldson and Scriven, 2008). This definition is similar to a definition from the online dictionary by Merriam Webster which defines evaluation as an act "to determine the significance, worth or condition of usually by careful appraisal and study" (Merriam-Webster, 2013). In Wikipedia, evaluation is a "systematic determination of a subject's merit, worth and significance using criteria governed by a set of standards" (Wikipedia, 2013) *w*hereas the free dictionary defines it as "to ascertain or fix the value or the worth of something" (The Free Dictionary, 2013). It shows that merit, worth and value cannot be separated from the term 'evaluation'. Scriven's idea of evaluation also rests on those three important terms or else evaluation does not make sense. So, from the definitions, it clearly shows that evaluation is a very general process that can be conducted by anyone without any special

purpose.

Stufflebeam created a new definition of evaluation when he realised the shortcomings of Tylers' definition as not fitting the needs of society, institutions and citizens, which led him to redefine evaluation as "the process of providing useful information for decision making" (Stufflebeam, 1983, p.120). This is similar to the definition by Gredler (1996, p.3) who stated evaluation as "the systematic collection of information to assist in decision making". Decision-makers or stake holders can come from various fields and organisations like commercial production, works of art, human services, individuals, facilities and equipment, business agencies, government agencies or schools (Gredler, 1996). Welch (2011) defined stakeholders as those people with a vested interest in evaluation. According to Welch, stakeholders can be grouped into three categories which are the 'upstream impactees' (tax payers, politicians, funders or policy makers); the 'midstream impactees' or 'primary stakeholders' (programme managers, staffs, teachers or administrators); and the 'downstream impactees' (people who receive the products like textbooks or curriculum or those who receive services). Decision-makers or stakeholders will differ depending on the institution, so in a university setting they could be faculty members, a curriculum committee, a technology roundtable, a faculty council or senior academic administrators. In the school setting, it could be a combination of parents, teachers, a school council and the district superintendent (Owston, 2008). Others could be clients or programme participants, journalists, students or even members of the general public (Patton, 2000).

Given the more comprehensive definition by Stufflebeam (1971a), evaluation is a systematic and continuing process of delineating the information to be collected, obtaining the information and providing them to the decision-makers. He believed that when the decision-makers like school administrators, project directors and school staff are provided with information from the programme evaluators, they are able to make reforms in education (Stufflebeam, 1983). This perception of serving the decision- makers with information is accepted by many researchers over the years (Cronbach, 1963; Alkin, 1969; Gredler, 1996). The process of providing useful information involves three steps (delineating, obtaining and providing) which in turn serve

four kinds of decisions (planning, structuring, implementing and recycling) in evaluation decisions. The steps and decisions regarding evaluation by Stufflebeam has been detailed in Chapter 1 on the CIPP model.

Stufflebeam and Shinkfield defined evaluation as "the process of delineating, obtaining and providing descriptive and judgemental information about the worth and merit of some object's goals, design, implementation and impacts in order to guide decision making, serve needs for accountability, and promote understanding of the involved phenomena" (Stufflebeam and Shinkfield, 1985, p.159). This definition is chosen because it presents evaluation as a continuous and systematic process that supports the process of decision-making, accountability and promoting an increased understanding of phenomenon. And at the same time, it represents a broad picture of evaluation which includes formative and summative evaluation aspects along with its stated goals.

3.6.2 Types of evaluation

There are different types of evaluation depending on the objects that are being evaluated or the role of evaluation. As already discussed, when referring to the role of evaluation, Scriven (1967) asserted that there were two types of evaluation; formative and summative. If evaluation is conducted in order to improve the programme (Scriven, 1967) or to develop current activities (Brinkerhoff *et al*., 1983) while the programme is still under development (Owston, 2008), that is a *formative evaluation.* Alternatively, if evaluation is conducted to look at the effectiveness of a programme or outcome (Scriven, 1967), for accountability, certification, selection or continuing a programme (Brinkerhoff *et al.,* 1983) of an already completed programme (Owston, 2008), then it is a *summative evaluation*. Similarly, proving the success or failure of a programme is also considered a summative evaluation (McNamara, 2002). But, both forms of evaluations are needed in the development of a product or service (Stufflebeam and Shinkfield, 2007, p.24). By means of summative evaluation, it might also be conducted to assess the quality, cost, utility and competitiveness of products and services or to provide evidence to tax payers (Stufflebeam, 2003b).

Alternatively, formative evaluation also known as a 'process-oriented evaluation' allows decision to be made during the implementation of a programme whereas summative evaluation, also known as a 'product-oriented evaluation', allows decision to be made about the effectiveness of a programme (Mitchem *et al.,* 2003). So, it can be concluded that both forms of evaluations could be used for the purpose of decision-making process. Summative evaluation also allows programme personnel or consumers to make decision whether to continue the programme or not by providing enough information (Worthen *et al.,* 1997). And both types of evaluation could also contribute to making programme improvements. Worthen *et al.* (1997) stated that formative evaluation provides feedback to improve programmes, whereas summative evaluation could play the same role when it provides feedback in the evaluation of teachers, courses or teachings at the end of the programme.

Examples of formative evaluation are the evaluation conducted during a curriculum development improvement, the evaluation on teacher self-improvement or the evaluation of the 'Eight-Year Study' by Ralph Tyler in 1930s (Gredler, 1996). Examples of summative evaluation are the evaluation of the demonstration projects for dropouts at high schools or the evaluation on kindergarten retention. Research has shown that recently, most programme evaluations are going through a shift towards being summatively evaluated rather than being formatively evaluated (Hogan, 2007) although Stufflebeam and Shinkfield (2007) believed that summative evaluation does limit the process of development as it judges programmes. Summative evaluation is sometimes seen as a form of evaluation that is conducted to serve purposes other than development purposes and is only appropriate for a well-defined intervention (Gredler, 1996). However, formative evaluation reports could contribute to summative evaluation as formative is normally conducted prior to summative evaluation (Stufflebeam, 2003).

Sometimes a programme can serve both purposes (Muijs *et al.,* 2004) such as a programme evaluating the impact of the continuing professional development of teachers on student learning. Its formative evaluation is trying to answer a question such as 'How can the programme be improved?' and its summative evaluation is focusing on answering 'Does the programme improve outcomes?'. In conclusion, any intention to come to an overall judgment in

making decisions of a completed programme is considered as a summative evaluation and those related to providing the information while the programme is still running in order to improve the programme is considered as a formative evaluation. Table 3.4 lists out the main differences between formative and summative evaluations (Welch, 2011).

Table 3.4: Main differences between formative and summative evaluation

FORMATIVE EVALUATION	SUMMATIVE EVALUATION
Primary purpose is improvement	Primary purpose is accountability
Used by midstream stakeholders	Used by upstream stakeholders
Summary of what is occurring	Summary of what has occurred
Decisions as a result of recommendations based on what is occurring	Decisions as a result of what has been completed
'Chef tastes the soup'	*'Customer tastes the soup'*

3.6.3 Programme evaluation and its purpose

Programme evaluation can vary, from assessing programme, project or system of services (Stufflebeam, 2001). Stufflebeam believes that evaluation of programme should be differentiated from evaluation of students, personnel, material or institutions. Programme evaluation has become a dominant feature of educational reform and has gained recognition internationally. It was first developed nearly 200 years ago but has become an established field for only 20 years ago. Examples of programme evaluations are activities involved in documenting the impact of alternative high-schools on at-risk students or strategies used by science students working through a computer simulation (Gredler, 1996) or assessing ongoing program like school curricula, assessing time-bounded projects like development of a new instrument or to assess state systems of services (Stufflebeam, 2001).

Patton sees programme evaluation as "the systematic collection of information about the activities, characteristics, and outcomes of programmes to make judgments about the programme, improve programme effectiveness, and inform decisions about future programming" (Patton, 2000, p. 426). In this study, I will be using Patton's definition because it is suitable for this study. As a researcher, I would systematically collect information on the SBA system implementation activities, characteristics and outcomes at the end of the

programme to look at the effectiveness of the programme and then provide information to decision makers to enable them to make suitable decisions. Moreover, it is believed that this definition reflects the CIPP approach (Robinson, 2002), the approach that I use as a framework in this study. Actually, Patton's definition is based on utilisation-focused evaluation and is grouped together with the CIPP model by Stufflebeam under the same evaluation approach which is the decision-making approach.

In practice, when evaluating programmes, there exist both benefits and challenges. Some of the benefits are that programme evaluation supports accountability, builds capacity, supports communication, increases understanding, helps in decision-making or increases improvement in processes, activities or outcomes (Kahan, 2008). To the contrary, evaluation also faces some challenges in terms of context like having pressure of conducting only one form of evaluation approach or insufficient resources in funding, time, energy, personnel or skills. Challenges can also occur during the collection and the analysis of data. Imbalanced treatment in the evaluation procedures like high preference on outcome evaluation over process evaluation or ignorance towards long-term outcomes could also pose problems. In fact, the task of evaluation itself is complex and poses its own challenges (Eseryel, 2002). Even worse, when the evaluation field itself is still being debated as whether it acts as a profession with developed competencies for practice as in, for example, the health care, teaching or counselling professions (Stevahn *et al*., 2005). However, be it benefits or challenges, both are important for an evaluation framework to successfully guide an evaluation when it comes to actual practice (Kahan, 2008).

In general, educational programmes are conducted to increase students' learning and achievement and can be understood as a set of activities or projects implemented to solve educational problems and challenges (Mohamad Najib, 2009). More than that, the implementation of an educational programme is also meant to make a change in students' knowledge, skills or attitude, change in education structure or education leaders' performance and many more (Frye and Hemmer, 2012). Generally, evaluation programme in education involves two main aspects of education which are the teaching process or the learning process. Examples of aspects related to the teaching process are the

teachers, the school facilities or the technology whereas those aspects related to the learning process are the students or their experiences. An example of a list of educational programmes in Malaysia is shown in Table 3.5 (Mohamad Najib, 2009).

Table 3.5: Educational programme

No		Specific programmes or projects
1	Students' performance and development	Tuition, motivation, extra classes, co-curriculum, moral values, camping
2	Teachers' performance	Seminar, workshop, courses, teachers' recruitment policy
3	Curriculum design and process	New subjects, time table, pedagogic strategy, ICT usage, examination and assessment system

Evaluation of an educational programme is one of the most important investments that can be implemented in education (Wholey, 1979). In some countries, educational programmes were evaluated to ensure funds were benefitted for the disadvantaged children, providing information for parents and communities or as a motivational source for schools to be upgraded (Fleischman and Williams, 1996). Over time, programmes have been evaluated for more comprehensive reasons.

Generally, the common role of programme evaluation is to look at the attainment of the programme's goals and objectives (Owston, 2008) and then make a decision whether to continue, improve, expand or reduce the programme implementation (Rossi *et al.,* 2004). Programme improvements could be undertaken because evaluation identifies the strengths and weaknesses of a programme (McNamara, 2002). Evaluation also provides information on the credibility of a programme before the programme can be implemented (Mohamad Najib, 2009). In addition, evaluation could also be a mean of determining how the various dimensions of a programme or project relate to each other, which then might assist evaluators to find ways to increase program effectiveness (NOAA, 2009). This is related to this study purpose which looks at the interrelationship of dimensions of the system in order to find ways to increase the effectiveness of the system. Moreover, evaluation helps in verification processes to see whether the programme works out as planned. It also helps in giving ideas to the management group on the

whole idea of any programme implementation including its goal, ways to achieve it and assessing the result of the programme (McNamara, 2002). It could also be helpful for the accreditation of the programme or to look at the overall impact of the programme (Owston, 2008). Moreover, the utility of a new programme along with the effectiveness of the management and administration or the programme sponsor's accountability requirements could be also assessed (Rossi *et al.,* 2004). Not only directly related to the programme, it could also be used in developing theory on intervention or to fulfil the needs of the audience (Owston, 2008). In conclusion, all the purposes of programme evaluation are related directly or indirectly to the merit, worth or value of a programme which is the reason why evaluation is conducted in the first place.

3.6.4 Models of programme evaluation

In order to evaluate the programme in a more complete and meaningful manner, it is advisable to follow certain models. As stated by Razali (1987) and discussed above, a model is a conceptual picture that shows the interrelationships between various elements involved in any activities. Experts in evaluation categorise different methods in evaluation based on the different approaches, strategies or models. According to Armstrong and Ogren (1986), the classification of evaluation models is based on the major evaluation methodology, target of evaluation, purpose of evaluation and the responsibility and accountability of evaluation. There are a wide variety of evaluation models. Some examples are the Provus Discrepancy Model, the CIPP Model, Stake's countenance approach, Scriven's concept of goal-free evaluation, judiciary models, Eisner's connoisseurship model, Stake's responsive evaluation perspective, illuminative evaluation (Gredler, 1996), Patton's utilisation-focused evaluation, ethnographic evaluation or Kirkpatrick four-level model (Owston, 2008). Even Zhang *et al.* (2011) state that there are twenty-six evaluation approaches categorised into five groups such as pseudo-evaluations, quasi-evaluation studies, improvement and accountability-oriented evaluation, social agenda and advocacy and eclectic evaluation that can be used to evaluate projects. Similar to Zhang, Stufflebeam (2001) suggested twenty-two evaluation approaches grouped into four main categories of pseudo-evaluations, questions and/or methods-oriented, improvement or accountability and social

agenda or advocacy. Sometimes, these models are grouped differently depending on the authors. Although there are many evaluation models with different purposes, they might have one factor in common. All evaluation models to certain extent should be able to evaluate in a systematic and reliable way to produce a solution.

For the purpose of this study, I will focus only on those approaches or models that are mostly applicable to educational programmes. These are i) Tyler's concept; ii) naturalistic evaluation approaches; iii) Kirkpatrick's Four Level; and iv) decision-making approaches, as shown in Table 3.6 (Owston, 2008; Zinovieff, 2008).

Table 3.6: Evaluation Approaches, Models and their Focus

Evaluation Approach	Orientation	Evaluation Model	Focus
a) Tyler's concept	Goal-based or objective-oriented	Tyler's model	Ways to determine whether a programme met its stated goals
b) Naturalistic	Participation-oriented	Stake's responsive evaluation	Evaluating programme based on the qualitative uniqueness of it
c) Kirkpatrick	Goal based or objective-oriented	Kirkpatrick's 4-Level model	Evaluating programme effectiveness related to corporate training
d) Decision-making	Management-oriented	CIPP Model; Patton's utilization-focused evaluation	Ways to provide meaningful and useful information for decision alternatives

What follows is a brief summary of the four models:

- Tyler's concept of programme evaluation was based on the belief that evaluation of a curriculum or a programme involves the activities used to determine whether a programme has met its stated objectives, is effective enough or needs improvement (Gredler, 1996).

- The naturalistic evaluation approach, also known as qualitative approach, believes that in evaluating a programme, evaluators should focus more on producing a well-rounded programme by using qualitative methodologies and full involvement of the evaluators at the actual site of the programme (Owston, 2008).
- Kirkpatrick's Four Levels is concerned with the four levels of assessment of reaction, learning, behaviour and results in evaluating the programme effectiveness related to corporate training.
- The decision-making evaluation approach believes that evaluators should provide meaningful and useful information for the educational decision-makers when evaluating programmes (Gredler, 1996).

The 'decision-making evaluation approach' is chosen for this study because this approach focuses on the decision-making process or in other words, it focuses on how results could be used towards decision-making (Owston, 2008). It serves the evaluation purposes that best suit this study which is to allow decisions to be made on the effectiveness of the programme or decisions whether to continue the programme or not. Almost all evaluation models are related to decision-making to certain extents but most seem not to be fully decision-oriented, rather tend to be goal oriented, goal-free oriented, transactional or adversary evaluation (Issac and Michael, 1982).

As for the other three approaches (by Tyler, Naturalistic and Kirkpatric respectively), they are not chosen for this study for several reasons. Tyler's concept of evaluation is not used in a broader context and perception as it focuses only on the stated and intended goals when evaluating a programme. This rigid concept was debated by Scriven after it had been implemented vigorously in the years from the 1930s to the 1950s. Scriven argued that when evaluating assessments, one has to assess the actual outcomes of the programme, both intended and the unintended. Furthermore, the naturalistic evaluation approach uses qualitative methodology so anything related to knowledge claims, strategies or methods used, such as data collection, data analysis and interpretation, should follow this approach. But, this study is using a full quantitative approach, so the natural evaluation approach is not suitable. Lastly is the Kirkpatrick model. This model is not chosen because it

focuses on the evaluation of corporate training programmes and does not concern with decisions made by decision-makers (Owston, 2008).

a) The Tyler evaluation model

Tyler evaluation or 'goal-based' evaluation model was developed by Ralph W. Tyler in the late 1940s. Tyler's model focused on the 'goals' such as the programme objectives, performance targets or expected outcomes (Kahan, 2008). It is also known as an 'objective-oriented' approach for the way in which it focuses on objectives rather than values (Isaac and Michael, 1982). The constructs involved in this model are the statements on the 'goals' in behavioural terms taking into account factors like pupil's entry behaviour, analysis of culture, school philosophy, learning theories or new developments in teaching. Evaluation is conducted based on pre-post measurement of student performance. The steps in the Tylerian Evaluation Approach are as follows (Lee, 2008):

1. Establish broad goals or objectives
2. Classify the goals or objectives.
3. Define objectives in behavioural-terms.
4. Find situations in which achievement or objectives can be shown.
5. Develop or select measurement techniques.
6. Collect performance data.
7. Compare performance data with behavioural-stated objectives.

The strengths of this model are that it is easy for evaluators to determine whether the goals have been met (Kahan, 2008) and it is easy to design evaluative studies using this model (Isaac and Michael, 1982). It is also scientifically acceptable, logic, easily adapted by the evaluators and has had a great influence on the evaluation approach that comes after it (Lee, 2008). On the contrary, its weakness is that it tends to focus on information on terminal rather than information on pre-programme or the on-going programme. Evaluation also focuses more on objectives rather than its value (Isaac and Michael, 1982). It is the fact that this approach does not consider the validity of the chosen 'goals' (Kahan, 2008) and, even if evaluation results show that they are aligned with predetermined goals that does not necessarily mean that they are caused by the programme implementation but could be due to other

factors. This phenomenon of ignoring processes seems to limit the approach in evaluating.

b) Stake's responsive evaluation model

Being a naturalistic approach, Stake's responsive evaluation model gives evaluators more freedom to choose which methodology to be used in the programme evaluation (Owston, 2008). It is 'natural' in the sense that evaluators act naturally like most of the people who evaluate things by observing and reacting (Stake, 1973). Stake proposes an approach that responds more to the needs of the audiences as he says "an evaluation probably will not be useful if the evaluator does not know the interests and language of his audiences" (Stake, 1973, p.4). That is why it is also known as 'client-centred study' (Stufflebeam, 2001) where evaluators have to work hand in hand with the clients and act as a counsellor or advisor to the clients in improving programmes during the evaluation process. Furthermore, being client-centred, it also rejects objectivist evaluations and moves more towards accepting open-ended and emergent designs of study, thereby producing a narrative description of study in order to create a rich set of information of the programme.

Evaluators under this type of assessment model should attend actual programme activities and not to focus only on the programme intentions. Evaluators should also present multiple perspectives from the audiences on the success or failure of the programme and should be focusing more on the methodology of anthropology and journalism in order to gain more information on the programme (Owston, 2008). In evaluating programmes based on the responsive evaluation approach, Stake lists the role of evaluators as follows (Stake, 1973, p.5):

1. To make plan of observations and negotiations.
2. To make an arrangement for the various observers of the programme.
3. To make a brief narratives, portrayals, product display or graphs.
4. To find out what values most to the clients.
5. To collect descriptive and judgmental data from various individuals.
6. To check the records quality.

7. To get programme personnel to react to the accuracy of his portrayals.
8. To get authority figures to react to the importance of findings.
9. To get audience members to react to the relevancy of findings.
10. To prepare for final written report if needed.

The strength of this model is that it provides a systematic method for the data matrices which contain descriptive data with intentions and observations and also judgment data (Issac and Michael, 1982). But still, it might not have enough methodology to gain information and it stirs up value conflicts. However, methods used still depend on the evaluation purposes (Stufflebeam, 2001). Examples of the methods used are the case study, expressive objectives, purposive sampling, observation, adversary reports, story-telling, socio-drama or narrative reports.

c) Kirkpatrick's 4-level model

Kirkpatrick's evaluation approach was developed by Donald J. Kirkpatrick in 1959 and takes an objective or goal-based approach. It favours neither the decision-making nor naturalistic approaches (Owston, 2008). Instead it consists of four levels of training evaluations which are reaction, learning, behaviour and results in a sequential manner (Kahan, 2008), with a fifth level added later to measure return on investment (McNamara *et al.,* 2010). The first level, the 'reaction level', is the easiest to conduct while the highest level is the hardest (Zinovieff, 2008). In general, this model is used to evaluate the quality, efficiency and effectiveness of the programme. The first level, reaction level measures the 'affection factor' of the participants towards the programme, learning level assesses knowledge, skills and attitudes of participants, behaviour level measures the change in participants' behaviour and results are dealing with the impact of the programme. In evaluating level 1, a positive reaction means that the participants are happy and satisfied with the training programme and are more likely to use the skills and knowledge they acquire (Mohamad and Alias, 2012) and vice versa for a negative reaction. Kirkpatrick and Kirkpatrick (2006) believed that the participants might be motivated to learn if they give a positive reaction. In level 2, evaluation determines how much knowledge, skills and attitudes that have been gained by the participants

throughout the training. The assessors can use pre-tests and post-tests to measure knowledge, performance test to measure skills and attitude surveys to measure attitude changed. In order to assess level 3, participants need to be given certain range of time for pre-post performance appraisal so that they have an opportunity to apply what they have learned (McNamara, 2010). An example of an evaluation made at the result level in the field of higher education is to assess students' performance at work following graduation, thus permitting an evaluation of the impact of the course or programme on the job performance of students (Owston, 2008).

Normally, levels 1 and 2 are considered as formative evaluation, while levels 3 and 4 are summative (Zinovieff, 2008). The first two levels are easy to assess because they involve the extent to which participants' satisfied and have changed their level of attitude, knowledge and skills after attending the programme. But, the latter levels are more difficult to assess as they involve assessing changes in participants' behaviour and also changes to the organisation as a result of the participants' attending the programme (Owston, 2008). Research in the USA has found that more than seventy-five per cent of organisations are using level 1 only in evaluating programmes, forty-four per cent are using level 2, twenty-one per cent are using level 3 and eleven per cent are using level 4 (Zinovieff, 2008). This figure is alarming because organisational performance should not be measured solely on trainees' perceptions. Hence, there is no proven causal link between level 1 and 2 which means an increase in how participants feel about the training does not necessarily cause an increase in their learning level. The same applies to levels 2 and 3. However, some trainers believe that level 3 evaluations does have some causal link to level 4 evaluations but to reach the upper level needs a lot of money. This condition seems to limit the rewards level 4 evaluations have to offer.

This model is considered good for its straight-forward procedures when it comes to implementation but it does not explore the 'why' or 'how' of the results (Kahan, 2008). Evaluators using Kirkpatrick's model have difficulties in determining the suitable evaluation methods and sometimes they end up conducting only the first two levels of evaluation because complexity increases as level increases (Eseryel, 2002). A suggestion by McNamara *et al.* (2010)

which is also consistent with Kahan and Eseryel's viewpoints is that results produced using this model of evaluation might not be significant enough although it represents evaluation as a straight forward process. Moreover, this model does not consider the intervening variables that can affect learning such as learner motivation or entry levels of trainers' skills and knowledge (Frye and Hemmer, 2012).

Table 3.7 simplifies the idea of Kirkpatrick model (Zinovieff, 2008, p.13). The questions posed by Kirkpatric allow evaluators to fit them in with their study interest so they do not have to do all levels if it is not related.

Table 3.7: Level and its measurement focus in Kirkpatrick's model

Level	Measurement focus	Questions addressed
Reaction	Trainees' perceptions	What did trainees think of this training?
Learning	Knowledge or skills gained	Was there an increase in knowledge or skill level?
Behaviour	Worksite implementation	Is new knowledge or skill being used on the job?
Results	Impact on organization	What effect did the training have on the organization?

d) Pattons' utilisation-focused evaluation model

There are two models in the decision-making evaluation approach; the CIPP Model and Patton's utilisation-focused model. Michael Quinn Patton developed the utilisation-focused evaluation model in the 1980s. By the name itself it means that evaluation is based on utility and actual use or in short, it is based on clients' requirements. It is a model of evaluation which focuses on "intended use by intended users" (Patton, 2000, p.425) which means focus is made on 'who' is going to use the evaluation results and 'how' they are going to use them (Kahan, 2008). Intended users or primary intended users are the specific groups from the list of potential stake holders related to the programme whereas the intended use is the information needed by the primary intended users (Patton, 2000). The primary uses related to evaluation findings can be to judge merit or worth of a programme as in summative

evaluation, to improve programme as in formative evaluation or to generate knowledge. And, the primary uses related to the evaluation process can be to increase shared understandings of participants, to reinforce interventions, to give full support to participants' engagement or to develop programmes. However, it is worth bearing in mind that uses are not fixed and they change by time as the programme develops. When conducting evaluations, evaluators should greatly consider the process and the design of evaluations and look at how they affect the way the evaluation process and findings are applied by the users. The use-oriented definition held by this model differentiates it from other models. Patton believed that in order to produce 'used' programme evaluation studies, decision-makers have to work together with the evaluators concerning research design including the data collection methods and the data analysis tools, together with the analysis or the dissemination of the data (Owston, 2008). Being flexible and having high probability of evaluation results to be reviewed seem to be the strength of this model (Kahan, 2008) although its main weakness stems from the fact that it ignores the role of groups other than the intended users.

Both models, the CIPP Model and Patton's model focus on how the evaluation results from the evaluation procedures could be used for the decision-making (Owston, 2008). Patton believed that evaluators have to play their role in working together with the intended users in making judgments and decisions without neglecting the evaluation's accuracy, feasibility and propriety factors (Patton, 2000). Since Patton's model does not propose exact models, exact methods or theory in evaluating programmes, evaluation should facilitate the intended users to decide on the most suitable evaluation purposes such as formative, summative or developmental, any method such as quantitative, qualitative or mixed methods, any research design such as naturalistic or experimental design or focusing on any possibilities such as focusing on processes, outcomes, impacts, costs and cost-benefit. Therefore, evaluators can act as facilitators in helping intended users to determine what kind of evaluation that they need.

However, Patton disregarded the idea of categorising his model into the decision-making approaches as he believed that the model not only serves the decision-makers but could also serve several other purposes intended by the

users. Instead, he categorised it as a 'user-focused' evaluation rather than a decision-making evaluation approach (Patton, 2000). But still, he believed that both models are "very concrete and focused, and also considered utilitarian" (p. 428). Furthermore, both models also give freedom to the evaluators in determining the value of a programme and suggest the best course of actions (Owston, 2008).

The difference between these two models is on the purpose of evaluation. If the purpose of evaluation is to look at the attainment of the goals and objectives of the programme, to improve the programme or to look at the overall impact of the programme then the CIPP Model is used whereas Patton's model is suitable if the purpose is to meet the needs of various audiences (Owston, 2008).

3.6.5 The CIPP Model

The CIPP Model was developed by Daniel Stufflebeam and his colleagues as a guidance to evaluate programmes, projects, personnel, products, institutions or systems from various disciplines such as education field, housing and community development, transportation safety and military personnel review systems (Stufflebeam, 2003a). It has been used to evaluate various educational programmes such as those related to science and mathematics education, rural education, educational research and development, school improvement, professional development schools and many more (Stufflebeam, 2002). It could be implemented by internal evaluations, self-evaluations or external evaluations. It is a model that can serve both formative and summative purposes of evaluation (Stufflebeam, 2003a). Stufflebeam believed that the CIPP model places priority for evaluation to function as a way to improve the programme rather than to prove. But still, proving can be conducted using summative evaluation. Recently, some people believe that the first three elements of the model (context, input and process evaluations) are more suitable for formative evaluations or improvement-focused and the last element (product evaluation) is suitable for summative evaluation (Frye and Hemmer, 2012). Stufflebeam also believed that by using his model, formative evaluation focuses on decision-making while summative evaluation focuses on accountability of the programme (Owston, 2008).

This model is based on the management-oriented approach, an approach that allows evaluators to provide enough information for the organisational leaders following information needed by the managerial decision-makers (Hogan, 2007). Being a decision or accountability-oriented evaluation, the CIPP Model is able to facilitate decision-making and quality assurance (Stufflebeam and Shienkfield, 2007). All these are done in order to fulfil the needs of managerial decision-makers and that is why the model itself comes under a management-oriented evaluation approach (Hogan, 2007) or decision-making evaluation approach (Owston, 2008). The four components of the model are the context, input, process and product evaluation. The relationship between the four components of evaluation and types of decision has been explained in detail in Chapter 1.

Context evaluation focuses on planning decision. Examples of specific evaluation questions that could be asked from context evaluation dimension are "What is the aim of the programme?", "What is the context of the programme and how will it match with the target population?" or "Who are the providers and [what is] their background information?" (Clinton, 2001). Even, during the evaluation process itself, evaluators can deal with the decision makers to identify emerging problems that might require changes in the programme objectives (Stufflebeam, 1971a). Data collection methods that could be used are surveys, document reviews, interviews, secondary data analyses, diagnostic tests, system analyses, hearings or the 'Delphi technique' (Dalkey and Helmer, 1963), and focus groups, website, journals or standardised testing (Clinton, 2001). According to Clinton (2001), the most used data collection method for context evaluation is existing documentation.

Input evaluation focuses on the personnel, resources and procedures in achieving the programme objectives or the desired results (Stufflebeam, 1971a). Not only does it involve assessing current system capabilities, Zhang *et al. (*2011) believed that input evaluation also involves examining potentially relevant approaches and then recommending alternative approaches. It serves 'structuring decisions' because the alternative strategies are structured to meet the needs of the given objectives (Stufflebeam, 1971a). Examples of input evaluation components are teachers' skill and knowledge, equipment,

workshop facilities, time table, funds allocation or the role of principal. Examples of specific evaluation questions are "What is the number of sessions?", "What input was received from MOE?" or "What professional development did the providers receive?" (Clinton, 2001). Data collection can come from surveys, interviews, websites, journals or existing documentation, although the majority of data collected comes from existing documentation (Clinton, 2001).

Process evaluation focuses on the process implemented to achieve programme's objectives. The information is needed occasionally to monitor the programme implementation (Zhang *et al.,* 2011). During the process evaluation, other than monitoring challenges, it also identifies programme adjustments, gets additional information for changes, documents the process and runs regular monitoring on the activities (Zhang *et al.,* 2011). It serves an 'implementing decision' as it provides information during the implementation stages of a programme or a complete description of the actual programme activities. This is to help indecision-making, helping to understand how the implementation could be strengthened, how programme replication could be assisted or the reasons why programme objectives that are not being achieved could be determined (Stufflebeam, 1971a). Examples of the process evaluation component of an evaluation of subject's effectiveness are teaching and learning methods practised by the teachers, methods used by the teachers to assess course work or courses attended by the teachers (Clinton, 2001). Methods such as questionnaires, on-site observation, interviews, rating scales, photographic records, records analysis, case studies of participants, focus groups, self-reflections with staff or tracking expenditures (Zhang *et al.,* 2011), survey, existing documentation, website or focus group could be used. Clinton (2001) believed that interviewing is the most suitable mode of data collection method for process evaluation.

Product evaluation assesses programme outcomes hence it is sometimes called an 'outcome evaluation'. After evaluation has gone through the three phases (context, input and process evaluation), at the end of it, of course it is important to know the extent to which the participants' needs are met by asking "Did the programme succeed?" (Stufflebeam, 2003a).

For product evaluation component, it is divided into four subparts of evaluations of impact, effectiveness, sustainability and transportability evaluations (Stufflebeam, 2003b). Impact evaluation assesses whether the programme reaches the target audience. These findings are very important for the stakeholders as an assurance as to whether the programme reaches intended beneficiaries or as a judgment on whether the programme serves the right beneficiaries and addresses the important community needs. Effectiveness evaluation assesses the effectiveness of the programme outcomes in terms of their quality and significance by looking at the positive and negative outcomes. The findings could be used by the stakeholders as judgments on the success or failure and the significance of the programme and then to come up with accountability reports of the outcomes. As Stufflebeam put it "Did it succeed in meeting the beneficiaries' needs?" or "Were their targeted needs met?" (Stufflebeam, 2003a).

Next are sustainability and transportability aspects. The findings from the sustainability aspect could be used to make decisions on whether to continue the programme or not. The findings from transportability aspects help in determining sustainability of the audiences, the information to be disseminated and the programme to be applied at other places. Respectively, it is like asking "Did it succeed in sustaining the beneficiaries' gains?" or "Were the gains for beneficiaries sustained?" (Stufflebeam, 2003a) and "Did it succeed in proving a transportable and adaptable gains for the beneficiaries and later is capable to be applied in other settings?" or "Did the processes that produced the gains prove transportable and adaptable for effective use elsewhere?" (Stufflebeam, 2003a). For example, the transportability aspect of the Teachers' Evaluation Programme has been found significant when other evaluators have shown interest in applying a similar methodology to their programme (Muhammad, 2008).

Examples of evaluation questions concerning product evaluation could be "Are the intended outcomes achieved?" or "What is the impact on the achievement outcomes?" (Clinton, 2001). Product evaluation could be conducted using techniques like rating scales, case studies, logs and diaries of outcomes, hearings, focus groups, document analysis, photographic records, achievement tests, trend analysis of longitudinal data, longitudinal or cross-sectional cohort

comparisons, and comparison of project costs or outcomes (Zhang *et al.*, 2011). Table 3.8 lists out the slight difference of the questions addressed dealing with formative and summative evaluations using the CIPP Model.

Table 3.8: Formative and summative evaluations

	Questions addressed	
Evaluation dimension	Formative evaluation report	Summative evaluation report
Context	What needs to be done?	Were important needs addressed?
Input	How should it be done?	Was the effort guided by a defensible plan?
Process	Is it being done?	Was the service design executed competently and modified as needed?
Product	Is it succeeding?	Did the effort succeed?

Hence, this model could be used for both purposes, formative or summative evaluation (Stufflebeam, 2003b). Although both purposes are sometimes interchangeable, the four dimensions of the CIPP Model could serve its purpose to a certain extent. The difference in actions taken following their roles in evaluation in all four dimensions of the model is shown in Table 3.9. It clearly shows that both forms of evaluation can be conducted using the CIPP Model. The main aim in formative evaluation is to improve programme. Since formative evaluation is meant to be conducted during the implementation process, all actions are in the forms of guidance for the evaluators to act upon at various stages based on certain conditions to help them in decision making and quality assurance. As for summative evaluation, it is the action of comparing to the targeted needs since it is conducted at the end of a completed programme to sum up the programme's merit, worth, probity and significance. Examples of summative evaluations are assessment on the quality, cost, utility and competitiveness of products and services. Actually, a formative evaluation report can be helpful in providing information for summative evaluation for the internal or external evaluators if it is effectively conducted and reported.

Table 3.9: The relationship of evaluation roles to context, input, process and product evaluations

Evaluation Roles	Context	Input	Process	Product
Formative Evaluation	Guide to identify needs or ranking of goals	Guide to choose strategy and examine work plans	Guide to implement plans	Guide to continue or terminate project
Summative Evaluation	Compare goals and priorities to assessed needs, problems, assets and opportunities	Compare strategy, design and budget to targeted needs	Compare designs and actual processes and costs	Compare outcomes and side effects to targeted needs

In conclusion, altogether there are four main evaluation approaches that have been discussed. Table 3.10 shows the difference of Tyler's model, Stake's model, the Kirkpatrick's model and the CIPP Model, a model from each evaluation approach (Isaac and Michael, 1982).

Table 3.10: Summary of the evaluation models

Evaluation Approach	Goal-based	Naturalistic	Kirkpatrick	Decision-making
Evaluation Model	Tyler	Stake	Kirkpatrick 4-level	CIPP
Definition	Comparing student performance with behaviourally stated objectives	Describing and judging an educational programme	Analysing the effectiveness and impact of the training	Defining, obtaining and using information for decision-making

Purpose	To determine the extent to which purposes of a learning activity are actually being realized	To describe and judge educational programmes based on a formal inquiry process	To evaluate the quality, efficiency and effectiveness of the corporate training programme	To provide relevant information to decision-makers
Key emphasis	Specification of objectives and measuring learning outcomes	Collection of descriptive and judgmental data from various audiences	It was created specifically for training interventions	Evaluation reports used for decision-making
Role of evaluator	Curriculum specialists who evaluates as part of curriculum development and assessment	Specialists concerned with collecting, processing and interpreting descriptive and judgmental data	Trainers do any level of evaluation according to the organization's selection	Specialists who provides evaluation information to decision-makers
Relationship to decision-making	Evaluation implies attainment of behavioural objectives stated at the beginning of the course	Descriptive and judgmental data result in reports to various audiences	Does not favour decision-making approach	Evaluation provides information for use in decision-making
Types of	Pre-post	Formal and	Reaction,	Context,

evaluation	measurement of performance	informal	learning, behaviour and results	input, process and product

3.6.6 Reason for choosing the CIPP Model as a Framework of this study

Generally, there is no one best model of evaluation that fits certain programmes. Research by Armstrong and Francis (2003) has shown that crime prevention programme evaluators have chosen the most common and the highest rated models as their evaluation models. It is suggested that in choosing the evaluation models, one has to consider whether the approach is plausible, practical or has been proven that it works. Until now, there is no evaluation model that has been extensively used to evaluate the SBA system implementation. It is important to bear in mind that to develop an evaluation framework is not an easy task for the evaluators since it deals with diverse and complex programmes in different contexts involving various groups of people with varying relationships (Armstrong and Francis, 2003). Although there are many evaluation models with different purposes, it can be said that they might have one factor in common: all evaluation models are supposed to evaluate in a systematic and reliable way to come up with an outcome (Robinson, 2002).

The CIPP Model by Stufflebeam is chosen as the framework of this study. Why? It is due to various reasons. Firstly, this evaluation model is based on the management-oriented evaluation approach. Generally, this approach helps managers to plan, implement and evaluate programmes (Hakan and Seval, 2011). In the field of education, evaluators use this approach to gain enough information to serve school managers, school administrators, teachers or those related for decision-making process. Unlike objectives-oriented evaluation approach, the main focus of this evaluation approach is not the objectives of the programme. Since this study is focusing on collecting data related to the strengths and weaknesses of the SBA system implementation to make sure that the stakeholders can make decisions weather to revise, continue or discontinue the system implementation, this approach is the best approach to be used. Being a framework to the evaluation studies, its main contribution is providing data to decision-makers and having a very good sensitivity upon feedback. Additionally, it is considered a holistic evaluation framework that enables

decision-makers to come up with a very intelligent judgment (Isaac and Michael, 1982). Even, each component in the model has its own relationship with decision-making process.

Secondly, this model seems to be the most suited approach for this study compared to the other three approaches as I have detailed out in the previous section in the comparison with Tyler's from goal-based approach, naturalistic approach and Kirkpatrick's approach. Similarly, it is more suited to this study compared to Patton's responsive model although they both come from the same approach. Tyler's evaluation model is goal-based evaluation or objective-oriented approach and it focuses more on objectives rather than value. Even if the objectives are achieved that does not mean it is due to the implementation so it is not suitable for this study. Natural approaches use qualitative methodologies whereas this study uses a quantitative methodology. Kirkpatrick's 4-level model is also an objective or goal-based approach and it focuses more on evaluating the programme effectiveness related to corporate training. For Patton's utilisation-focused evaluation model, although it is under the same approach as the CIPP Model, it is not chosen because it is more of a user-focused evaluation, focusing more towards intended users of the evaluations and the way users are going to use it. If the purpose of the evaluation is to meet the audiences' needs then Patton's model is the most suitable but, this study focuses on looking at the attainment of the programme's objectives and the overall impact of the programme.

Thirdly, the CIPP model is believed to be the evaluation model that is the most widely used by evaluators (Zhang *et al.,* 2011). It has gone through a period of evolution since the 1960s, and has been tested for various contexts and also has strong support from theoretical and pragmatic literature (Stufflebeam and Shienkfield, 2007). Frye and Hemmer (2012) believe that this model offers a powerful approach to evaluate a complex, dynamic and having nonlinear relationships between elements in a programme and, since the SBA system is an educational programme with those features so choosing this model seems to be a wise decision. Fourthly is the four dimensions of evaluation that correspond to four types of decisions: context, input, process and product evaluation with planning, structuring, implementing and recycling decisions, accordingly (Gredler, 1996). The good thing about the CIPP Model is that

research could be done using all the four dimensions of the model or one of the combinations of the dimensions depending on the programme's requirements (Azizi *et al.,* 2010). Furthermore, it considers several environmental factors, such as politics or personalities (Kahan, 2008).

Fifthly, it could be used for both evaluation purposes, either formative or summative and could be conducted by various disciplines especially the social and educational projects in the dynamic social context (Stufflebeam and Shrienkfield, 2007). Since this study is evaluating on the summative basis, it is appropriate to use this model. Lastly, Owston (2008) asserted that the CIPP Model is suitable if the purpose of the evaluation is to look at the attainment of the programme's goals and objectives, to improve the programme or to look at the overall impact of the programme. Since it agrees with the focus of this study, the CIPP Model once again seems the right choice.

In conclusion, the CIPP Model is the most readily suited model to be used as a framework in evaluating the system implementation with limited time and access. This study involves an evaluation of the SBA system implementation in Malaysian primary and secondary schools which began in 2011 (primary schools) and 2012 (secondary schools) and is ongoing. Research shows that the implementation of the system has not been evaluated ever since so it will serve a very high potential in benefitting stakeholders if the evaluation is conducted. In context dimension, the system is implemented in all schools from various school category and school types. Its input dimension covers the personnel and resources in the system implementation, its process dimension includes the process and challenges faced by the teachers during the implementation process and its product dimension covers the students' learning outcome.

The following section will discuss the CIPP dimensions in more detail.

3.7 Context Dimension of SBA Implementation Evaluation

Context dimension focuses on the environment where the changes happen so in this study it is the changes distinguished by the school context which involves schools from different category and different types of school.

3.8 Input Dimension of SBA Implementation Evaluation

As mentioned by Stufflebeam (1971a), input evaluation is used to assess strategies for achieving the desired result by focusing on certain personnel, resources or procedures. So, in this study three types of input dimensions are considered, which are the material and personal needs in SBA, appropriateness qualification of personnel and suitability of physical infrastructure and ICT. Input evaluation is assessing the extent of necessity for a complete assessment document, teaching assistant, training of the personnel, teachers' qualifications, teachers' skills in assessment, physical equipment in the classroom, space and also the ICT hardware. Examples of questions for the first type of input dimension is: 'to what extent do you agree that a teaching assistant is needed to help teachers in assessment?' or 'to what extent do you agree that training of the personnel involved should be properly planned and implemented?'. For the second type, it is: 'to what extent do you agree that teachers are sufficiently qualified to implement assessment activities?' and for the third type, it is: 'to what extent do you agree that the space can be adapted to assessment activities?'.

As for teaching assistants, there was no teaching assistant position in the Malaysia schooling system up until 2014 (KPPK, 2014b). In the United Kingdom, either in primary, secondary or special schools, there are many teacher assistants (Teaching-Assistants.co.uk, 2013). Their main role is to increase education standards in the classroom by giving their full support towards schools, teachers and students. Their roles are as follows:

a) Perform routine tasks (getting things ready for lessons, putting up displays around the school, administration task like photocopying or collecting dinner money).
b) Work with individual children, small groups or the whole class.
c) Help children with special needs like those with learning difficulties, physical impairment or those for whom the first language is not English.

There is a research conducted in the north-east of England looking at the effectiveness of teaching assistants in improving the achievement of students in mathematics using a quasi-experimental design (Muijs and Reynolds, 2003).

The researchers found that the use of teaching assistants did not really help in improving low achieving students in primary schools in their mathematics achievement.

Next, 'web-based assessment' is a general term used to explain the use of computer and internet in the classroom assessment process in managing and preparing test questions and assessing student learning (Hamsiah and Raja Maznah, 2011). This is concordance with the National Education Assessment System in Malaysia which suggests that technology is to be used widely in SBA. It is believed that SBA will not achieve its goal if ICT is not well-implemented in schools (Marzita, 2010). ICT support not only helps in developing questions and analysis but also to help students in the assessment process. ICT is much more important when online examination is implemented although schools have to ensure there are sufficient numbers of computers and accompanying hardware for pupils.

3.9 Process Dimension of SBA Implementation Evaluation

In this section of process dimension of evaluation, the aim is to assess the process used to achieve the desired result when the programme is implemented. It consists of constructs that facilitate the SBA implementation such as teachers' attitude (belief, feeling and readiness), understanding, effect of SBA courses, IHT, administration, moderation and monitoring process followed by section on challenges faced by the teachers and the role and importance of SBA towards school improvement.

3.9.1 Teacher's attitude towards SBA

Successful implementation of a system entails complete participation of practitioners. Thus, positive attitudes of teachers towards SBA must be nurtured so that they would have a strong inclination towards SBA and a strong tendency to embark in the SBA implementation. There are several dominant theories emerged in defining attitudes. In the early nineties, Thomas and Znaniecki (Koszalka, 2000) defined attitudes as 'a mental and neural state of readiness, organized through experience, exerting a directive or dynamic influence upon the individual's response to all subjects and situations with

which it is related'. Over time, attitudes have focused more to the thoughts, feelings and intentions for action. It consists of three constructs; i) the cognitive or as some call it the 'beliefs component'; ii) the affective or the 'feelings component'; and iii) behavioural or the 'readiness component' (Zimbardo and Leippe, 1991). Although, some researchers believe that attitude comprises of affective construct only (Fishbein and Ajzen, 1975) or attitude comprises of both cognitive and affective constructs (Narayanan, Rajasekaran and Iyyapan, 2008), these are still debatable. That is why in some writings, belief and attitude or readiness and attitude are written as different constructs. In this study, I use all the three constructs to define attitude towards the SBA implementation. All the three constructs would be a very good measurement of one's attitude towards an educational system implementation. But, it is more complex than that because as Pramanik (1988) pointed out, the attitude of students towards second language learning would not only consider the attitude of the learners themselves but also the attitude towards the second language learning or the attitude towards second language teachers, peers and their parents.

Koszalka (2000) defines beliefs as "a trust, confidence or perception that something would provide benefits". These belief components underpin questions in my questionnaire, such as 'Using SBA in class is efficient in improving students learning' or 'Using SBA in class can help students to become more independent learners'. Belief of someone could be influenced by family and cultural background, social peers, interpretations of prior experiences and individual differences such as gender and personality (Bernat and Gvozdenko, 2005).

Feeling is an "emotional, affective or internal response toward or away from something that is generally predicated on participatory experience" and not relating it to providing any benefits (Koszalka, 2000). To assess teacher's internal feelings toward or away from SBA implementation, the items such as 'Using SBA is not burdening me' or 'Using SBA is not boring' are used. Readiness is defined as a state of mentally and physically prepared toward behaviour (Walinga, 2008). In order to assess teacher's level of agreement, the examples of items are 'I prepare files for my subject before conducting SBA' or 'I always follow the planning instructed by the MOE'. Although belief is

normally seen as central construct in human behaviour and learning related discipline (Bernat and Gvozdenko, 2005), readiness for change plays a very important role in any change of management strategy (Walinga, 2008). So, in order for teachers to implement the SBA system successfully, they have to be ready for change to avoid resistance.

3.9.2 Teachers' Understanding of SBA

In the most basic sense, teachers have to understand that SBA must be implemented effectively as it could lead to increased students' outcome. In the Malaysian context, teachers have to understand several main issues relating to the SBA system implementation like the SBA implementation policy, its general rules, its operational schedules, its four main components together with its characteristics and objectives, the instruments for each components, quality assurance manual, the SBA evidence management, the SBA management system and a lot more. A very important message that has been delivered by the Education Minister of Malaysia, Tan Sri Muhyiddin Yassin at the Specific Message Ceremony conducted at the Putrajaya International Convention Centre (12th of January 2012) clearly indicated that the ministry is moving to a more holistic assessment system to ensure students' potential could be determined and treated as a whole (Muhyiddin, 2012). So, the most important message here is that the Malaysian education system is moving or transforming from an examination-oriented assessment to a more holistic, integrated and balanced assessment. This assessment transformation is supposed to be understood by all teachers who are the major players in the implementation of this new system.

In general, understanding of the SBA system is very important because teachers are the most responsible person to assess students' performance and to understand the context of the subject area (Mansor *et al.,* 2013). Misunderstanding may lead to wrong instructions and actions. Examples of misunderstanding of teachers may be when teachers believe that weekly tests producing a record of marks without providing learning actions following the outcome or of coursework assessment that forms part of some examination without an active feedback in improving pupils' work is a formative assessment (Black, 2007).

No doubt, knowledge and understanding on assessment processes, assessment criteria and standards are very important and they need an active engagement and participation in acquiring it (Rust *et al.*, 2005). This could be achieved through discussion with peers on the meaning of the criteria, the marking procedures using the criteria or the moderation process. According to Heritage (2007a), the four basic elements are required by the teachers to effectively implement formative assessment are domain knowledge, pedagogical content knowledge (PCK), knowledge of students' previous learning and knowledge of assessment. Domain knowledge includes the concepts, knowledge and skills which enable teachers to define learning progressions and then guide assessment and instruction. Furthermore, PCK includes knowledge on multiple models of teaching metacognitive processes and self-assessment skills. Next, teachers also have to understand their students' level of knowledge, their understanding of concepts, their level of skills, their attitudes and also their level of language proficiency. Knowledge of assessment includes all the formative assessment strategies, ways to align them with the instructional goals and also in interpreting evidences.

So, do teachers have a good understanding of this new assessment system? It is well-known that student teachers are not taught much on assessment practises compared to teaching practises (Heritage, 2007a). To make matters worse, the administrators are lacking in the knowledge and skills, so could not offer much help to student teachers to improve their own assessment practises. Some teachers understand most of the benefits but are still not clear with its concept (Wang, 2008). Furthermore, some teachers still want to carry on with their traditional concept and cannot cope with the educational reform. As for teachers in Malaysia, Norizan (2012) found that only 9006 primary school teachers (10 per cent) teaching Year 1, 2 and 3 and 7979 secondary school teachers (17 percent) teaching Year 7 are clear on School Assessment concept. But they do not understand how to integrate the concept of assessment into the teaching and learning process (the formative assessment concept). Teachers are also having problems in developing various assessment instruments as required by the system. However, research by Hamzah and Paramasivan (2009) shows that the English language teachers in the rural secondary schools in Sandakan Division, Sabah, West Malaysia, understood the

purpose of formative assessment. A more detailed perception on teachers' knowledge was discovered by Chan, Sidhu and Yunus (2006) and it is found that most teachers are not exposed to SBA but in terms of knowledge, the teachers believe that they are good in aspects like planning, reviewing and preparing questionnaires but not in score interpretation, item analysis and item bank formation aspects.

In order to assess the knowledge of teachers towards the SBA system, few items are developed. An example is 'SBA is conducted during the teaching and learning processes'. This item is the most basic theme of SBA which differentiates it from the traditional assessment. Another two items which are more specific to the concept of assessment in the Malaysian context are, 'Educational assessment transformation through SBA leads to changes from examination-oriented assessment to an assessment which is more integrated' and 'For centre assessment in SBA, assignments are prepared by the Examination Board'.

3.9.3 Effect of SBA courses in improving teachers' skills

What are the skills that teachers really need to know to be able to successfully implement SBA in schools? Heritage (2007a) listed the specific teacher skills for the effectiveness of formative assessment. First is the skill in creating 'a suitable classroom condition' that allows students to actively involved in assessment. It includes a supportive classroom culture which enables self-assessment and peer-assessment amongst students, a community of learners who respect each other and classroom norms where students can work collaboratively.

Second is the skill in assisting students on how to practise self-assessment. Students might not have the knowledge of how to undertake self-assessment so teachers are responsible to train them. This could be done by advising students to make reflections on their performance. This includes the setting of goals and criteria for success or learning to evaluate their learning performance against specific criteria. In addition, teachers have to know the strategies to develop their students' skills in giving constructive feedback to their peers. Since students normally learn from the models around them then,

the way teachers provide feedback would be set examples for them to follow. This scenario also creates an environment that allows students and teachers to work collaboratively in developing a shared understanding of the students' current learning status.

Third is the teachers' skill of interpreting evidences from students' responses in relation to the criteria for success without taking into account the assessment strategy used. This is then followed by modifying their instruction to 'fill in the gap' that exists between the students' current learning status and the desired goal. All of these processes should be based on strong domain knowledge or else it would lead to misinterpretations of students' responses. Lastly is the skill in matching the teachers' instructions to the learning needs of the students. If the instruction is too difficult for the students, it causes frustration whereas if it is too easy, it causes boredom amongst them. And, theoretically, instructions are supposed to be unique for each student but in practical terms it is very difficult to apply. But, teachers can still practise strategic questioning, form subgroups for instruction or assign different activity to different students in order to match those two.

Generally, teachers' skill could be improved by staff development and in this study it refers to the courses held by the schools and the ministry. There are several ways in which SBA courses can facilitate the improvement of teachers' skill in SBA. In school settings with the formal preparation programmes on SBA courses which sometimes award training certificates, teachers' skills may serve as an anchor for determining required courses. Without extensive professional development, teachers' skill in SBA maybe either assumed or overlooked, leaving teachers without a full complement of skills.

3.9.4 In-house training on SBA system

Hopkins and Levin (2000) believed that workshops should be a place where teachers gain understanding on certain skills or being exposed to demonstration of teaching strategies. In this study, IHT is an example of the workshop held in schools to give knowledge of SBA to teachers. Hence, to ensure that teachers gain full understanding on the SBA system, they should be provided with theories as well as the practical part. Training to teachers in the

Malaysian context has been detailed out in chapter two on the assessment system in Malaysia.

3.9.5 Encouragement by administration

Administration teams should play their role in the SBA implementation. In general, administration should prepare instrumentation, material, time table, facilities, spaces and method to store records and conduct the monitoring and moderation process (MPM, 2012). All of these tasks, if implemented effectively, could inspire teachers to keep learning and improving their SBA practices.

3.9.6 Teachers' commitment towards the Moderating Process

There are two types of moderation process dealing with internal assessment; external and internal moderation. Generally, external moderation is a set of processes to ensure that all the evidences or students' assignments assessed by teachers across the state are following a common set of standards (QSA, 2010). It is also one way of checking whether assessment instruments match with the syllabus requirements and to demonstrate the transparency and accountability of the assessment system (Galloway, 2008). It might involve schools panels, district review panels and state review panels. And internal moderation, as the name suggests, is a process conducted by the nominated teachers in the school to make decisions on the internally moderated assessment (SQA, 2008).

In the Malaysian context, moderation in SBA is one of the QA process to ensure the standardisation of the implementation of tasks given to students and also the scoring given by teachers are based on the Performance Standard or rubric (Lembaga Peperiksaan, 2012). The moderation process is conducted all throughout the SBA implementation period following the schedule set by the SBA committee. The objectives of conducting the moderation process concerning the standardisation are to:

i) Standardise teachers' understanding on tasks requirements and scoring criteria.

ii) Standardise SA scores given by teachers based on standards statement, descriptor and evidence in the PSD.
iii) Standardise CA scores given by teachers based on the prescribed scoring criteria for each task.
iv) Standardise tools, materials and situations according to prescribed criteria to produce an adjusted score.
v) Ensure the validity and reliability of scores.
vi) Ensure that the assigned scores met the capability and performance of students.
vii) Ensure that the assigned scores are fair to all students.

'Reliability of scores' (see point v) above) refers to the extent which the score of the assessment is precise and consistent (Lembaga Peperiksaan Malaysia, 2011). 'Validity of scores' refers to the extent which the assessment score is able to portray the relevancy and meet the needs of what is being assessed. In general, assessment assessed directly by the students' own school teachers is said to be reliable since teachers are the most qualified person to assess their students and they are the subject matter experts (MOE, 2013). In addition, they have received practical on assessment, are able to monitor students' development in continuous basis and to give constructive feedback in improving students' learning, have better understanding on the suitable context for students' development or assessing and are able to give feedback following the standard performance. Furthermore, the reliability of teacher judgement on the quality of students' work could be built and improved by them having discussion and making analysis together in a partnership or team situation (QSA, 2010; HKEAA, 2013). This is consistent with what is being practised during the moderation process in the SBA system implementation in Malaysia. Similar to the issue of reliability, validity of assessment also requires teachers to work as assessors because some skills could only be assessed by teachers and not by external examinations (CSEC, 2010).

There are two types of moderation process implemented in Malaysia, external and internal moderation as shown in Table 3.11. The implementation process is shown in Table 3.12.

Table 3.11: The difference between external and internal moderation

External moderation:	Internal moderation:
To standardise the understanding of subject teachers all over the country concerning assignments and scoring of SA and CA	To standardise the understanding of subject teachers from the same school concerning assignments and scoring of SA and CA

Table 3.12: The implementation of the moderation process

Before moderation: MES: 1. Prepare CA moderation schedule. 2. Send letter and SBA Implementation schedule to JPN. 3. Appoint national assessors. JPN: 1. Receive letter from MES on CA. 2. Prepare CA moderation schedule at state level. 3. Appoint state assessors. PPD: 1. Receive letter from JPN on CA. 2. Prepare CA moderation schedule at district level. 3. Appoint district assessors.	School: 1. Receive letter from PPD on CA. 2. Head teacher appoints SBA committee. 3. SBA committee prepare moderation schedule at school level. 4. Appoint teachers as SBA assessors. 5. Conduct item developing training, scoring regulations and instrument development to teachers.
During moderation: MES: 1. Conduct meeting with Head National Assessor and Head State Assessor on the standardised task understanding and scoring in February. 2. Clarify task requirement and scoring for each task. 3. Discuss with JPN on moderation meeting schedule at state level. JPN: 1. Conduct meeting with Head State	School: For CA: 1. Conduct meeting with Head of Panel and subject teachers in February. 2. Moderate marking: i) Conduct training for teachers. ii) Choose 10% samples of scripts randomly according to the ratio of 3:4:3 for the achievement level. iii) If the difference between teacher score and assessor score is more

<table>
<tr>
<td>Assessor and Head District Assessor.
2. Clarify task requirement and scoring for each task.
3. Discuss with PPD on moderation meeting schedule at district level.
PPD:
1. Conduct meeting with Head District Assessor, District Assessors and Head of Panels.
2. Clarify task requirement and scoring for each task.
3. Discuss with Head District Assessor and District Assessor on moderation task schedule at school Level.</td>
<td>than 5, then teachers have to recheck the scripts.
iv) Teachers have to recheck all scripts if 25% or more scripts show a different score of 5 or more.
3. Score verification:
i) Ensure that the answers are authentic.
ii) Ensure that the given scores are based on the given answers and meet the true capability of students.
iii) Ensure the consistency of marking.
iv) Validate score.
v) Send score to MES.

For SA:
1. Coordinate instruments.
2. Choose evidence based on students' achievement.
3. Compare score given by teachers to those given by moderators.
4. Standardised score for score validation.</td>
</tr>
<tr>
<td colspan="2">After moderation:
1. Give new score to students after score standardisation.
2. Inform teachers on any errors in giving scores.
3. Guide teachers on how to give an authentic score.
4. Giving students a chance to improve their task to improve their score.
5. Head teacher validates score before sending them to MES by keying in the SPPBS system.</td>
</tr>
</table>

To make sure that the moderation process can be conducted effectively, teachers have to be well-prepared and very responsible in contributing scores from SA and CA. They should also prepare instruments and manage evidence in an organized manner so that moderators could coordinate them wisely. In

order to assess the commitment of teachers towards the moderation process, a few items have been formed in this study such as 'Each teacher prepares students' scores for his or her subjects', 'Each teacher prepares evidence according to students' achievement' and 'Each teacher records the marks of evidence in the SPPBS application'.

3.9.7 Teachers' commitment towards the Monitoring process

Generally, monitoring is a process of looking at the implementation of the whole course of the study and the assessment system by the review panels (QSA, 2010). Normally this is conducted after half-way through the whole course of the programme.

In Malaysia, the monitoring is one of the QA processes to ensure the SBA implementation follows the procedures and guidelines set by the MES (Lembaga Peperiksaan, 2012). It is conducted throughout the SBA implementation range of time following the schedule set by the SBA committee. The main aims in the monitoring processes are as follows:

i) To make sure that the instruction on the SBA implementation is adhered to.
ii) To determine challenges faced by the school teachers.
iii) To solve problems faced by the teachers.
iv) To avoid fraud.
v) To prepare SBA implementation reports for improvement purposes.

There are two types of monitoring, external and internal monitoring as shown in Table 3.13.

Table 3.13: The difference between external and internal monitoring

External monitoring:	Internal monitoring:
External monitors from MOE, MES, JPN and PPD check whether the implementation of SBA follows the procedures by MES	Internal monitors in school to check whether the implementation of SBA follows the procedures by MES at least two times a year

Generally, there are three stages involved in the monitoring process (Lembaga Peperiksaan, 2012) as listed in Table 3.14.

Table 3.14: The implementation of monitoring process

Stage	The implementation process
Before monitoring	1. SBA committee plan SBA monitoring process. 2. Prepare the monitoring documents.
During monitoring	1. To review materials and information which are monitored. 2. To observe SBA implementation in classroom. 3. To give advice on the completed monitoring documents.
After monitoring	1. To report the findings. 2. To list the strengths and weaknesses of the implementation 3. To suggest follow-up actions.

The monitoring documents includes survey on the SBA Management Master File, SBA Document Management, Special Case Management, SPPBS Application, Student Evidence Management, Communication Plan, QA-mentoring, QA-moderation, QA-monitoring, Psychometric Assessment, Psychometric Data Management and Psychometric Evidence Management (Lembaga Peperiksaan, 2012). For CA, monitoring is also focusing on the printing and storage process of the assessment instruments (Lembaga Peperiksaan, 2014). Some items have been formed in this study such as 'School SBA committee appoints certain teachers as internal monitors', 'Heads of panel of each subject monitor SBA activity implementation' and 'Evidence of CA is kept in a systematic way in a safe place before it is returned to the students'.

In the context of this study, only moderation and monitoring processes from the QA activities are assessed and not the tracking and mentoring process. This is because the duration required to implement tracking and mentoring process is not fixed to a particular schedule. Teachers might not have experienced them when this study is conducted so it is not possible to assess teachers on those two processes.

3.9.8 Challenges in SBA

Currently, the need to assess students effectively has become increasingly important. However, it is not easy to implement SBA effectively in schools. For example, in order to gain an effective formative assessment, various

relationships and interactions between the students and the teachers need to be changed and managed (Learning Point Associates, 2009). And, in order to change practices in classrooms, it requires a change in teachers' habits rather than changing only their knowledge-transmission teaching tradition (Wiliam, 2009). It is also known that, with this new assessment system, policy-makers claim that teachers are not prone to changes, whereas teachers feel that the administrators do not really understand what is going on within the system (Fullan, 2002). This confusion of attempts at reform seems to challenge the process of implementation. Furthermore, 'most schools are not looking for change' because most of them feel that they have 'settled down' with whatever they have been doing (Levin, 2001, p.3). Although there are popular calls towards changes in students' assessment, it looks like it is difficult for schools to accept changes.

Some of the expected challenges are to gain support from head teachers and cooperation from teachers, lack of knowledge and confidence to implement SBA, problems getting related resources on SBA, school climate (see below for meaning) seems to hinder the effectiveness of SBA, not enough financial resources, extra workloads, no special recognition to teachers and problems with integrity and fairness. Generally, in implementing any innovation in schools, the head teachers' role is very important. SBA requires full support from the head teachers who are expected to understand in depth the concept, principles and procedures of the SBA implementation. They have to give full support to all the teachers and even more than that is to try to solve all the problems that arise (Boon and Shaharuddin, 2011). Support from head teachers, though, can vary. Research has found that in order for teachers to make radical changes in assessment, teachers should be given a support system in terms of availability of time for them to share experiences and advices with other teachers and availability of opportunities to develop their assessment skills within a certain period of time. Availability of time and opportunities should be given until SBA is well established without being interrupted by other innovations (Black, 2007). Hence, all of these practises could be successfully implemented if head teachers are willing to give full support to their staff because they involve support from the administration side and also the leadership factor of head teachers.

Research conducted in Kenya on a 'School-based Teacher Development Programme' shows that the problems in implementing the programme are the poor attitude of the head teachers, lack of support from the MOE, inadequacy in teaching and learning resources and lack of time (Omuga, 2013). Additionally, Boon and Shaharuddin (2011) discovered that if the level of understanding of head teachers on SBA is high, the level of support is also high and the level of challenges faced by the head teachers in implementing SBA is moderate.

Teachers' lack of knowledge about SBA can cause problems for the implementation process. Fullan (2001) believed that in order to improve the education reform process, there seems no doubt that one has to be able to gain an appropriate knowledge and understanding. Although AfL is seen as an assessment which is integrated in the teaching and learning process, its implementation in the classroom cannot be treated as simple. Actually, there are four basic types of knowledge that teachers really need to know if they were to implement formative assessment successfully in their teaching process. They are domain knowledge, pedagogical content knowledge, the knowledge of students' previous learning and knowledge of assessment (Heritage, 2007a). The knowledge that is needed by teachers has been given in the section on teachers' understanding on SBA. In terms of knowledge of assessment, teachers are still lacking. When it comes to actual classroom practises, they tend to use rote learning and recalls rather than conceptual learning (Young and Giebenhaus, 2005) and use normative-based-assessment rather than criterion-based-assessment. This may be due to the fact that knowledge on formative assessment is not well developed in teacher preparation programmes.

Black and Wiliam (1998a) believed that teachers have insufficient knowledge in assessing students, thus affecting their assessment practises due to the high expectation in the scope of formative assessment itself. In order to assess effectively using formative assessment, teachers have to have initial knowledge on the material that will be given to the students and have to be aware of the different responds that they may encounter when dealing with students. These could be gained either from experiences in assessing and evaluating students' work or knowledge on students' beliefs and misbeliefs (Black and Wiliam,

1998a). In practical terms, there are not many teachers who have enough opportunities to gain that experience and knowledge. With insufficient knowledge, teachers are not keen in adopting the innovations in the SBA system. This issue cannot be separated from the issue of training or courses that teachers are currently working on. Black (2007) conducted research upon teachers from six schools and found that they need at least one year of training just to show a slight change in teachers' classroom practises. His research involved a whole-day meeting with all the teachers once every five weeks and visits by research staffs for observation and giving feedback. So, it looks as if it is not an easy job to impart knowledge to the teachers because it involves rethinking of their roles in their core task as a teacher.

'School climate' is a school environment with multidimensional constructs from physical, social and academic dimensions which affects one's feeling and attitudes (Loukas, 2007) as shown in Table 3.15. The school environment can develop a sense of belonging amongst students and feeling of closeness with each other which is called 'school-connectedness' and in turn could improve students' outcome. Since school climate affects behaviour and emotional aspects of teachers and students, it could affect the SBA implementation in schools. Rust *et al.* (2005) believed that teachers and students have to develop a good relationship between them to achieve a meaningful understanding of assessment and to allow teachers to train students in the feedback interpretation. Furthermore, students must learn to work *as a group* and not just *in a group* (IowaCORE, 2011). If the environment is not supportive enough, the implementation of the SBA system will be affected.

Table 3.15: Dimensions in school climate

Physical dimension	Appearance of the school building and its classrooms; School size and ratio of students to teachers in the classroom; Order and organisation of classrooms in the school; Availability of resources and Safety and comfort.
Social dimension	Quality of interpersonal relationships between and among students, teachers, and staff; Equitable and fair treatment of students by teachers and staff; Degree of competition and social comparison between

	students; and Degree to which students, teachers, and staff contribute to decision-making at the school
Academic dimension	Quality of instruction; Teacher expectations for student achievement; and Monitoring student progress and promptly reporting results to students and parents

Extra workload is also a factor that impedes teachers' efforts in implementing SBA in schools. As formative assessment is meant for instruction (Ginsburg, 2009), it requires the assessor to prepare the instruction before-hand for each individual, plus there is the time required for the administrative work (Hunt and Pellegrino, 2002). On the other hand, formative assessment could be viewed as tasks that do not take much of the teacher's time when they are handled in an organised manner. They could happen in an informal form when teachers pose spontaneous questions towards students on their methods of solution, for example, or when teachers use the prepared materials to assess their students (Hunt and Pellegrino, 2002). So, formative assessments should not be viewed as assessments with formal practises or pre-set materials only.

In the Malaysian context, most researchers in Malaysia have found that time-consuming workloads seem to be a big issue in the SBA implementation (Mukhari and Md Amin, 2010; Faizah, 2011) The full attention of teachers is expected in practising SBA during the teaching and learning process but teachers are still burdened with large amounts of administration work. For example, a teacher teaching 24 slots with 8 different classes per week with 36 to 45 students per class (Hailmi, 2013) would have limited time to assess effectively. After assessing one written assignment, for example, a teacher has to keep 36x8 evidences into the file. So, there will be 288 files altogether. Normally, there will be more than 50 evidences per subject so there will be more than 50x8x36 = 14,400 evidences altogether. If all of them were to be printed, then each teacher needs more than 14,400 papers. That is still not alarming. For example, a teacher has one subject with 30 topics. One topic needs six evidences for six bands (Ibnuabbas, 2013). So, there will be 36 pieces of evidences for each topic. Since there are 30 topics, there will be 30x36 = 1,080 pieces. If there are 40 students from eight classes so there will be 1,080x40x8 altogether which is nearly 345,600 pieces of papers altogether.

Teachers then have to key-in thousands of data according to the band of each student in a system which can be slow depending on the computer server; indeed, some teachers feel that they are burdened with the keying in process since the computer server is so slow (Utusan Online, 2012). Furthermore, practising SBA in a classroom itself is time-consuming (Chun, 2006). Preparing for providing individual oral feedback, recording it as to provide evidence to the administration, giving and writing comments on individual students is time-consuming. Generally, all teachers are responsible for all the three portfolios which are students' development portfolio, showcase portfolio and teacher portfolio which are time-consuming in order to prepare them (Norizan, 2012).

The integrity of teachers is also being questioned. Parents and students might not trust teachers' assessment in SBA. Most agree that some teachers tend to give high marks to their favourite pupils but as Chun (2006) suggested, teachers should have consensus between them on the marking criteria. For example, marking criteria for certain subject is based on content structure, writing styles, use of phrases and weightings. When teachers assess accordingly, assessments with very high or low marks will be re-marked by a different teacher and if comes out as a different mark, then the marks will be changed.

From the discussion above, it is obvious that there are a lot of challenges faced by teachers in implementing SBA. With all these challenges, it is clear as why it is still not up to the standard. That is the reason why Levin (2010a) stated that it is easier to create certain attributes to change a complex system but sustaining them is a problem.

3.9.9 The Role of SBA in School Improvement

As was identified in the earlier section on the key characteristics of school improvement from the model and the framework given, it is clear that assessment is one of the important elements in improving schools. However, as Stiggins (2002, p.758) observed, in order to finally connect assessment to school improvement in meaningful ways, assessment must be seen through

new eyes. The recent widespread use of SBA in schools plays a vital role in driving school improvement all over the world. SBA, be it formative or summative, is an assessment that is expanding with more and more research being conducted to explore its impact on student learning. Generally, SBA is seen as a catalyst for education reform by policy makers and educators out there (Chan and Gurnam, 2012). The focus will now shift to consider how assessment practises could improve schools.

If school improvement is to succeed, updated assessment practises are essential (Partners in Learning, 2009). This view is supported by Wiliam (2011) who states that assessment, to a certain extent, could influence students' learning. A good and well-developed formative assessment or AfL has a great impact on student achievement (Hopkins, 2009; Wiliam *et al.,* 2004) with strong evidence from research showing this (Black and Jones, 2006, p. 4; Black and Wiliam, 1998b, p.3). So this implies that formative assessment is the key to school improvement. In addition, summative assessment can also have a positive impact on student learning, as students work harder for assessment that produces higher grades or marks (Queen's University Belfast, 2013).

Formative assessment allows students' thinking to be elicited and displayed. It also allows them to extend, refine and deepen their understanding and then able to move to higher levels of understanding (Herman *et al.,* 2006). Furthermore, good feedback practised in formative assessment also assists students to improve their level of expertise and also provides opportunities for them to clarify their misconceptions. This shows that formative assessment facilitates direct involvement of students in learning. Direct involvement of students (Levin, 2000) or students' voice (Harris, 2003b) does have its contribution to school improvement. So, by means of students' direct involvement in learning, assessment practises have ability in improving school.

Levin (2000) suggested that direct involvement of students could influence the improvement processes in many ways. Firstly, students' understanding and commitment towards school improvement can affect teacher's commitment towards it. Secondly, students' knowledge, ideas and perspective on school improvement could give much information for teachers to implement changes. Thirdly, students are a source of data so the data on their achievement or

attitude at a particular school can improve teachers' willingness to implement changes at the school. From the views presented, it appears as if there is movement from implementing reform 'by adults for students' to something like 'by adults and students for students'. But still, this is not always the case. Even the involvement of students does not really contribute to the success of the improvement process.

Furthermore, the results or data gained from the teachers' feedback on the progress of their students during formative assessment practices could be benefitted by various parties such as teachers, parents or administrators (Young and Giebelhaus, 2005). Teachers might use it in assisting them in their teaching, making decisions on students' learning and encouraging student-centred learning. Parents might use it in setting expectations and helping them with the best instructions and suitable learning environments for their kids. And, for administrators, it provides them with information to plan professional development programs or remediation programs or to develop any suitable educational programme. The involvement of teachers, parents and administrations in student learning can contribute greatly to the success of schools in their improvement initiatives.

Furthermore, SBA is an assessment system with the potential to reduce absentee rate, provide new learning opportunities, increase public satisfaction and confidence towards school, contribute strongly to the quality assurance process, help students improve generic skills to build on their strength and interest (Thimmappa and Sharma, 2003), reduce pressure in public examinations (Esther, 2012) and improve literacy and numeracy skills (Department of Education and Skills, 2011). All of these factors seem to be the major factors which contribute to school improvement. And by implementing SBA in the Malaysian education system, there is an effective way of keeping informed of students' learning status which in turn could enable key actors to find ways in improving students' achievement. And this could contribute to the success of the improvement initiatives.

3.10 Product Dimension of SBA Implementation Evaluation

When talking about product dimension in SBA evaluation, particularly on students' outcome, Levin and Fullan's conclusion appears highly relevant:

> Large scale, sustained improvement in student outcomes requires a sustained effort to change school and classroom practices, not just structures such as governance and accountability. The heart of improvement lies in changing teaching and learning practices in thousands and thousands of classrooms … (Levin and Fullan, 2010, p. 291).

Increasingly, evidence has shown that formative assessment does have its impact on students' outcomes (Leahy and Wiliam, 2009) and can improve student achievement or increase their learning rate up to somewhere between 50 to 100 per cent. The following section on product evaluation discusses the extent to which participants' needs are met. In this study, discussions are focusing on the programme outcomes on students only which are students' knowledge in SBA, their attitudes towards SBA and their motivation to learn. Actually, there are a lot of positive outcomes that could be gained following effective implementation of SBA. As Henrichsen argued (cited in Cheng *et al.,* 2011), the success of any educational innovation is dictated to some extent by learner-relevant factors, for example, students' perceptual, cognitive and affective characteristics.

When practising self-assessment in SBA, students are encouraged to develop their skills in making judgments about their own achievements and learning process (Nicol and Macfarlene-Dick, 2006). Peer-assessment, by contrast, rather than assisting their own self-assessment, involves students making judgments on their peers' performance by looking at the relationship between certain criteria with the goals of the work (Deakin-Creek *et al.,* 2005). Both types of assessment require students to have knowledge in assessment. Furthermore, students must not only have some knowledge of SBA, but also their attitude towards the practises is also important. Research has shown that the positive outcome from the implementation of SBA in schools in Malaysia is when students develop a positive attitude towards learning and assessment (Mansor *et al.,* 2013).

Motivation to learn is the meaningfulness, values and benefits of academic tasks to the learner, regardless whether or not they are intrinsically interesting (Shabait, 2010). This includes components like 'effort', 'self-regulation', self-

esteem' (ARG, 2002). Motivation comes from inside oneself seems to be the most important factor in determining success (The Northern Ireland Curriculum, 2009). Motivation for learning can be affected by various factors such as assessment practises, peer culture, pedagogy, curriculum, school ethos or home support (ARG, 2002).

SBA is seen as a medium to strengthen and sustain students' motivation to learn and gain confidence (Wei, 2010; Tan, 2004). This is supported by the fact that a climate which encourages motivation can be developed when students understand that they are responsible for their own learning, they have choice of what and how to learn, they have opportunities to reflect their own thinking and experience challenges in learning or they understand the link between feedback and the success criteria (The Northern Ireland Curriculum, 2009). These are all important elements that could be gained during the implementation of formative assessment. Research by Maria *et al.* (2009) also found that formative assessment increases students' motivation to learn as it increases involvement in the learning process. There is also research conducted in England which found that when practising self-assessment, students became responsible for their own learning, increased their self-esteem, and became more independent, positive and motivated (Qualifications and Curriculum Authority, 2003). In addition, the type of feedback given by teachers could also improve students' motivation especially for low-achievers (Black, 2004). Students believe that in order to improve their learning, getting only comments as a feedback from teachers is more preferable than getting grades.

Nevertheless, it is still unclear that formative assessment has a positive impact on students' outcomes and the empirical evidence on the advantages of formative assessment is also lacking (Maria *et al.,* 2009). They also believe that to the date, there is no study looking at the impact of formative assessment on students. This may be due to the difficulties in dealing with experimental studies. On the contrary, there is a negative impact of high-stakes tests on the students' motivation for learning but still, the impact varies with the characteristics of students and the conditions of testing (ARG, 2002). In general, tests might lower self-esteem for those students who prefer a creative learning experience, lower the self-image of the lower-achieving students,

limits students' performance and increase anxiety, especially among female students (Harlen, 2005).

3.11 Summary

In this chapter, the most important findings are examined involving assessment, in general and the school-based assessment, in particular. This includes the meaning of the terms and their concepts. Learning theories together with their implications to this study and the critical demonstration of the assessment models are discussed. The knowledge on programme evaluation and its models are well-developed in previous research. The comparison between the evaluation models have been made to determine the best model that fits this study. Some of the most pertinent dimensions that contribute to the effectiveness of the implementation of the SBA system (context, input, process and product dimensions) have also been examined, following the CIPP framework by Stufflebeam. The previous research on the school improvement models is linked to the strategies taken by the Malaysian government as to compare them. The role of SBA in improving schools has also been looked into, to see how it relates to the development of the school improvement process. This chapter is guided by the conceptual framework of this study. All the reviews from this chapter have provided appropriate information to meet the needs of the main purpose of this study which is to evaluate the implementation of the SBA system in schools. The following chapter moves away from such theoretical and conceptual consideration to focus on the research methodology.

Chapter 4: Research Methodology

4.1 Introduction

This chapter addresses the knowledge claims or paradigms of this study and the theories underpinning it, the research design, the population, the samples and also the methods of data collection. There then follows a discussion on the reliability and validity of the instrument used and sets out the pilot study, the data collection procedures and ethical issues.

4.2 Research Paradigm

Paradigms, as central concepts in social science research methodology, play a very important role in preparing a foundation for knowledge claims upon which research is based. Paradigms act as belief systems or assumptions made by researchers as "how they will learn and what they will learn" (Cresswell, 2003, p.6) during their research. The ontological assumption underpinning realism is one of a theoretical perspective that truth in this world can be unravelled by the action of a researcher with a minimum involvement of a researcher to allow reality to be uncovered (Muijs, 2011b). Realism gives rise to positivism. Positivists believe that researchers can be certain in their research as natural science does (Cohen *et al.,* 2007) but the problem with this paradigm is that humans as researchers cannot avoid themselves not influencing their own research.

This study is based on an epistemology or knowledge claim of 'post-positivist assumptions', an epistemology which states that although absolute truth cannot hundred percent be unravelled due to the fact that humans are part of the world, it is still possible to predict outcomes of research findings (Muijs, 2011b). There are four major elements that are related to post-positivism: i) determinism; ii) reductionism; iii) empirical observation; and iv) theory verification (Creswell, 2003, p.7). It is deterministic in that there must be some causes or sources that determine outcomes in a research. Similarly, scientists with their scientific faith also assume that there must be some circumstances that determine events in the natural world (Cohen *et al.,* 2007). Next, in

conducting good research, the big sets of ideas, have to be reduced to a small and discrete set of ideas such as those variables for hypothesis testing and research questions, numeric measures of observation has to be developed and theories have to be verified. Normally, these groups of researchers start with a theory followed by data collection, and end up with either supporting or refuting the theory (Creswell, 2003).

In post-positivist study, the focus is to get an objective and generalisable result by using a valid and reliable instrument, a suitable sampling method and an appropriate statistical method to the collected data (Steenhuis and Bruijn, 2006). Moreover, it does not need an interpretation or a rich description from the interactions between both sides as with interpretivist approaches. The survey or experimental designs are linked with quantitative approaches based on an epistemology of being realist or sometimes positivist (Muijs, 2011b) but are not subjectivist like qualitative approaches. Furthermore, the quantitative paradigm is employed in this research because it has been used by many researchers previously with regards to formative assessment studies (Black and Wiliam, 1998b; Wiliam *et al.,* 2004; Wiliam, 2011) or SBA studies (Reyneke *et al.,* 2010; Cheng *et al.,* 2011). Those studies provide very good sources in supporting this study in terms of their literature review, variables used, theories and models developed.

4.3 Research Design

Research design is an important aspect in conducting research. Research design is "the arrangement of condition for collection and analysis of data in a manner that aims to combine relevance to the research purpose with economy in procedure" (Kumar, 2008, p.31). In this study, a survey research design has been selected based on research purpose and objectives. By definition, a survey research design is a "procedure in quantitative research in which investigators administer a survey to a sample or to the entire population of people in order to describe the attitudes, opinions, behaviours, or characteristics of the population" (Cresswell, 2002, p. 396).

By means of quantitative research, survey design involves collection of numerical data to be analysed statistically in order to explain phenomena

(Muijs, 2011b). When a sample is used, then it has to be a valid sample which represents the population that is the focus of the study. Surveys should be systematic in their planning and execution, have a representative sample, either by using census study data or using the right sampling method, being objective with the observable data and is quantifiable with its data (Isaac and Michael, 1982). Furthermore, a survey can either be a survey of records, mailed questionnaires, telephone surveys, group interview or individual interviews. However, in describing attitudes and behaviours, surveys are capable in answering questions raised, solving problems, accessing needs and setting goals, determining whether objectives have been met or analysing trends across time (Isaac and Michael, 1982).

There are several reasons why survey research design is used in this study. Survey research design allows collection of data from large range of topic from large number of respondents with standardised information (Cohen *et al.,* 2007). It is also very flexible which makes it capable of answering various types of research questions (Muijs, 2011b). In addition, survey research design is frequently been used in programme evaluations in evaluating the success or effectiveness of a programme (Cresswell, 2002). So, this makes it suitable to be used in this study which involves programme evaluation with large-scale data gathering.

Secondly, survey research design is the most suitable research design in measuring opinion about issues and identifying practises (Muijs, 2011b; Creswell, 2002). This study involves measuring teachers' beliefs and opinions on the SBA implementation and their practises in schools, so this is the most appropriate design to use. Teachers' beliefs and opinions are important and the idea of SBA has already been imparted by the ministry to the teachers over recent years. In this study, in order to gain views and opinions of respondents, only pencil-and-paper questionnaires were used and not telephone interviews or face-to-face interviews. Questionnaires when compared to telephone interviews or face-to-face interviews, are less biased, less intrusive and therefore induce more cooperation from samples (Muijs, 2011b). Questionnaires when compared to interviews are able to access greater numbers of respondents and have a higher overall reliability (Cohen *et al.,* 2007). Furthermore, by using questionnaires, sources of error can only come

from instruments and samples whereas in interviews it could come from interviewer, instruments, coding and samples.

Thirdly, although experimental research design also takes a quantitative approach, it involves treatment and manipulation of conditions (Creswell, 2002) so it is not suitable for this study. Experimenting is not practical because there will be no control over the implementation of the system and cannot be undertaken since this research focuses on existing programme, so practically, it is not possible. Furthermore, experimental designs also limit generalisability of findings to real-world settings as it sets its own situations (Muijs, 2011b). However, experimental designs still present a more rigorous explanation between the variables involved (Creswell, 2002) and are more suited to answer causality-type research questions (Muijs, 2011b) compared to survey design. Finally, survey research design is normally used to "support or refute hypotheses about the target population" (Cohen *et al.,* 2007, p.206; Muijs, 2011b) so this is the best research design that could be used. However, survey research design is not suitable for an in-depth exploration of problems, to develop theories or discover the meaning of events (Muijs, 2011b), all of which are outside the remit of this study.

In general, for programme evaluation, quantitative or numerical data can be collected in various forms, such as frequencies, ratings, ranking or test scores, while for qualitative or non- numerical data like words, objects or pictures are used (Gredler, 1996). Furthermore, evaluators might be interested in using quantitative data to summarise data, assess the relationship among variables, interpret group differences or generalise sample results to the population on any evaluation issues.

Although survey research design is seen as the most appropriate design for this study, there are some advantages and disadvantages related to it. By using survey as a means of research design, it saves time and money, is able to reach geographically dispersed sample and participants are not influenced by the researcher (Creswell, 2002). However, it can have low response rate, have less flexibility in answering unless open-ended questions are used and is less rigorous in explaining findings compared to experimental research designs. Instead, the nature of responses in questionnaires which are limited in terms

of length and depth could not produce a deeper understanding of the study (Muijs, 2011b). It also invites slanted or biased responses due to unnatural feeling of respondents or proneness to agree with the items and having a high probability of being distributed to the accessible and cooperative respondents only (Isaac and Michael, 1982).

4.4 The Population and the sample

In Malaysia, there are three types of schools; public or government-sponsored schools, private schools and home schooling. Federal government is responsible for education in Malaysia along with the Education Department in each state coordinating educational matters in its territory following the Education Act of 1996. In the government primary schools, there are two categories of schools which are the Malay-medium national schools, which use Malay language as its medium of instruction and the vernacular schools, using Mandarin-medium or Tamil-medium. Malaysia is a multi-racial country with its major ethnic group the Malays and the Bumiputras (58.1 per cent), the Chinese (32.1 per cent) and the Indians (8.3 per cent) and that is why the government has taken the initiative to develop schools for all major groups of people.

There are nearly 10,099 schools under the government schools with 5,222,531 students and 417,749 school teachers all over Malaysia from the latest data as in 31st of October 2013 (Emisportal, 2013). It shows that the ratio of teachers to students in Malaysia is nearly 1:13. Table 4.1 shows the teacher-student ratio for the last three consecutive years (EPRD, 2013). It clearly shows that in general, the ratio decreased from year to year for both types of schools in Malaysia. Since, student-teacher ratio is related to education quality (Department of Statistics, 2010), this shows a good sign.

Table 4.1: The teacher-student ratio in Malaysia

School	2010	2011	2012
Primary	1:12.97	1:12.60	1:12.29
Secondary	1:13.39	1:13.12	1:13.11

4.4.1 The Population of the Study

Muijs (2011b, p.33) defines population, a group of individuals with the same characteristics as "the group you want to generalize your findings to". This is important for a researcher to be clear of what it is because in sampling, defining the population seems to be the first thing to do (Entwistle and Nisbet, 1972). In this study, the population consists of school teachers working in the government primary and secondary schools in Kelantan Darul Naim. Kelantan, which is situated in the north-east of Peninsular Malaysia, is one of the states in Malaysia. Why Kelantan? Altogether Malaysia has 13 states and three federal territories. In Malaysia, all schools throughout the country are practising the same education system controlled by the MOE based in Putrajaya. Considering practicality factors, the research could not be conducted all over the country as it has certain limitations in terms of funding, time, facilities and accessibility. Kelantan is chosen due to its convenience factor. In Kelantan, different types of schools, either primary or secondary schools and rural or urban types, are almost evenly spread in all the ten main districts so this enhances the generalisability of the findings later. According to the information from the Kelantan Education Department, by the year 2014, there were 591 government schools in the ten major districts in Kelantan (JPN Kelantan, 2015). Details are shown in Table 4.2. It clearly shows that for Kelantan, there is approximately one teacher for every 11.39 students.

Table 4.2: Number of schools, teachers and students

	School	Teacher	Student
Malaysia	10,099	417,749	5,222,531
Kelantan	591	29,748	338,771

4.4.2 The sampling method

After deciding on the population, it was necessary to decide on how to choose a sample from a population using a suitable sampling method. It had to be undertaken in a valid manner as samples are supposed to represent population and the result from the samples will be generalised to the population; but at the same time, sampling could not avoid sampling error. There are two types of sampling method in educational research, the probability sampling method and non-probability sampling method. Probability sampling method involves

"selecting a relatively large number of units from a population, or from specific subgroups (strata) of a population, in a random manner where the probability of inclusion for every member of the population is determinable" (Tashakkori and Teddlie, 2003, p.713). Being the most rigorous sampling method in a quantitative research, it is the best method to avoid a biased sample that does not represent the population being studied (Muijs, 2011b). And with these representative samples, generalisations of the results made to the population will be stronger (Creswell, 2002). Being less rigorous and not so strong to claim on generalising results to a larger population, non-probability sampling method could still be used. It is a sampling method where samples are selected from the population due to availability, convenience or representative (Creswell, 2002).

In this study, a stratified sampling method was chosen in sampling. According to Creswell (2002), stratified sampling is a method whereby the population is stratified according to specific characteristics, such as gender, race or type of school and then simple random sampling is used to choose a sample from each stratum. In this study, the sample was divided into two subcategories which are the sample of the schools and the sample of the teachers.

4.4.3 The sample of the schools

Since this study is looking at the differences of the implementation according to different type of schools and different school categories, stratified random sampling is seen to be the most suitable method. Although simple random sampling is the most unbiased form of sampling, it might not help if generalisation are to be made to a specific subpopulation that is too small to be reliably chosen. According to Mohd Majid (1990), if simple random sampling is applied to a population with an imbalance in a trait of interest, simple random sampling may cause increased sampling error. For example, from Table 4.3, the total number of schools in Kota Bharu district is 144 schools whereas in Jeli district there are only seventeen schools. From the number of schools, it can be predicted that if simple random sampling is used, more schools from Kota Bharu district will be chosen as samples. If that is the case, samples will not present Kelantan school teachers fairly. Using stratified

random sampling, provided a good chance that each stratum in the population had enough representatives to contribute.

There are two ways of selecting samples from each group in the stratum: disproportionate sampling or proportionate sampling (Sekaran, 1992). Disproportionate sampling is a selection of individuals not proportional to their representation in the total population, whereas proportionate sampling is vice versa. Disproportionate sampling is suitable for the stratified population which is too small or too large and it is more convenient and saves costs when conducted in research studies (Sekaran, 1992).

A stratified sampling frame was used where schools were first divided by districts, followed by type of school and type of school category. Disproportionate sampling was used to choose school samples in this study since the number of secondary schools and primary schools and also the rural and urban schools were not the same and the difference is quite big. Two secondary schools were randomly chosen from each district, one from urban area and one from rural area. The same goes for primary schools. For each district there were four schools chosen altogether. So, the samples of schools include forty schools from all over Kelantan.

The names of the schools are gathered from the Kelantan Education Department. The ten districts in Kelantan are Kota Bharu, Pasir Mas, Tumpat, Bachok, Pasir Puteh, Tanah Merah, Kuala Krai, Machang, Gua Musang and Jeli. The number of school samples involved in this study is shown in Table 4.3.

Table 4.3: The number of secondary and primary schools chosen

District	Number of Secondary Schools	Number of Secondary Schools Samples	Number of Primary Schools	Number of Primary Schools Samples	Total number of schools
Kota Bharu	48	2	96	2	144
Pasir Mas	25	2	56	2	81
Tumpat	13	2	35	2	48
Bachok	18	2	34	2	52
Pasir Puteh	16	2	39	2	55
Tanah Merah	16	2	34	2	50
Kuala Kerai	13	2	41	2	54
Machang	10	2	29	2	39
Gua Musang	8	2	39	2	47

Jeli	5	2	15	2	17
Total	172	20	418	20	590

Table 4.3 presents the number of schools which were planned to be sampled during the final study. However, the actual school samples gained were from 37 schools. Data from two secondary schools from Jeli district and one primary school from Bachok district was not available. Out of 37 schools, 16 schools responded to the questionnaire posted from UK. Other schools were selected again when the researcher went back to Malaysia for data collection. Questionnaires were then posted or sent directly to the selected schools. One primary school from Kuala Krai district was selected through snow ball sampling with the help of the head master from the nearest selected school. Four schools were chosen through convenience sampling from four different districts as there were no responses from the selected schools from that area.

The study focuses on looking at the differences between urban-rural because it is an established issue with the government policies concerning the differences between the two areas. Moreover, the study also focuses on the primary and secondary schools because the SBA implementation started with the Year 1 students (primary) in January 2011 and the Year 7 students (secondary) in the following year. The information on the differences between the two aspects is very much needed by the authorities. Furthermore, many researchers and practitioners have identified the differences in academic achievement (Mohd Burhan, 2005), working conditions (Vegas, 2007) or educational quality (Othman and Muijs, 2013) based on the different type of schools (urban-rural) or school category (primary-secondary). So, the researcher feels that there is a need to conduct a research on primary-secondary and urban-rural schools differences on the SBA system implementation in her own country. Even, the Preliminary Report of Malaysian Education Blueprint 2013-2025 states that the Malaysian government is really concerned with the international assessment like PISA and TIMSS and also students' achievement regarding the gap between rich and poor students, those from urban and rural and also between states in Malaysia (Malaysian Education Blueprint, 2012).

4.4.4 Teachers' samples

In this study, teachers are the respondents. It was important to gain the opinions of teachers on the SBA system implementation. This is due to the fact that SBA is a newly-implemented system in the Malaysian education system so teachers are the most involved and the most concerned with the system compared to other parties. Teachers are seen as the most important stakeholder in ensuring the success of this new system implementation (Salmiah, 2013). Furthermore, time was limited for the data collection so there was not enough time to gain information from other important people in the system like the officers from the ministry or the head teachers. In recent studies in Malaysia, there are considerable numbers of studies on SBA conducted using teachers as respondents (Mansor *et al*., 2013; Othman *et al.,* 2013; Salmiah, 2013) because of the recognition of the importance of teachers' views on this issue.

Teachers were randomly selected from the randomly chosen schools. For some schools, the head teachers or the SBA teachers were charged to select teachers, the researcher having previously explained the procedure to them. The same process applied to the questionnaires, as these were posted to the head teachers whereby the procedures were detailed in the letters sent to them.

According to Muijs (2011b), it is better to choose all pupils from the randomly selected schools as samples when looking at the school effects. This is supported by Creswell (2002) who stated that a large sample is needed if research uses powerful statistics and then generalise its result to a population. But the problem is, how large is large? It seems like 'the larger, the better' although Creswell does give a rough estimate on the number of respondents for a survey which is approximately 350 respondents. Similarly, Cohen *et al.* (2007) also agreed with the idea of 'the larger, the better' although they do state that sample size can be determined by, inter alia, the study purpose, the nature of the population, the kind of analysis that are used, the size of the population, the number of strata required and the style of the research. Larger samples give greater reliability in research, allowing for the use of more sophisticated statistics and allow a better way to overcome problems of sampling error. In order to choose as many samples as required, there are

several limitations such as the availability of samples, access, funding, the overall size of the population, the number of variables involved in the survey, time, administrative support, the number of researchers and resources (Creswell, 2002; Cohen *et al.,* 2007). In this study, the population involved was nearly 30 000 teachers and the samples were expected to be nearly 1000. If this is the case, it means that the samples are approximately 3.33 percent of the population. The number of schools, teachers and students in Kelantan are given in Table 4.4 (JPN Kelantan, 2015).

Table 4.4: Number of school, teacher and student in Kelantan

	Secondary School	Primary School	Total
Number of School	173 (29.3%)	418 (70.7%)	591
Number of Teacher	12,946 (43.5%)	16,802 (56.5%)	29,748
Number of Student	143,945 (42.5%)	194,826 (57.5%)	338,771

Following the issue of sample size is the issue of representativeness of sample. Representativeness of sample is important in making sure that the sample is valid (Cohen *et al.,* 2007). Are the school and teacher samples truly representing the whole population? Cohen *et al.* (2007) state a large sample size does not necessarily guarantee representativeness; the same can be said of small sample sizes. But for random sampling, the larger the sample size, the greater the chance there is of it being representative. So, how to ensure that those samples really represent the total population? The sampling method used is important in making sure that the sample fairly represents the population (Muijs, 2011b). In this study, to achieve a fairer representativeness of the population of teachers, stratified random sampling was used whereby weighting was made to the sample according to type and category of schools as those two factors are related to the study purpose.

4.5 Methods of Data Collection

There are several methods of collecting quantitative data such as by using questionnaire, closed-ended interview, closed-ended observations or documents such as census or attendance records. Normally, when a survey research design is used, data collection is conducted using a standard questionnaire form either pencil-and-paper questionnaires, phone interviews, face-to-face interviews or web-based and email forms (Muijs, 2011b). In this

study, surveys are carried out using pencil-and-paper questionnaires and are self-administered without the presence of the researcher. The self-administered questionnaire is a method where instruments are given out to the respondents by the researcher or someone in an official position and the respondents are free to complete and return the questionnaire later to the researcher (Oppenheim, 1996). According to Cohen *et al.* (2007), the use of questionnaire with the presence of the researcher enables questionnaire to be distributed to a large number of samples in a shorter period of time and could get a good response rate if, for example, when questionnaires are given straight away after staff meetings. Alternatively, in the case where the researcher is not present, teachers may have ample time to respond, feel less threat and can be more honest to answer sensitive issues. In some cases where the researcher is not present, a head teacher or a teacher might represent the researcher in dealing with the teachers' queries. Furthermore, self-administered questionnaires, in any form of dissemination either by hand, postal or e-mail allows low cost of administrating and less intrusive towards respondents upon sensitive questions compared to face-to-face interview (Eiselen and Uys, 2005). It is easier to administer and analyse and could also reduce the possibility of interview bias. By using postal survey, a large number of respondents could be accessed especially those living in remote areas and respondents will have their own time and convenience to answer and this will in turn, could increase their potential and knowledge to respond (Cohen *et al.,* 2007).

In this study, questionnaires were used to gather numeric scores data for the following reasons:

1. A large amount of information was needed from a large number of respondents.
2. There was a limited time frame, energy and money available to collect and analyse data and this study deals with a large number of respondents. If interviews or observation were used, there would be insufficient time to cover all the respondents in the range.
3. Interviews might cause difficulties in the analysis stage due to the volume of answers given by the respondents, and because it involves many variables from all the four dimensions of evaluation (context, input, process and product dimensions).

4. Since this study is interested to see the differences between the respondents in different type of school and different school category, using questionnaires increased the probability of getting feedback from a large geographical area rather than using closed-ended interviews or observations.

However, questionnaires also pose disadvantages, such as getting a low response rate especially from those questionnaires which are too long or too complicated and there is also a lack of control on the response itself (Eiselen and Uys, 2005). Questionnaires can also produce major biases if great care is not given during a briefing period (Oppenheim, 1996) and too much time needed for follow-up and data entry (Muijs, 2011b). In particular, postal questionnaires are not suitable for poor literacy respondents, visually handicapped people or people with language difficulties (Oppenheim, 1996). However, Oppenheim's observations are not applicable to this study because the respondents are all teachers teaching government schools and all of them are not in this category. Although research has shown that the response rate can be as low as twenty to thirty percent, initial steps can be taken such as to pre-notify potential respondents, make sure to include a prepaid-returned stamped addressed envelope with a clear and detailed instruction inside or personally delivering them to the selected schools (Cohen *et al.,* 2007). The researcher undertook all those steps including getting teachers' hand phone number so that follow-up could be done easily. In this study, pencil-and-paper survey is more preferable than online survey because the culture of most of the teachers in Malaysia is not really exposed to online questionnaire. Indeed, some of the senior teachers are still having problem with ICT. So, if online questionnaire were used, the expected return rate would be very low. To worsen the condition, some teachers still have technophobia or some schools have less access to the internet.

4.5.1 Instrument of the study

An instrument of the study could be a test, scale, observation procedure, questionnaire or interview schedule (Isaac and Michael, 1982). In this study, questionnaire is used, designed by the researcher. According to Cohen *et al.* (2007), there are four major steps that have to be followed before one comes

out with the items in the questionnaire. The researcher has to decide the objectives of the questionnaire, to generate constructs and the required data to meet the objectives of the research, to decide the kinds of measures, scales or questions required in the questionnaire and lastly is that the population and the sample must agree with the items in the 'personal details' in the questionnaire. Once these decisions have been made, a researcher can then write the items. In planning a questionnaire, Cohen *et al.* (2007) consider the process of operationalising the questionnaire. '*Operationalising*' is a process involving a shift from generalised purpose of questionnaire to a specific set of features so that empirical data could be gathered from the items in the questionnaire.

In this study, the constructs were developed from past literature reviews. And then, based on the operational development for each construct, items were developed. Initially the researcher looked at the characteristics listed by Stufflebeam in the CIPP Model in evaluating any programme. Then, constructs in several instruments related to SBA implementation were determined. Up to this point, there were no instruments on SBA from previous research which followed the dimensions suggested by Stufflebeam, so the researcher matched any context, input, process and product dimensions whenever appropriated. Instruments from Asian countries like those from Hong Kong, Singapore, Australia and Malaysia were more focused on to make sure that the constructs were more suited to an Asian context. For example, use was made from the items on 'impact on motivation to learn' in Cheng *et al.* (2011) such as 'My child reads more English books in S4 than before or My child is becoming more interested in English in S4 than before'. Another example of items are in Faizah (2011) such as 'Not having enough time' or 'Time spent on non-academic matters related to SBA'. The researcher also refered to the constructs developed in the documents from the ministry to choose the most suitable ones for this study. Then, the questionnaire had to go through the process of operationalising it, as mentioned by Cohen *et al.*

According to Chua (2006), good characteristics of a questionnaire is those which are suitable for the respondents' readiness level, has a systematic format, has clear instructions, come with a cover letter and consent form and have undergone validity and reliability processes during pilot study. A cover

letter (Appendix G) contained the purpose of the study and the researcher contact information and also the importance of the participants' participation to the study. The consent form (Appendix C) explained the purpose of the study, ethics approval and the assurance of anonymity. In addition, it was personalised by having a hand-signed signature of the participants.

In this study, the questionnaire consisted of four pages of A4 paper, the maximum number of pages recommended for questionnaires (Muijs, 2011b). The main section of the questionnaire covered three evaluation dimensions of the SBA system implementation (input, process and product evaluation) in six sections with 83 items altogether. The main sections were:

- Teachers' attitudes and understandings of SBA and the input evaluation constructs;
- Teachers' statements on the courses they have attended and their views on courses on SBA, IHT on SBA, administration side, monitoring and moderation process;
- Challenges of SBA;
- The role and the importance of SBA in school improvement
- Teachers' perceptions on students' attitudes, knowledge and motivational towards learning.

The demographic information was a closed question asking about gender, race, school grade, type and category of school, age, teaching experience, assessment experience and academic achievement.

The questionnaire, as shown in Appendix A, contained six sections, Section A to Section F. Section A contained some items on process evaluation and all items on input evaluation. Section B focused on items on information of SBA courses. Section C considered the challenges teachers were facing in the SBA implementation. Section D contained items on process evaluation, whereas Section E centred on items on product evaluation. The last section, Section F contained demographic information. It clearly shows that the items on process evaluation are a lot more than other dimensions as shown in Table 4.5. These are due to the fact that, in evaluating the implementation of SBA, constructs on process dimension are important and are more readily to be improved by the decision-makers compared to other dimensions. As such, despite criticism and complaints from various parties and agencies, after improving some of the

processes and the related challenges determined by recent relevant research, the Malaysian government decided to continue SBA with few improvements without suspending or abrogating it (Rahim, 2012). Furthermore, the reason why the product dimension is looking at those three factors – attitude and knowledge of SBA and the motivational towards learning, is because SBA itself is an assessment designed not only meant to assess the mastery of knowledge (cognitive process) but also assessing students' feeling, behaviour, interests and attitude (Wei, 2010). SBA is therefore looking at the overall picture of students' potential (Lembaga Peperiksaan, 2014).

Table 4.5: Contents of the questionnaire

Factors	Item section	Total Item
Context evaluation	F2 – F3	2
Input evaluation	A14 – A21	8
Process evaluation	A1 – A13; B24 – B28; C29; D30 – D31	52
Product evaluation	E32 – E34	8

In this study, a five-point Likert scale is used for marking questionnaires' statement. It is a five-point scale with a statement of strongly agree, agree, neutral, disagree and strongly disagree for the respondents to choose according to their level of agreement on given statements. All items are designed on a five-point Likert scale of agreement apart from the demographic information. The rationale of using five-point Likert scale is that not only being the most common scale used in programme evaluation (Losby and Wetmore, 2012) but has the advantages of using scales which, compared to other scales, are easy to construct, provide more precise information on respondents' degree of agreement or disagreement and also the possibility of including any item unrelated to the study (Oppenheim, 1996). On the other hand, its disadvantages are the lack of reproducibility, the problem of demonstrating validity, having central tendency bias, acquiescence bias or social desirability bias (Bertram, 2006).

4.6 Validity and Reliability of the questionnaire

Validity of an instrument shows the extent of an instrument measuring what it is supposed to measure (Muijs, 2011b) which means that any instrument that is able to accurately measure any prescribed variable is considered a valid

instrument for that variable. Furthermore, any instrument or test is said to have high validity if the ability to measure what should be measured is high. That is why validity seems to be the most important concept when talking about the design of an instrument in educational research exceeded the importance of our research design and statistical analyses. Results are only meaningful when they measure that which they are supposed to measure.

There are three major types of validity; those of criterion validity, content validity and construct validity (Muijs, 2011b). All are somewhat related to the theoretical knowledge of the concept. Criterion validity 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" (Muijs, 2011b, p.59). Content validity is a concept whereby the content of the items in the questionnaire is actually measuring the concept being measured (Muijs, 2011b). In relation to content validity is face validity, which is the concept of "whether or not the test looks valid on its surface" (Jackson, 2003, p.44).

In this study, after the questionnaire was developed, it went through face validity and content validity checking before going through pilot testing. Since face validity could be assessed by obtaining comments on the questionnaire from the respondents (Muijs, 2011b), a few school teachers who were practising SBA in schools were asked to judge on the items to ascertain whether it looked valid or not. Their comments seem to be very useful because teachers were the ones who were going to respond to the questionnaire later. A few changes were made upon their comments. For example, for the item 'SBA is non-threatening me' is changed to 'SBA is not burdening me' when the teachers feel uncomfortable with the word 'non-threatening'. What is needed to achieve content validity is an extensive search of the literature on the concepts that we are trying to measure (Muijs, 2011b) or referring to the experts in the area of interest (Jackson, 2003). I have searched literature on most of the concepts related to this study from the theories, previous instruments and past research findings and also I have referred to experts in measurement and evaluation, SBA and experts in the Malaysian curriculum to ensure that content validity of the questionnaire is achieved. A professor in the field of measurement and evaluation in education examined the items to make sure

that the items represent relevant information of the concept in this thesis. Lastly is the construct validity. It is a concept to assess "the extent to which a measuring instrument accurately measures a theoretical construct or trait that it is designed to measure" (Jackson, 2003, p.45). Constructs as defined by Hair *et al.* (2006) are unobservable or latent factors, which cannot be measured directly but can be measured using measurement on one or more variables or indicators to get a reasonably accurate measure of those constructs. Examples of constructs are attitude, creativity, motivation, performance, intelligence, verbal fluency, neuroticism or scholastic aptitude (Jackson, 2003). To achieve construct validity, a strong theoretical knowledge is necessary of the concept to be measured (Muijs, 2011b), so a wide-ranging review of the literature was undertaken to achieve these. And, to check for construct validity in this study, exploratory factor analysis (EFA) was used. EFA is a technique used to explore the interrelationships among a set of variables (Pallant, 2007, p.179). EFA is a technique that is used to "identify the underlying factors or latent variables for a set of variables" (Harrington, 2009, p.9). Although it is less complex and sophisticated than confirmatory factor analysis (CFA), it can still be used to confirm or test specific hypothesis or theories on certain constructs. On the contrary, some believe that EFA is more on exploring patterns in the data whereas CFA is more on testing hypothesis (Indiana University, 2006). In other words, EFA is data driven when factor structure is interpreted post-hoc based on the results and CFA is theory driven when factor structure is specified priory. Statistical analysis software like SPSS, SAS or STATA is used for EFA whereas Amos, LISREL or Mplus is used for CFA. SPSS calculates the factor loadings, a value range from -1 to +1 to show the strength of the correlation of variables to the factor.

Next is the concept of reliability of an instrument. Reliability is referred to stability or internal consistency of an instrument in measuring certain concepts (Jackson, 2003). It is defined as "the extent to which test scores are free of measurement error" (Muijs, 2011b, p.61) or stable in repeated measurement and consistent (Creswell, 2002). There are five types of reliability depending on the number of times the instruments are administered, the number of different versions of the instrument and the number of individuals who provide information (Creswell, 2002). They are test-retest reliability, alternate forms reliability, alternate forms and test-retest reliability, internal consistency

reliability and inter-rater reliability. In order to check for test-retest reliability, one has to administer the same test to the same group of participants on two different occasions and then check the correlation between the two scores (Pallant, 2007). If the correlation value is reasonably high, it means 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 but with the same concept (Creswell, 2002). The third type of reliability combines both the two earlier concepts, where two different versions of instrument with the same concept are administered twice at two different time intervals, and then the stability of scores are administered. Internal consistency reliability is the extent to which scores from an instrument are internally consistent across the items. Inter-rater reliability is the extent to which scores from one sample are stable when more than one observer records the behaviour of respondents at the same time using the same instrument.

Since this study involves an instrument with one version and administered once only to all the participants, internal consistence reliability was a practical decision to measure reliability. Creswell (2002) asserted that internal consistence reliability is used when there is only one version of instrument or questionnaire involved, the questionnaire is administered once and each individual in the sample completes the questionnaire. But still, this study lacks aspects of reliability gathered from test-retest reliability since internal consistency is only looking at the correlation between all items that make up the constructs to ensure that the items are measuring the same concept (Muijs, 2011b).

There are two ways of measuring internal consistency reliability either using split-half reliability or alpha coefficient of reliability. In this study, alpha coefficient of reliability or Cronbach alpha is used to measure internal consistency of all the concepts or among the items. Its values can vary between 0.00 and 1.00. The value represents the correlation value of each item with the sum of all the other items that make up the scale, hence it is known as inter-item correlations coefficient (Cohen *et al.,* 2007). It has to be over 0.7 for the test to be internally consistent (Muijs, 2011b; Nunnally, 1978).

I will now consider the relationship between these two concepts of validity and reliability. "A test can be reliable without being valid" but a test cannot be valid without being reliable (Jackson, 2003, p.45). Or in other words, "if the test is valid, it must be reliable" *(*p.46*).* So, it is clear that any test that shows a high value in reliability does not necessarily have a high value in validity and vice versa. This is due to the fact that validity is measuring data related to knowledge and facts whereas reliability only concerns with the consistency of scores. That is why the procedure of assessing validity of an instrument is more difficult than assessing reliability.

4.7 Pilot Testing

Although pilot studies are sometimes expensive and time consuming, avoiding them may cause more problems as claimed by Oppenheim (1996, p.48) that "almost anything about a social survey can and should be piloted". The importance of pilot studies is to help the researcher with validity, reliability and the practicality of the instrument (Cohen *et al.,* 2007). It is to make sure that the questions are well-understood by the participants so that wrong responses can be minimised and the question sequences are in a good order (Oppenheim, 1996). It is also a form of checking the time taken to complete the questionnaire, checking the suitability of the questionnaire in terms of length or eliminating difficulties in wording (Cohen *et al.,* 2007). Checking superficially could be achieved by gaining feedback from a limited number of respondents but then it has to go through statistical analysis of the data to check for its reliability and validity. The sample has to be good-sized and representative in order to develop reliable data for the statistical analysis.
All of the items in the questionnaire were developed by identifying the potential and wide-ranging reviews of the themes related to SBA and the educational programme evaluation. Identification was achieved by reviewing the literature and also comparing it with questionnaires that have been developed previously. Items were developed for each sub-construct where a few sub-constructs contributed to certain constructs. They were then presented to the expert in this field. After the items has been validated by the professor who is expert in measurement and evaluation, few changes were made to the items and sub-constructs. Then it went through translation process. Since the study took place in the east of Peninsular Malaysia, it was translated from

English language to Malay language. The translated version was then checked by the lecturer from language department to correct any grammar mistakes. Then, two lecturers who are teaching measurement and evaluation were chosen to validate the questionnaire in the Malay version after the instrument was rewritten.

A pilot study was conducted in June 2012 and sixty teachers from four different types of schools was the total response. Teachers were selected using convenience sampling method. Twenty teachers were from a secondary school in urban and rural area respectively and ten teachers were from a primary school in urban and rural area respectively. The schools are located in the south of Peninsular Malaysia which were not the place of the actual research. According to Cohen *et al.* (2007), piloting is supposed to be conducted to the representative sample who will not receive the final version of the questionnaire. A number of participants involved to represent a good-sized sample with similar backgrounds to the respondents of the actual study and the area where the schools were located are also considered representative for this study. Two things that were conducted during the pilot study: (i) the researcher's representatives administered the complete instrument to the teachers and determined the length of the time taken for them to complete answering (ii) the researcher's representatives asked the teachers for comments and suggestions after they have finished answering the questionnaire. All of the comments from pilot study were referred to the experts to make changes to the items. All of the feedback was very important to increase the validity and reliability of the instrument. The question on SBA implementation for pilot study is shown in Appendix K.

4.7.1 Findings of Pilot Study of the Questionnaire

a) Validity and Reliability for the Input Dimension of 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. Cronbach's Alpha was used to give a measure of internal consistency reliability of each construct as in Table 4.6. Cronbach's alpha for the first construct which was the 'material and personal needs in SBA'

was 0.779, for the second construct was 0.517 and for the third construct was 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 something had to be done to the second construct. A18 was deleted and the alpha for the second construct was increased from 0.517 to 0.675.

Table 4.6: 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	

Next, to certain extent, factor analysis could be used to assess the validity of a measurement scale (Muijs, 2011b). An independent exploratory factor analysis (EFA) using the principal components analysis (PCA) with a rotation called Direct Oblimin in IBM SPSS was conducted on the questionnaire. Eigenvalue or variance extracted by the factor greater than 1 was used. Only factor loadings more than 0.3 were counted towards any factors but still, meaningful interpretations had to be carefully been done when double loadings occur (Muijs, 2011b). PCA was run after the deletion of A18. From Table 4.7, two factors were formed. Realising the fact that factor analysis provided an indicator of construct validity and not much on content validity, I chose to continue with the three constructs that contributed to the input evaluation because I wanted the constructs on 'personnel' and 'physical infrastructure' to be separated.

Table 4.7: 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

As the default setting was changed, as in Table 4.8, and the number of factors were specified to three, it clearly showed that the type of items that contributed to the factors were the same as had been presented earlier by the theory.

Table 4.8: 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

b) Validity and Reliability for the Process Dimension of Evaluation

Process dimension of evaluation consists of ten main constructs altogether (belief, feeling, readiness, understanding, skill, IHT, administration, moderation, monitoring, challenges). Teachers' attitude construct consists of three sub-constructs which were belief, feeling and readiness towards SBA. Considering the value of Cronbach alpha in Table 4.9 and the three-factor pattern matrix of factor analysis in Table 4.10, few decisions had been made:

i) For belief sub-construct, all the three items, A1, A2 and A5 were maintained and one item, item A7 from 'feeling' construct was added to it. Item A7 states that 'SBA can be effectively incorporated into existing lessons'. Item A6 was deleted to increase the Cronbach alpha value. The Cronbach alpha value for the four items, A1, A2, A5 and A7 was improved from 0.671 to 0.749.

ii) For feeling sub-construct, only one item, item A4 was maintained and item A3 was deleted. Since the Cronbach alpha was low, another two items were added, which were 'SBA is not boring' and 'SBA is compulsory'.

iii) For readiness sub-construct, A8 was deleted and the Cronbach alpha value for the three items, A9, A10 and A11 was improved from 0.581 to 0.654.

Finally, there were 10 items contributing for teachers' attitude towards SBA.

Table 4.9: 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	

Table 4.10: Three-factor pattern matrix showing the factor loadings of teachers' 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

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 4.9 and the three-factor pattern matrix factor analysis of items in Table 4.11, few decisions had been made as follows:

i) For the fourth construct, 'teachers' understanding on SBA', all the three items, A12, A13 and A14 were maintained with the Cronbach alpha value was 0.644.

ii) For the fifth construct, 'effect of SBA courses', all the four items, B26i), ii), iii) and iv) were maintained with the Cronbach alpha value was 0.992.

iii) For the sixth construct, 'IHT on SBA', all the three items, B27i), ii) and iii) were maintained with the Cronbach alpha value was 0.973.

iv) For the seventh construct, 'encouragement by administration', item B28i) was deleted and the Cronbach alpha value increased from 0.543 to 0.610.

Table 4.11: Three-factor pattern matrix showing the factor loadings items

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 4.9, the three-factor pattern matrix factor analysis in Table 4.12 and the two-factor pattern matrix in Table 4.13, all items were maintained according to the theory.

Table 4.12: Three-factor pattern matrix showing the factor loadings of items

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 4.13: Two-factor pattern matrix showing the factor loadings of items.

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

c) Validity and Reliability for the Process Dimension (School Improvement Constructs)

There are six items for the first construct, the 'role of SBA' and six items for the second construct, the 'importance of SBA'. Looking at the Cronbach alpha value in Table 4.14 and the two-factor pattern matrix in Table 4.15, two items had been taken out from the construct 'role of SBA' which were item D32ii) (reduces pressure in public exam) and item D32iii) (improves literacy and numeracy) and moved them to the second construct, which was the 'importance of SBA'. The list of items for process dimensions before and after going through validity and reliability procedures were shown in Table 4.18.

Table 4.14: 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 4.15: Two-factor pattern matrix showing the factor loadings of items

	Component	
Item	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

d) Validity and Reliability for the Product Dimension of Evaluation

Looking at the Cronbach alpha value in Table 4.16 and the three-factor pattern matrix in Table 4.17, all items were maintained. Altogether, there were eight items for product evaluation constructs.

Table 4.16: 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 4.17: Three-factor pattern matrix showing the factor loadings

	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 4.18 showed all the items on input, process and product evaluation dimensions before and after undergo validity and reliability process. Finally, out of 71 items, 68 items were retained.

Table 4.18: 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
1.Material and personal needs in SBA system	A15, A16, A17	A15, A16, A17
2. Appropriateness of personnel's qualifications	A18, A19, A20	A19, A20
3. Suitability of physical infrastructure and ICT	A21, A22, A23	A21, A22, A23
Process evaluation construct		
1.Teacher's attitude: teacher's belief	A1, A2, A5, A6	A1, A2, A5, A7
2.Teacher's attitude: teacher's feeling	A3, A4, A7	A3, new item, new item
3.Teacher's attitude: teacher's readiness	A8, A9, A10, A11	A9, A10, A11
4.Teacher's understanding on SBA	A12, A13, A14	A12, A13, A14
5. Effect of SBA courses on improving skills of SBA	B26i, B26ii, B26iii, B26iv	B26i, B26ii, B26iii, B26iv
6. IHT on SBA	B27i, B27ii, B27iii	B27i, B27ii, B27iii
7. Encouragement by administration	B28i, B28ii, B28iii	B28ii, B28iii
8. Moderation process	B29i, B29ii, B29iii	B29i, B29ii, B29iii
9. Monitoring process	B30i, B30ii, B30iii	B30i, B30ii, B30iii
10. Challenges	C1 – C12	C1 – C12
School Improvement Construct		
1. Role of SBA	D32i, D32ii, D32iii, D32iv, D32v, D32vi	D32i, D32iv, D32v, D32vi
2. Importance of SBA	D33i, D33ii, D33iii, D33iv, D33v, D33vi	D32ii, D32iii, D33i, D33ii, D33iii, D33iv, D33v, D33vi
Product evaluation construct		
1. Student's attitude towards SBA	E34i, E34ii, E34iii	E34i, E34ii, E34iii
2. Student's knowledge in SBA	E35i, E35ii	E35i, E35ii
3. Motivational source towards learning	E36i, E36ii, E36iii	E36i, E36ii, E36iii
TOTAL ITEMS	71 items	68 items

4.8 Data Collection Procedures

The data collection procedure involved the distribution of questionnaires to the selected primary and secondary schools all over the state of Kelantan. Before the main research was conducted, the researcher sent a letter to the educational institutions in Malaysia, those being the MOE, the Ministry of Higher Education and the Kelantan Education Department. The letter requested permission to carry out research in forty primary and secondary schools in

Kelantan. The researcher also gained ethical approval from the School of Education of the University of Southampton. Lists of schools together with their addresses and phone numbers were gathered from the ministry website. Questionnaires were posted to the selected schools in early January 2013.

In February 2013, the researcher collected the questionnaires from the schools. It happened that there were not enough questionnaires collected as expected so the researcher went to another schools to gain additional data. On visiting the schools, the researcher was not allowed to wait and give briefings to the teachers as the teachers were busy with their duties. The head teachers suggested that the questionnaires were left at the school for about one week and they would post them to the given address or call the researcher to come and collect them. The researcher had prepared the envelope together with the stamp and the researcher's address written on it to ease teachers' work.

Out of 1000 sets of questionnaires posted from the United Kingdom, only 342 sets (34.2 percent) were received back. The response rate of the actual study is discussed in section 5.4. According to Cohen and Manion (1994), the return rate of questionnaires between seventy to eighty percent is enough. Kerlinger (1970) also stated that the return rate which exceeded eighty percent seems to be a good return rate. Hence, the return rate of this study from postal method did not show a good return rate. In order to improve the not so good return number of questionnaires, the researcher went to schools herself to discuss directly with the head teachers. An additional number of questionnaires were also posted to some schools. Some of the questionnaires were also given to the teachers and head teachers who were willing to help distribute questionnaires to their nearby schools.

4.9 Ethical Issues

The UK's Economic and Social Research Council (ESRC) devised the Research Ethics Framework to improve social science research in the United Kingdom and it has proven to be up to the standard (ESRC, 2012). Research ethics as defined by ESRC is "the moral principles guiding research, from its inception through to completion and publication of results and beyond". Ethical issues in research are concerned with the rights of the participants in retaining their

autonomy. An example of those rights can be seen following the guidelines developed by the Federal Drug Administration (Creswell, 2002) with three main principles which are the respect of participants concerning their consent, their right to privacy and anonymity, having the benefits of research outweighs the participants' risk and also the issues of justice towards all participants.

The research ethics guidelines are very important as a way to try and ensure research is conducted with respect to persons, animals and the environment, respect for knowledge and also respect for justice and compliance with law (HPRC, 2012). Furthermore, ethical issues should also consider responsibility to the research profession, the participants and the public. Whenever there is a conflict or crisis between the three components which seem not to fit with one another, a professional ethical judgment is needed by the researcher. As such, ethical practises actually have to be considered in all steps of the research process (Cohen *et al.,* 2007).

The researcher has followed all the ethical procedures that have been stipulated by the Faculty of Social and Human Sciences, University of Southampton and also by the Ministry of Education in Malaysia. Furthermore, before the researcher conducted the main research, the proposal of the study had been checked and approved by the Ethics Committee from Electronic Research Governance Online (ERGO) system team from the University of Southampton. In this study, several steps were taken to ensure that the ethical issues were preserved by:

i) Gaining ethical approval from all relevant parties before the main study was conducted.
ii) A full explanation on the research purpose and objectives was provided to the head teachers and teachers so that they understood the overall picture of the study.
iii) Each questionnaire was provided with an 'informed consent' form stating the objectives of the study, the right to withdraw at any time, the procedures taken by the researcher, the right to ask any questions, the benefits from the study, the assurance of no known risks that the respondents might encounter and the information about the researcher so that they are guaranteed to be protected.

iv) Each questionnaire was provided with a cover letter stating the information of the study together with the researcher's name, phone number and address and also expected amount of time to complete the questionnaire. It also stated that the survey was for research purpose only and would be kept confidential.
v) All of the respondents were asked to sign the informed consent form and a copy given to them to show that they agreed to participate with full knowledge and they understood that their rights were protected.
vi) In order to retain anonymity of respondents, respondents were not asked to provide their names and personal information on the questionnaire.

The above steps are mostly concerned with the anonymity of the participants. Other than that, the researcher is expected to be very honest with the study as well as being responsible and respect the participants' culture, religion, gender or any relevance differences which might arise at the field work. And, the researcher should also be aware of the potential risk when conducting research or after it has been done. It includes the safety at the fieldwork or the possible distress that might occur after committing oneself in the research (HPRC, 2012). In this study, the potential risk assessment has been submitted and approved by ERGO committee of the university.

In addition, a researcher is also expected to be responsible to the research profession and to the public. A researcher should present a full and detailed report on the study without any misinterpretations which should be made accessible to various parties especially the stakeholders or other researchers. It is also advisable to present this information in a way that is easily understood by the public. There should also be a right balancing between the participants' or data's confidentiality and the extent to which the researcher communicates the findings of the study. Not forgetting the fact that the important issues in the research study is to avoid plagiarism and maintained the ownership in any manner. According to Cohen *et al.* (2007), plagiarism could be avoided by citing the author's name together with the date of the publication and listing their names in the reference section. Not only respecting the participants, research profession and the public, consideration should also be given towards the research sites. As in this study, permission was gained from the head

teachers before the study was conducted and the teachers were not disturbed during their teaching sessions. This study complies with Ethical Guidelines for Educational Research by British Educational Research Association, BERA (www.bera.ac.uk/system/files/3/BERA-Ethical-Guidelines-2011.pdf).

4.10 Summary

This chapter explained how research methodology was designed to answer the research questions of the study. A cross-sectional survey design was conducted to evaluate the implementation of the SBA system in schools in Malaysia by examining the current attitudes, beliefs and opinions of the school teachers. It also included explanations on populations, samples, how samples are sampled and also the use of questionnaires in collecting data. The concepts and constructs were developed following the research questions. A questionnaire with six sections was developed and then piloted. Exploratory factor analysis and internal consistency reliability were used to analyse the data from the pilot study. The results of the pilot test confirmed the validity and reliability of the items from the instrument. The detailed findings of the main study are discussed in the following chapter.

Chapter 5: Analyses and Results

5.1 Introduction

This chapter explains the analyses and the results of the study. It consists of twelve sections. The second section following this introduction provides an overview of the data analysis. It follows with the section on the requirement for SEM analysis and the response rate. The fifth section provides preliminary data analysis which includes missing data, outliers and assessing normality. Then, it follows with the profile of the respondents. The seventh section includes analysis and results of the descriptive statistics. Sections eight and nine include the measurement models and the structural models formation in testing the hypotheses. Then, it follows with a section on moderating effects of moderation variables. The eleventh section provides the summary of the hypothesis testing and the last section is the conclusion of this chapter. Most importantly, this chapter describes how the research questions stated in Chapter 1 are answered.

5.2 Overview of the data analysis

Before continuing with the analysis of the data, the outcome of the data collection is presented. Responses from 776 teachers from 37 schools all around the state of Kelantan have contributed to the final study. The schools involved were the government-funded schools. Out of 776 teachers, 350 teachers were from primary schools and 426 were from secondary schools. Or, if referring to school types, 396 teachers were from rural schools and 380 teachers were from urban schools.

Data analysis in this study consisted of two main phases. The first phase was the preliminary data analysis involving screening and cleaning process and checking for missing data, outliers and normality. These were undertaken using the Statistical Package for Social Science (SPSS) version 21 software. The second phase included the use of a technique called structural equation modelling (SEM). SEM is a statistical analysis technique developed to analyse the inter-relationships among multiple variables in a model (Zainuddin, 2012).

It is a multivariate technique which combines both factor analysis and multiple regression analysis to simultaneously examine a series of interrelated dependence relationships among the latent and observed variables or between the latent variables themselves (Hair *et al.,* 2006). Latent variables, which are known as factors, constructs or unobserved variables, are the hypothetical constructs of interest in a study which cannot be measured directly; these include attitude or knowledge whereas observed variables, which are known as latent indicators, manifesting variables or measuring items, are variables that can be measured directly using developed instruments or tests. Observed variables are used to define or infer latent variables. SEM could be seen as a statistical methodology which takes a confirmatory approach rather than an exploratory to analyse a structural theory involving certain phenomenon (Byrne, 2010). Compared to ordinary least square (OLS) methods, SEM has the ability to gain a more efficient and accurate findings because it analyses the structure of mean, variance and covariance simultaneously and it is also more efficient in making estimation for multiple variables (Zainuddin, 2012 and Nazim and Ahmad, 2013). On the contrary, for OLS method, it analyses the variation of mean score and is more suitable in making estimation for a single item variable. Furthermore, SEM is capable of correcting for measurement error by providing estimates of these error variance parameters which could not be done by traditional multivariate procedures (Byrne, 2010). The disadvantages of SEM are that the process might be more tedious in order to achieve model fitness, the risk of having to delete some items due to uni-dimensionality issues and there is also a significant element of researcher judgment involved. SEM procedures are also limited in terms of dealing with nonlinear relations and multilevel problems (Lei and Wu, 2007). Figure 5.1 shows the two phases and sub-stages of the data analysis used in this research.

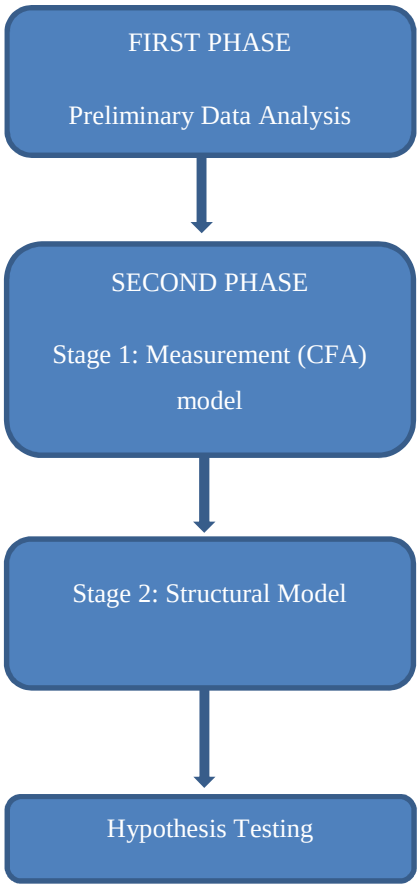


Figure 5.1: Data analysis procedure

The second phase of the analysis consists of a two-stage approach to SEM analysis to test hypothesis in order to answer the research questions of this study.

a) First stage: Assessing the Measurement (CFA) model

The measurement model is a model which specifies the relationship between latent and observed variables. Confirmatory factor analysis (CFA) is a SEM technique which is theory-driven and it is used to confirm a hypothesised measurement model by determining the goodness of fit between the model and the sample data (Byrne, 2010). This technique is an inferential statistics as it estimates a population covariance matrix from the observed covariance matrix from the hypothesised model and it minimises the difference between the two matrices (Schreiber *et al.,* 2006). It is used within the framework of SEM and it is appropriate when the researcher has had some knowledge on the

latent variable structure (Byrne, 2010). Since this researcher has had some knowledge from the theory and empirical research, the link between the latent variable and observed variable is modelled. The CFA procedure is conducted to assess the measurement model (Zainuddin, 2012). Then, it follows with a confirmatory assessment of the measurement model for uni-dimensionality, validity and reliability requirements for all observed variables for each latent variable. At this stage, some hypotheses of the study are tested.

b) Second stage: Analysing the SEM

The second stage involves the analysis of the structural model based on the theoretical interrelationships among the latent variables and follows with the interpretation of the text output. Then, further hypotheses of the study are tested.

SEM is used in this study because it is a comprehensive way of assessing and modifying theoretical models and come out with a meaningful conclusion to the interrelations of the theoretical constructs. Or, in other words, it is suitable for theory testing and development (Anderson and Gerbing, 1988). The two-stage approach is used here. According to Hair *et al.* (2006), the two-stage approach allows validation to be conducted to the measurement model before the researcher could proceed with the full structural model to avoid bad measures. The second reason for using the two-stage approach is that it provides an accurate value of reliability of items in each construct. By conducting it in two different stages, the interaction between the measurement and structural model could be avoided.

There are two main types of SEM; covariance-based SEM and variance-based SEM (Henseler, 2010). Covariance-based SEM is used in theory-testing, particularly in three main types, either strict confirmatory theory-testing situation, forming alternative models or generating models to fit the given data. It also allows for model measurement error. On the other hand, variance-based SEM is more useful in causal predictive analysis. For covariance-based SEM, software such as AMOS (Analysis of Moment Structures), LISREL, EQS, MPLUS or Mx is used whereas the latter uses software such as PLS-Graph or Smart PLS. The good thing about AMOS software especially AMOS Graphics is

its convenience in drawing graphics identical to the schematic diagram of a model in the study rather than working with equation statements in computer programmes (Byrne, 2010). Since the purpose of this research is to test theory and to form alternative models if possible, covariance-based SEM using AMOS Graphics software was chosen to test the hypotheses in this study.

5.3 Requirement for SEM analysis

In order to employ SEM, there are few procedures or assumptions that have to be made regarding sample size and the normal distribution of the data. The size of sample has to be taken into consideration before the measurement model analysis is performed so as to obtain a reliable and meaningful estimate (Hair *et al.,* 1998). Small sample sizes can cause less stable estimates of the covariance and correlations, limit the power of statistical tests in SEM (Kline, 2005) or can create problem with the goodness-of-fit indices (Quintana and Maxwell, 1999). Some researchers suggest a different sample size that suits SEM analysis. For example, according to Anderson and Gerbing (1988), 100 to 200 is the minimum sample size whereas Kline (2005) stated that sample size exceeding 200 is considered large enough to run SEM analysis. Furthermore, a minimum sample size for a stable maximum likelihood estimation procedure is 100 to 150 (Hair *et al.,* 2006). However, Hair *et al.* (2010) also came out with the suggestions as in Table 5.1. Even so, most researchers believe that the number of sample depends on the model complexity and basic measurement model characteristics.

Table 5.1: Minimum sample required

Number of latent constructs and item	Item communality	Minimum sample
Five or less latent constructs with more than three items per construct	0.6 or higher	100
Seven or less latent constructs with more than three items per construct	0.5 or higher	150
Seven or less latent construct with some constructs with less than three items	0.45 or lower	300
More than seven latent constructs with some constructs with less than three items	0.45 or lower	500

In this study, the sample size is 776 so it meets the suggested recommendations. Therefore, SEM analysis can be conducted. Normal distribution of data is explained in detail in the following section.

5.4 Response rate

During the final study, nearly 2500 questionnaires were distributed to 50 schools all around Kelantan altogether as shown in Table 5.2. 30 to 50 questionnaires were distributed to each school. Of the 2500 questionnaires, 826 were returned. This was equivalent to 33.04 percent of response rate. Unfortunately, some questionnaires were found to have a lot of unanswered items or same responses written on the same page. So, out of 826 questionnaires, 50 were discarded. Finally, there were 776 usable questionnaires which contributed to 31.04 percent of the final response rate or the effective response rate. Between 9 to 38 teachers responded per school and they came from a total of 37 schools altogether. According to Randall and Gibson (1990), the common range of response rate is between 21 to 50 percent. So the response rate in this study was considered appropriate. The final samples were considered representative of the population since they came from both types of schools from both categories of schools from the ten major districts in Kelantan. The sample size was deemed adequate for the application of structural equation modelling to address the research objectives.

Table 5.2: Rate of return of questionnaires

	Number of questionnaire	Response rate (in percent)
Total questionnaires distributed	2500	100.00
Completed questionnaires received	826	33.04
Unusable questionnaires	50	2.00
Usable questionnaires	776	31.04

5.5 Preliminary Data Analysis

Before analysing the data, it has to go through screening and cleaning process. This is to make sure that all errors made during data entering can be removed. These errors include out-of-range values (outliers) which are the data that are well below or well above other scores. Screening processes involves error

checking for each variable and correcting the error in the data file (Pallant, 2007). To screen for out-of-range values during error checking, frequency analyses are conducted for each variable. According to Pallant (2007), outliers can be deleted from the data set or the value is changed with a high score but not too different from the rest of the score for that particular variable. Appropriate changes have been made to the data in the data file such as the value of 6 or 55 is changed to 5 or value of 11 is changed to 1.

5.5.1 Missing Data

Missing data refers to data missing from the data file. This is considered normal when conducting research with human beings (Pallant, 2007). The frequency and the percentage of missing data is shown in Appendix J. The maximum percentage of missing data was 1.3 percent. According to Cohen and Cohen (1983), missing data up to 10 percent of the total data is acceptable as it does not really affect the interpretation of findings later. Overall, 0.293 percent of the data values in this study were missing. Out of 68 items, 21 items had complete data. Many items had missing data but not many data missing from each item. Three items (a13, b26ii and b28i) had the highest percentage of missing data with 1.3 percent. In this study, the amount of missing data is minimal, so when the data is analysed for missing value analysis through single imputation method using SPSS, Ttest and Crosstab, tables are not produced. So, the researcher decided to replace the missing value with the mode of that particular variable as suggested by Humphries (2013).

5.5.2 Outliers

Outliers are cases which data is substantially different from the other data with an extreme value on one or more variables (Hair *et al.,* 2006). Outliers due to data entry error or mistakes in coding were identified during the data cleaning stage. The presence of outliers can be beneficial or not. It can represent some characteristics of the respondents which might not be determined with normal analysis, although it can also seriously distort the statistical results. According to Pallant (2007), outliers can be checked using techniques such as histograms, boxplots or the 5% trimmed mean. In this study, outliers were

identified using the histogram and normal Q-Q plot. The histogram and normal Q-Q plot for each variable is shown in Appendix I. Looking at the tails of the distribution of each histogram, there is no data point sitting on their own, out on the extremes. So, there is no potential univariate outlier.

5.5.3 Assessing normality

Normal distribution of data occurs when all the data or scores of the variables are normally distributed or form a symmetrical and bell-shaped curve with the greatest frequency of data in the middle and smaller frequency towards extreme (Pallant, 2007). There are two types of non-normality, univariate and multivariate non-normality. If univariate data is non-normal, meaning that each and every variable is not normally distributed, then automatically multivariate distributions are also non-normal (Newsom, 2012). But, in case the univariate distributions are normal, it does not mean that multivariate distributions are also normal so the researcher has to check. Univariate normality of data can be assessed by various methods such as inspecting the shape of the distribution in histogram especially for large samples, normal probability plots or looking at skewness and kurtosis values (Pallant, 2007). The histogram and normality probability plots for each variable are shown in Appendix I. Skewness is the measure of the symmetry of the distribution whereas kurtosis is the measure of the 'peakedness' of the distribution (Pallant, 2007). If the data is normally distributed, the value of skewness and kurtosis will be 0.00. Table 5.3 shows the value of skewness and kurtosis of the variables involved in this study.

Table 5.3: Skewness and kurtosis values of variables

Variable	Item	Mean	SD	Skewness	Kurtosis	Min	Max
Input evaluation							
Material and personal needs in SBA system	a14 a15 a16	4.25 4.22 4.41	0.81 0.90 0.74	-1.16 -1.10 -1.43	1.60 0.83 2.56	1 1 1	5 5 5
Appropriateness of personnel's qualifications	a17 a18	3.79 3.38	0.90 1.01	-0.64 -0.40	0.21 -0.44	1 1	5 5
Suitability of physical infrastructure and ICT	a19 a20 a21	3.10 3.31 3.30	1.06 0.99 1.11	-0.35 -0.67 -0.61	-0.78 -0.24 -0.48	1 1 1	5 5 5
Process evaluation							
Teacher's attitude: belief	a1 a2 a3 a4	3.27 3.30 3.56 3.60	1.02 0.98 0.93 0.88	-0.61 -0.55 -0.76 -0.84	-0.37 -0.40 0.19 0.57	1 1 1 1	5 5 5 5
Teacher's attitude: feeling	a5 a6	2.56 3.25	1.17 0.99	0.25 -0.58	-0.98 -0.27	1 1	5 5

	a7	3.44	1.06	-0.58	-0.24	1	5
Teacher's attitude: readiness	a8	3.82	0.89	-0.91	0.90	1	5
	a9	3.38	0.99	-0.42	-0.41	1	5
	a10	3.74	0.85	-0.59	0.36	1	5
Teacher's understanding on SBA	a11	3.54	0.99	-0.66	0.01	1	5
	a12	3.82	0.82	-0.97	1.29	1	5
	a13	3.86	0.87	-0.56	0.23	1	5
Effect of SBA courses on improving skills of SBA	b24i	3.71	0.78	-1.14	1.88	1	5
	b24ii	3.67	0.72	-1.09	0.98	1	5
	b24iii	3.60	0.79	-0.80	0.34	1	5
	b24iv	3.71	0.72	-0.96	1.03	1	5
IHT on SBA	b25i	3.53	0.91	-1.01	0.73	1	5
	b25ii	3.28	1.04	-0.60	-0.42	1	5
	b25iii	3.22	1.01	-0.38	-0.65	1	5
Encouragement by administration	b26i	3.07	1.17	-0.27	-0.85	1	5
	b26ii	3.10	1.17	-0.25	-0.82	1	5
Coordination process	b27i	3.76	0.86	-1.24	1.90	1	5
	b27ii	3.85	0.80	-1.24	2.51	1	5
	b27iii	3.87	0.83	-1.24	2.47	1	5
Monitoring process	b28i	3.65	0.88	-0.88	0.75	1	5
	b28ii	3.50	0.92	-0.74	0.14	1	5
	b28iii	3.93	0.81	-1.28	2.90	1	5
Challenges	c1	2.53	1.05	0.40	-0.47	1	5
	c2	2.55	1.05	0.54	-0.39	1	5
	c3	3.43	1.10	-0.57	-0.37	1	5
	c4	3.15	1.05	-0.19	-0.65	1	5
	c5	3.46	1.13	-0.48	-0.62	1	5
	c6	2.96	1.13	0.11	-0.80	1	5
	c7	3.39	1.18	-0.34	-0.82	1	5
	c8	3.89	1.11	-0.90	0.09	1	5
	c9	3.65	1.04	-0.61	-0.14	1	5
	c10	3.45	1.09	-0.39	-0.51	1	5
	c11	3.48	1.08	-0.43	-0.55	1	5
	c12	3.50	1.13	-0.36	-0.74	1	5
School Improvement						1	5
Role of SBA	d30i	2.82	1.04	0.08	-0.72	1	5
	d30ii	3.05	1.01	-0.30	-0.60	1	5
	d30iii	2.86	0.98	-0.04	-0.60	1	5
	d30iv	3.23	0.96	-0.54	-0.31	1	5
Importance of SBA	d31i	3.35	0.95	-0.58	-0.31	1	5
	d31ii	3.43	0.91	-0.81	0.01	1	5
	d31iii	3.48	0.91	-0.84	0.24	1	5
	d31iv	3.54	0.91	-0.85	0.35	1	5
	d31v	3.36	0.99	-0.66	-0.29	1	5
	d31vi	3.35	0.91	-0.67	-0.06	1	5
	d31vii	3.38	0.99	-0.55	-0.23	1	5
	d31viii	3.42	0.95	-0.81	0.16	1	5
Product evaluation							
Student's attitude towards SBA	e32i	3.14	0.98	-0.40	-0.75	1	5
	e32ii	3.03	1.02	-0.16	-0.87	1	5
	e32iii	3.18	1.02	-0.36	-0.60	1	5
Student's knowledge in SBA	e33i	3.39	1.02	-0.77	-0.09	1	5
	e33ii	3.33	1.03	-0.64	-0.38	1	5
Students' motivational source towards learning	e34i	2.88	1.03	-0.16	-0.84	1	5
	e34ii	3.10	0.99	-0.32	-0.44	1	5
	e34iii	3.10	1.06	-0.36	-0.76	1	5
Total	68						

For the distribution to be considered normally distributed, Table 5.4 shows the range of values of skewness and kurtosis from several researchers like Tabachnick and Fidel (2001), Curran *et al.* (1996) and Kline (2005). The results for this study from Table 5.3 indicated that the value of skewness and kurtosis

for most variables were within the range of -2.00 to +2.00 and all the kurtosis value are less than 3.00. In conclusion, looking at the values of skewness and kurtosis, the results indicated that all the variables could be considered univariate normally distributed. Multivariate normality distribution is presented later when AMOS is applied.

Table 5.4: Range of values of skewness and kurtosis

Type of distribution	Skewness value	Kurtosis value	
Normally distributed	[-2.00, 2.00]	[-2.00, 2.00]	Tabachnick and Fidell (2001)
Normal distributed Moderately non-normal Extremely non-normal	<2.00 [-2.00, 3.00] >3.00	<7.00 [7.00, 21.00] >21.00	Curran, West and Finch (1996)
Extremely skewed or kurtosis	>3.00	>8.00	Kline (2005)

5.6 Profile of Respondents

This section presents the general profile of respondents of the survey questionnaire. Table 5.5 presents the respondents' profile in terms of school location, school category, gender, race, age, teaching experience, experience in practising SBA, number of times attending courses on school-based assessment, teaching their option or not and their highest academic achievement in addition to the profile of schools involved according to school location and category of school. The sample consisted of 776 respondents.

The profile of respondents according to school location and category in this study were found to be promising. The number of respondents from urban and rural type of schools was almost the same. Nearly half (49 percent) of them were from urban area and the rest were from rural area. Similarly, nearly half of them are from secondary schools which contribute 54.9 percent and the rest are from primary schools. Nearly two-third of the respondents were female (74.7 percent) and the remaining were male (24.6 percent). The majority of the respondents were Malay (93.6 percent) followed by Chinese (3.5 percent), Indian (1.8 percent) and other races (0.5 percent). As for age, nearly half of the respondents (49.7 percent) were below 40 years old and the other half were above 40 years old. The mean and standard deviation for age were 40.23 and 8.38 respectively. Nearly one-fourth of the respondents (25.4 percent) had teaching experience of fewer than 10 years similar to those who had more than

20 years of teaching experience (24.1 percent). Nearly half of the respondents (49.1 percent) had 10 to 20 years of teaching experience. The average year of teaching experience was 14.75 with its standard deviation 7.90. Nearly half of the respondents (49.9 percent) experienced one year in practising SBA and almost all (93.3 percent) practiced SBA in the range of 0 to 3 years.

Looking at the number of times respondents attended SBA courses, nearly half of them (53 percent) attended courses once, 15. 3 percent of them attended two times and 13.4 percent of them never attended the course. But, there was a small portion of them (0.4 percent) who attended the maximum number of times (eight). Next, nearly two-thirds were teaching their own option with the rest not doing so. Finally, the largest group of respondents (71.4 percent) were degree holders, followed by diploma holders (23.8 percent), master holders (3.4 percent) and none held PhDs. So, the majority of the respondents were Malay, female, degree holder, had 10 to 20 years teaching experience and 1 year's experience in SBA implementation and had attended courses on SBA once.

Table 5.5: Profile of respondents (N=776)

Variable	Category	Frequency	Percentage (%)
School Location	Urban Rural	380 396	49.0 51.0
School Category	Secondary school Primary school	426 350	54.9 45.1
Gender	Female Male	580 191	74.7 24.6
Race	Malay Chinese Indian Others	726 27 14 4	93.6 3.5 1.8 0.5
Age	40 and below Above 40 years old Mean = 40.23 SD = 8.38	386 381	49.7 49.1
Teaching experience (in years)	< 10 years 10 to 20 years > 20 years Mean = 14.75 SD = 7.9	197 381 187	25.4 49.1 24.1
Experience in practicing SBA (in years)	0 1	88.0 387.0	11.3 49.9

	2	179.0	23.1
	3	70.0	9.0
	5	7.0	0.9
	6	1.0	0.1
	9	1.0	0.1
	10	1.0	0.1
	12	1.0	0.1
Attending courses on SBA (in times)	0	104	13.4
	1	411	53.0
	2	119	15.3
	3	48	6.2
	4	6	0.8
	5	13	1.7
	6	2	0.3
	8	3	0.4
Teaching own option	Yes	600	77.3
	No	164	21.1
Highest academic achievement	Diploma	185	23.8
	Degree	554	71.4
	Master	26	3.4
	PhD	0	0.0

5.7 Analyses and Results

Results were based on the percentage of respondents' views on each item. The scale ranges from 1 to 5 which is from 'strongly disagree' to 'strongly agree'. Tables 5.6, 5.7, 5.8, 5.9, 5.10 and 5.11 show the frequency counts and percentages of responses to each item of the questionnaire. This section relates to the first, second and third research questions proposed in Chapter 1. The research questions were as follows:

RQ1: What are the teachers' perceptions on the input dimension of the SBA system implementation?

RQ2: What are the teachers' perceptions on the process dimension of the SBA system implementation?

RQ3: What are the teachers' perceptions on the product dimension of the SBA system implementation?

Input dimension of evaluation consists of material and personal needs in SBA, appropriateness of personnel qualifications and suitability of physical infrastructure and ICT. Overall, data from Table 5.6 indicated that the

resources, procedures and personnel qualifications involved in this study have met the desired result. Greater percentage of the respondents chose 'strongly agree' or 'agree' compared to the other three choices ('strongly disagree', 'disagree' or 'neutral').

Material and personal needs such as a complete assessment document, a teaching assistant and properly planned training were seen as greatly needed if the SBA system were to be implemented successfully, where over three-quarter of the respondents 'agreed' and 'strongly agreed' with this. A very high percentage of respondents (91.6 percent) thought that in order to implement SBA effectively in teaching, 'training of the personnel involved should be properly planned and implemented' (item a16) with only a minority who did not think so (3.0 percent). Similar results were obtained for the need of 'complete assessment document' (a14) or 'teaching assistant' (item a15).

Teachers were perceived as having sufficient and adequate qualifications and skills in implementing assessment activities. The majority of respondents (68.5 percent) thought that teachers were sufficiently qualified (item a17) and only minority (9.2 percent) thought otherwise and the rest (22.2 percent) were unsure about it. Concerning teachers' skills, a slightly different scenario was seen. Half of the respondents thought that skills were adequate (item a18) and the rest of the respondents, either thought that the skills are inadequate or unsure about the issue in a similar proportion. In terms of the suitability of physical infrastructure and ICT, the survey confirmed that physical equipment and the space in classroom or the ICT hardware were perceived suitable for the SBA implementation. However, half of the respondents, in similar proportion, neither agreed nor disagreed or they thought that all the three main constructs involved were not suitable. In particular, quite a high percentage of respondents, nearly one-third (30.7 percent), perceived that the physical equipment in classroom (item a19) was not suitable for SBA activities.

Table 5.6: Teachers' responses on input evaluation of SBA (frequency and percentage)

Item	Strongly disagree	Disagree	Neutral	Agree	Strongly agree	Total	Missing
a14: It is easy to implement assessment if	6 (0.8)	23 (3.0)	79 (10.2)	330 (42.5)	336 (43.3)	774 (99.7)	2 (0.3)

teachers are supplied with a complete assessment document							
a15: A teaching assistant is needed to help teachers in assessment	8 (1.0)	30 (3.9)	112 (14.4)	259 (33.4)	366 (47.2)	775 (99.9)	1 (0.1)
a16: Training of the personnel involved should be properly planned and implemented	3 (0.4)	20 (2.6)	42 (5.4)	301 (38.8)	410 (52.8)	776 (100.0)	0 (0.0)
a17: Teachers are sufficiently qualified to implement assessment activities	10 (1.3)	61 (7.9)	172 (22.2)	373 (48.1)	158 (20.4)	774 (99.7)	2 (0.3)
a18: Teachers' skills in assessment are adequate for implementing assessment activities	30 (3.9)	131 (16.9)	216 (27.8)	316 (40.7)	83 (10.7)	776 (100.0)	0 (0.0)
a19: The physical equipment in classroom is suitable to conduct SBA activities	62 (8.0)	176 (22.7)	200 (25.8)	299 (38.5)	39 (5.0)	776 (100.0)	0 (0.0)
a20: The space can be adapted to assessment activities	46 (5.9)	120 (15.5)	198 (25.5)	365 (47.0)	44 (5.7)	773 (99.6)	3 (0.4)
a21: The ICT hardware is suitable to conduct SBA activities	68 (8.8)	118 (15.2)	170 (21.9)	350 (45.1)	69 (8.9)	775 (99.9)	1 (0.1)

Process dimension of evaluation consists of several constructs. The first one discussed here is on teachers' attitude – their belief, feeling and readiness towards SBA as shown in Table 5.7. Overall, data indicated that more than half of the respondents believed that their attitude towards SBA was positive while it was discovered that higher percentages of respondents thought that they had positive beliefs towards SBA and were ready to implement SBA rather than having positive feeling towards SBA.

In terms of teachers' belief towards SBA, nearly half of the respondents thought that SBA was efficient in improving learning and helped in building independence in students. Larger numbers of respondents thought that SBA does play a role in offering more opportunities to students in terms of assessment and the role of SBA as being incorporated into existing lessons. However, nearly one-fourth of the respondents were not sure on their belief towards SBA.

Looking at their feeling towards SBA, it was quite different especially on the issue of burden. Nearly half of the respondents (51.8 percent) believed that SBA was burdening them. However, similar percentages of respondents also thought that SBA was not boring and it was compulsory for them to practise SBA in their teaching. Similar to the numbers related to belief factor, nearly one-fourth of the respondents were not sure of their feelings towards SBA. However, the highest was the case which 30.2 percent of respondents were not sure whether SBA was boring or not which were higher than the percentage of respondents (21.8 percent) who feel that SBA was boring. Similarly, more respondents were not sure whether SBA was compulsory or not (25.6 percent) compared to those who felt that SBA was not compulsory (18.7 percent). The third factor related to the teacher readiness towards SBA. The majority of respondents thought that they were highly satisfied with this components of the programme; they were ready to implement SBA, either to prepare students' files, having discussion with their friends or following instruction by MOE.

Table 5.7: Teachers' responses on process evaluation of SBA in terms of teachers' attitude towards SBA (frequency and percentage)

Item	Strongly disagree	Disagree	Neutral	Agree	Strongly agree	Total	Missing
a1: SBA is efficient in improving students learning	51 (6.6)	127 (16.4)	202 (26.0)	349 (45.0)	44 (5.7)	773 (99.6)	3 (0.4)
a2: SBA can help students to become more independent learners	36 (4.6)	138 (17.8)	203 (26.2)	350 (45.1)	47 (6.1)	774 (99.7)	2 (0.3)
a3: It is good that students have more than one opportunity to be assessed	21 (2.7)	97 (12.5)	164 (21.1)	412 (53.1)	82 (10.6)	776 (100.0)	0 (0.0)

a4: SBA can be effectively incorporated into existing lessons	18 (2.3)	79 (10.2)	174 (22.4)	431 (55.5)	74 (9.5)	776 (100.0)	0 (0.0)
a5: SBA is not burdening me	169 (21.8)	233 (30.0)	176 (22.7)	164 (21.1)	33 (4.3)	775 (99.9)	1 (0.1)
a6: SBA is not boring	49 (6.3)	120 (15.5)	234 (30.2)	329 (42.4)	41 (5.3)	773 (99.6)	3 (0.4)
a7: SBA is compulsory	45 (5.8)	100 (12.9)	199 (25.6)	326 (42.0)	102 (13.1)	772 (99.5)	4 (0.5)
a8: I prepare students' individual files for my subject before conducting SBA	15 (1.9)	57 (7.3)	131 (16.9)	421 (54.3)	149 (19.2)	773 (99.6)	3 (0.4)
a9: I have plenty of opportunities to discuss SBA implementation with my colleagues who are teaching the same subject with me	27 (3.5)	130 (16.8)	214 (27.6)	326 (42.0)	78 (10.1)	775 (99.9)	1 (0.1)
a10: I always follow the planning instructed by the MOE in SBA implementation	8 (1.0)	54 (7.0)	192 (24.7)	397 (51.2)	125 (16.1)	776 (100.0)	0 (0.0)

Following teachers' attitude towards SBA, there were also several factors that contribute to the process dimensions of evaluation which are i) teachers' understanding, ii) effect of SBA courses on skills, iii) IHT, iv) encouragement by administration, v) moderation (coordination) process, and vi) monitoring process in Table 5.8.

The majority of the respondents believed that teachers understood the main idea about SBA, its function, when to implement and how it is produced. Similarly, a large number of respondents believed that courses on SBA had upgraded their skills on SBA. But still, nearly one-fourth of the respondents were unsure about that. Looking at their perspective on the IHT that is conducted in their schools, more than half of the respondents agreed that the IHT in their schools was conducted accordingly – conducted by the experts several times and cover both aspects, theoretical and practical forms. Similarly, nearly one-fourth of respondents still were unsure about those matters. On the administration side, not many respondents (41.2 percent) agreed with the

provision of reinforcement or special recognition to teachers by the administrators. A relatively high percentage of respondents were unsure about this, more than one-third of the respondents. The next point focused on the moderation and the monitoring process. Most respondents (more than three quarter) agreed that the moderation process was implemented – students' scores and evidence were prepared and marks were recorded. Few respondents were unsure about moderation process, only less than 15.0 percent of them. Looking at the monitoring process, the majority of the respondents understood the process; however the percentage was a bit less compared to moderation process. Only one item - item b28iii, looking at how the 'evidence of centre assessment are handled' was perceived a bit differently by the respondents. A large number of respondents (nearly 80.0 percent) agreed about it and only 5.1 percent disagreed.

Table 5.8: Teachers' responses on process evaluation of SBA (frequency and percentage)

Item	Strongly disagree	Disagree	Neutral	Agree	Strongly agree	Total	Missing
a11: Educational assessment transformation through SBA leads to changes from examination-oriented assessment to an assessment which is more integrated	28 (3.6)	95 (12.2)	181 (23.3)	369 (47.6)	102 (13.1)	775 (99.9)	1 (0.1)
a12: SBA is conducted during the teaching and learning process	10 (1.3)	54 (7.0)	122 (15.7)	466 (60.1)	120 (15.5)	772 (99.5)	4 (0.5)
a13: For centre assessment in SBA, assignments are prepared by the Examination Board	8 (1.0)	36 (4.6)	190 (24.5)	351 (45.2)	181 (23.3)	766 (98.7)	10 (1.3)
b24i: to assess students using the assessment instrument	12 (1.8)	39 (6.0)	132 (20.3)	412 (63.4)	55 (8.5)	650 (100.0)	0 (0.0)
b24ii: to assess students documentation following stated criteria in the assessment document	2 (0.2)	59 (9.1)	122 (18.8)	437 (67.2)	30 (4.6)	650 (100.0)	0 (0.0)
b24iii: to explain assessment criteria in	4 (0.6)	70 (10.8)	148 (22.8)	386 (59.4)	42 (6.5)	650 (100.0)	0 (0.0)

details to students							
b24iv: to conduct assessment activities effectively	2 (0.3)	51 (7.8)	125 (19.2)	427 (65.7)	45 (6.9)	650 (100.0)	0 (0.0)
b25i: IHT is conducted by the experts from the ministry or the panels of teachers	26 (4.0)	66 (10.2)	142 (21.8)	372 (57.2)	44 (6.8)	650 (100.0)	0 (0.0)
b25ii: IHT has been conducted more than once to increase teachers' understanding	45 (6.9)	109 (16.8)	156 (24.0)	297 (45.7)	43 (6.6)	650 (100.0)	0 (0.0)
b25iii: Training includes practical forms not only theoretical	32 (4.9)	139 (21.4)	170 (26.2)	265 (40.8)	41 (6.3)	647 (99.5)	3 (0.5)
b26i: Administrators provide reinforcement such as giving gifts or praise to those teachers that practice SBA in their classrooms	97 (12.5)	143 (18.4)	207 (26.7)	250 (32.2)	70 (9.0)	767 (98.8)	9 (1.2)
b26ii: Administrators give special recognition to teachers performance for conducting SBA	89 (11.5)	144 (18.6)	213 (27.4)	243 (31.3)	77 (9.9)	766 (98.7)	10 (1.3)
b27i: Each teacher prepares students' scores for his/her subjects	23 (3.0)	50 (6.4)	115 (14.8)	487 (62.8)	96 (12.4)	771 (99.4)	5 (0.6)
b27ii: Each teacher prepares evidence according to students' achievement	16 (2.1)	35 (4.5)	110 (14.2)	494 (63.7)	116 (14.9)	771 (99.4)	5 (0.6)
b27iii: Each teacher records the marks of evidence in SPPBS application	20 (2.6)	29 (3.7)	115 (14.8)	474 (61.1)	134 (17.3)	772 (99.5)	4 (0.5)
b28i: The school SBA committee appoints certain teachers as internal monitors	17 (2.2)	70 (9.0)	161 (20.7)	432 (55.7)	86 (11.1)	766 (98.7)	10 (1.3)
b28ii: Heads of panel of each subject monitor SBA activity implementation	22 (2.8)	100 (12.9)	180 (23.2)	402 (51.8)	64 (8.2)	768 (98.9)	8 (1.0)
b28iii: Evidence of centre assessment is kept in a systematic way in a safe place before it is returned to the students	18 (2.3)	22 (2.8)	105 (13.5)	476 (61.3)	147 (18.9)	768 (98.9)	8 (1.0)

Next, the process dimension of evaluation included challenges faced by the teachers in implementing SBA. There are twelve main challenges altogether included in this study as listed in Table 5.9. In general, survey data showed that most respondents (over 50 percent of the total respondents) agreed that all factors seem to challenge the implementation except for the two factors – getting support from head teachers and getting cooperation from teachers, which showed a very low percentage of respondents who thought that as a challenge. Interestingly, over half of the respondents thought that it was easy to gain support and cooperation from the head teacher and their peers. The most challenging factor in SBA was the extra work loaded to the teachers with 70.0 percent of the total respondents in the survey 'agreed' and 'strongly agreed' with it. 'No special recognition for teachers' performance' was rated as the second most challenging by 62.0 percent of respondents. 'Problem to get related resources' and 'insufficient knowledge' were considered challenging by 56 percent of the respondents for both cases. It is worth pointing out here that there were generally similar numbers of respondents who agreed, disagreed and were unsure in perceiving lack of confidence and school climate as a challenge to them.

Table 5.9: Teachers' responses on process evaluation of SBA in terms of challenges faced (frequency and percentage)

Item	Strongly disagree	Disagree	Neutral	Agree	Strongly agree	Total	Missing
c1: Difficult to gain support from the head teacher	125 (16.1)	287 (37.0)	215 (27.7)	114 (14.7)	31 (4.0)	772 (99.5)	4 (0.5)
c2: Difficult to gain cooperation from teachers	100 (12.9)	339 (43.7)	176 (22.7)	122 (15.7)	36 (4.6)	773 (99.6)	3 (0.4)
c3: Insufficient knowledge to implement SBA process	53 (6.8)	106 (13.7)	183 (23.6)	323 (41.6)	110 (14.2)	775 (99.9)	1 (0.1)
c4: Lack of confidence to conduct SBA	48 (6.2)	169 (21.8)	241 (31.1)	250 (32.2)	65 (8.4)	773 (99.6)	3 (0.4)
c5: Problem to get related resources on SBA	45 (5.8)	129 (16.6)	159 (20.5)	302 (38.9)	135 (17.4)	770 (99.2)	6 (0.8)
c6: School climate	72	219	229	178	76	774	2

seems to hinder SBA implementation	(9.3)	(28.2)	(29.5)	(22.9)	(9.8)	(99.7)	(0.3)
c7: Insufficient financial resources	51 (6.6)	143 (18.4)	177 (22.8)	254 (32.7)	147 (18.9)	772 (99.5)	4 (0.5)
c8: Extra workload	33 (4.3)	64 (8.2)	134 (17.3)	267 (34.4)	276 (35.6)	774 (99.7)	2 (0.3)
c9: No special recognition for teachers performance for conducting SBA	27 (3.5)	83 (10.7)	184 (23.7)	317 (40.9)	162 (20.9)	773 (99.6)	3 (0.4)
c10: Students and parents may not trust teachers' assessment in SBA	38 (4.9)	113 (14.6)	218 (28.1)	272 (35.1)	133 (17.1)	774 (99.7)	2 (0.3)
c11: Students may memorize responses and reproduce them during SBA sessions	32 (4.1)	124 (16.0)	188 (24.2)	297 (38.3)	132 (17.0)	773 (99.6)	3 (0.4)
c12: The existing number of teachers is not sufficient to implement SBA activities	33 (4.3)	133 (17.1)	187 (24.1)	260 (33.5)	161 (20.7)	774 (99.7)	2 (0.3)

Lastly is the process dimension of evaluation related to school improvement. As shown in Table 5.10, there are two main constructs involved – 'role of SBA' and the 'importance of SBA' in school improvement. Looking at 'role of SBA', generally it clearly showed that the number of respondents who 'agreed', 'neutral' or 'disagreed' were similar (around one-third of the respondents), except for the item d30iv, 'SBA provides new learning opportunities for building on students' strengths and interests' which showed a higher percentage. For this particular item, nearly half of the respondents thought that SBA provided new learning opportunities. Another interesting point to make is that nearly half of the respondents felt that absentee rate reduction and public satisfaction increment were not affected by SBA implementation and nearly one-third of the respondents were not sure about them. Concerning 'importance of SBA', a higher percentage of the respondents chose 'agree' and 'strongly agree' as compared to 'disagree' and 'strongly disagree'. And, nearly one-third of the respondents perceived that they were unsure whether SBA could influence teachers on their willingness to take decisions related to assessment reform (item d31vi) or reduce pressure in public examination (item d31vii).

Table 5.10: Teachers' responses on process evaluation of SBA in terms of school improvement process (frequency and percentage)

Item	Strongly disagree	Disagree	Neutral	Agree	Strongly agree	Total	Missing
d30i: SBA reduces absentee rate	77 (9.9)	241 (31.1)	240 (30.9)	184 (23.7)	34 (4.4)	776 (100.0)	0 (0.0)
d30ii: SBA contributes strongly to the quality assurance process	60 (7.7)	166 (21.4)	261 (33.6)	256 (33.0)	33 (4.3)	776 (100.0)	0 (0.0)
d30iii: SBA increases public satisfaction and confidence towards school	65 (8.4)	217 (28.0)	280 (36.1)	189 (24.4)	25 (3.2)	776 (100.0)	0 (0.0)
d30iv: SBA provides new learning opportunities for building on students' strengths and interests	42 (5.4)	130 (16.8)	245 (31.6)	323 (41.6)	34 (4.4)	774 (99.7)	2 (0.3)
d31i: SBA helps students improve generic skills to build on their strengths and interests	27 (3.5)	132 (17)	201 (25.9)	365 (47.0)	45 (5.8)	770 (99.2)	6 (0.8)
d31ii: SBA helps teachers to improve students assessment practices	23 (3.0)	117 (15.1)	168 (21.6)	424 (54.6)	41 (5.3)	773 (99.6)	3 (0.4)
d31iii: SBA helps teachers to actively develop their professionalism	25 (3.2)	101 (13.0)	174 (22.4)	424 (54.6)	51 (6.6)	775 (99.9)	1 (0.1)
d31iv: SBA helps teachers to develop students various learning opportunities	22 (2.8)	95 (12.2)	165 (21.3)	431 (55.5)	62 (8.0)	775 (99.9)	1 (0.1)
d31v: SBA makes teachers and students willing to discuss learning problems in non-threatening manners	38 (4.9)	129 (16.6)	175 (22.6)	381 (49.1)	53 (6.8)	776 (100.0)	0 (0.0)
d31vi: SBA makes teachers willing to take decisions on assessment reform	28 (3.6)	114 (14.7)	226 (29.1)	372 (47.9)	36 (4.6)	776 (100.0)	0 (0.0)

d31vii: SBA reduces pressure in public examinations	36 (4.6)	115 (14.8)	213 (27.4)	341 (43.9)	71 (9.1)	772 (99.5)	4 (0.5)
d31viii: SBA improves literacy and numeracy learning	36 (4.6)	99 (12.8)	194 (25.0)	398 (51.3)	48 (6.2)	775 (99.9)	1 (0.1)

The last dimension of evaluation is product or outcome dimension. In this study, it is assessing the students' positive attitude towards SBA, their good knowledge in SBA and their motivational towards learning as shown in Table 5.11. In general, the survey results suggested that SBA implementation has, to some extent, reached the target audience, showing a positive result in terms of all the three constructs. However, poorer results were shown in terms of motivating students towards learning, especially for the item e34i – 'SBA encourages students to read more books than before' which showed that most respondents (37.0 percent) felt that SBA did not encourage students to read more books than before. In terms of students' interest in subject and their understanding on their strengths and weaknesses, there were generally very few 'strongly agree' but also very few who responded with 'strongly disagree' responses on these two items.

Table 5.11: Teachers' responses on product evaluation of SBA (frequency and percentage)

Item	Strongly disagree	Disagree	Neutral	Agree	Strongly agree	Total	Missing
e32i: Students practice SBA in their study	39 (5.0)	183 (23.6)	207 (26.7)	321 (41.4)	25 (3.2)	775 (99.9)	1 (0.1)
e32ii: The students are motivated to prepare for SBA.	47 (6.1)	217 (28.0)	209 (26.9)	268 (34.5)	34 (4.4)	775 (99.9)	1 (0.1)
e32iii: Practising SBA in their study is NOT burdening them at all	46 (5.9)	159 (20.5)	224 (28.9)	298 (38.4)	48 (6.2)	775 (99.9)	1 (0.1)
e33i: SBA tasks are part of teaching and learning of their respective subjects	48 (6.2)	107 (13.8)	170 (21.9)	390 (50.3)	58 (7.5)	773 (99.6)	3 (0.4)
e33ii: Feedback from teachers helps them develop skills that may not be	47 (6.1)	131 (16.9)	168 (21.6)	372 (47.9)	54 (7.0)	772 (99.5)	4 (0.5)

reflected in public examinations							
e34i: SBA encourages students to read more books than before	76 (9.8)	211 (27.2)	238 (30.7)	231 (29.8)	20 (2.6)	776 (100.0)	0 (0.0)
e34ii: The students are becoming more interested in my subject than before	53 (6.8)	148 (19.1)	279 (36.0)	257 (33.1)	39 (5.0)	776 (100.0)	0 (0.0)
e34iii: SBA helps students to understand more on their strengths and weakness in each subject	62 (8.0)	172 (22.2)	205 (26.4)	299 (38.5)	38 (4.9)	776 (100.0)	0 (0.0)

Overall, the descriptive showed that the SBA system has a positive input, process and product dimension, although the elements of this evaluation dimension cannot demonstrate them conclusively. Survey data demonstrated that the teachers consistently rated above the midpoint of the scale for the factors observed, indicating quite a high level concerning the effectiveness of the system implementation.

To summarise this section of the study regarding the findings of the descriptive analysis, the respondents were quite satisfied with input dimension especially with qualification and skills of teachers and also the ICT hardware supplied and space provided, but not with the physical equipment provided in the classroom. They also felt that they really needed proper training, a complete assessment document and teaching assistants. Regarding process dimension, overall, teachers had a positive attitude towards SBA, but that they felt SBA was burdensome. They understood the main goals of SBA and the purpose of the in-house training. Administration as a side factor was not perceived as very helpful indeed. Teachers supported moderation and monitoring process. The most challenging issues in SBA were the issue of workload or time-constraint followed by no special recognition and not enough resources provided. In terms of the process of improving school, SBA was seen as a medium in developing teachers' professionalism and assessment practises and in assisting students in terms of improving their generic skills and literacy and numeracy skills, and also to reduce pressure towards public examinations. However, SBA did not really help in reducing absentee rate and increasing

public satisfaction towards school. Finally, regarding product dimension, which focused on student outcome, it showed that the respondents were satisfied with the attitude and knowledge of the students but were not so satisfied with the students' motivational level, especially with their reading interest.

5.8 Analysis and Results of Measurement (CFA) Models: Stage 1

Before continuing with the development of measurement models, it would be beneficial to look at the details related to SEM, such as estimation method, goodness-of-fit indices, unidimensionality, validity and reliability used in this study.

a) Estimation method

In this study, maximum likelihood (ML) estimation method is used. ML is an estimation method used in generating parameter estimates of structural equation models. It is an estimation method that works simultaneously where estimates of model parameters for all variables are calculated all at once (Kline, 2005). Being the most widely used, it is an iterative procedure that seeks to minimise a discrepancy between the model and the sample covariance. It assumes that the raw data to be analysed has no missing values and the population distribution for the variables is multivariate normal. Furthermore, with ML estimation, estimates from a given sample are considered to be "random but known" whereas the true values of the model parameters are considered to be "fixed but unknown" (Byrne, 2010). The good thing about this estimation method is its ability to deal with complex models and its robustness to non-normality data (Brown, 2006).

b) Goodness-of-Fit Index Assessment

Goodness-of-fit indices are fit indices of SEM. The values represent "how well the specified model reproduces the covariance matrix among the indicator items" (Hair *et al.,* 2006, p. 745). There are many measures developed but only some will be utilised in this study, as shown in Table 5.12 (Hooper *et al.,* 2008;

Kline, 2005; Hair *et al.,* 2006). In general, there are three groups of measures: practical fit measures, absolute fit measures and incremental fit measures. It is suggested that the study should report at least three fit indexes with at least one from each category (Hair *et al.,* 2010). Chi-square statistics (X^2), also known as CMIN (minimum discrepancy), is a value representing the discrepancy between the unrestricted sample covariance matrix and the restricted covariance matrix (Byrne, 2010). But, it has limitations. With large sample sizes, it is common to get a large X^2 value relative to degrees of freedom. So, large samples mean poor fit. However, with small samples, the Chi-square statistics lack power and yet could not discriminate between good and poor fitting models (Hooper *et al.,* 2008). A fitness index developed to minimise the sensitivity of sample size on the Model Chi-Square is normed chi-square (X^2/df) and it is grouped under practical, subjective or ad hoc indices of fit. Normed chi-square is a value representing value of X^2 divided by the degrees of freedom resulting in a lower value (Kline, 2005). And, the recommended value for a reasonable fit model is between 2.0 and 5.0.

Next is an absolute fit index which measure of how well the hypothesized model fits in comparison with no other models (Byrne, 2010). They estimate the proportion of variability in the sample covariance matrix explained by the predicted covariance matrix in the model (Kline, 2005). Included in this category are GFI (Goodness-of-Fit), AGFI (Adjusted GFI) and RMSEA (Root Mean Square Error of Approximation) (Hooper *et al.,* 2008). GFI is an index measuring the amount of covariance between the latent variables in the model (Kline, 2005). For data with low factor loading and sample size, the cut-off is suggested to be 0.95 instead of 0.9 (Hooper *et al.,* 2008). Furthermore, with a large number of degree of freedom compared to the number of sample size, its value tends to have a downward bias. Recently, this index is becoming less popular due to its sensitivity. It is even recommended not to be used. AGFI also adjusts the GFI based upon degrees of freedom in the specified model. There will be greater reduction of values of GFI when it comes to more complex models (Kline, 2005). However, AGFI is becoming less popular nowadays as it is not well-performed in some computer simulation study. RMSEA is an index measuring the discrepancy between the observed and estimated covariance matrices per degree of freedom (Hoe, 2008). It is not sensitive to sample size as the discrepancy measured is in terms of the population and not the sample.

However, it is quite sensitive to the number of estimated parameters in the model (Byrne, 2010) and it favours models with the lesser number of parameters (Hooper *et al.,* 2008).

Next are incremental or comparative fit indices, NFI (Normed Fit Index), CFI (Comparative Fit Index) and TLI (Tucker Lewis Index). These fit indices compare the chi square value of the hypothesised model against the chi square value of some standard, such as the null model. The null model assumes zero population co-variances among the observed variables (Kline, 2005). In other words, the null model assumes that all measured variables are uncorrelated. Since incremental indices measure the increase in fit relative to a null model, higher value of the indices indicate larger improvement in fit (Lei and Wu, 2007). CFI is actually a revised form of NFI. When choosing the index, CFI should be given priority compared to NFI, as suggested by Bentler in Byrne (2010) because NFI tends to underestimate the fit of a small sample sized model (Kline, 2005) whereas CFI is the least effected by sample size (Hooper, Coughlan and Mullen, 2008). With NFI, the fit of a sample size less than 200 will be underestimated. CFI is also the most widely used due to its relative insensitivity to model complexity (Hair *et al.,* 2006). Furthermore, CFI provides a measure of complete co-variation in the data (Byrne, 2010). With regards to which indices should be reported, Hooper *et al.* (2008) believe that CFI and RMSEA are preferred over other indices for they are the most insensitive to sample size, model misspecification and parameter estimates.

Table 5.12: Goodness-of-fit Indices

Goodness-of-fit Index	Acceptable Value	Comments
Chi-square (X^2)	$p > 0.05$ (non-significant)	Indicates exact fit of the model. Value is sensitive to large sample size
Practical indices of fit:		
Normed chi-square (X^2/df)	[2.00, 5.00]	This is to reduce the sensitivity of X^2 to sample size $X^2/df < 3.0$: good fit
Absolute fit index:		
The Goodness-of-Fit Index(GFI)	[0.00, 1.00]	GFI = 1.00: perfect fit GFI > 0.9: good fit
The Adjusted GFI (AGFI)	[0.00, 1.00]	Values close to 1.00: good fit GFI > 0.9: good fit

Root Mean Square Error of Approximation (RMSEA)	RMSEA ≤ 0.08	RMSEA < 0.05: good fit RMSEA 0.05 - 0.08: adequate fit Values up to 0.10: poor fit
Incremental fit indices:		
Normed Fit Index (NFI)	NFI ≥ 0.90	NFI = 1.00: perfect fit Values close to 0.00: poor fit
Comparative Fit Index (CFI)	CFI ≥ 0.90	0.00 > CFI > 1.00 for acceptance
Tucker-Lewis Index (TLI)	TLI > 0.90	0.0 > TLI > 1.00 for acceptance

c) Uni-dimensionality

For items to be uni-dimensional, they have to be strongly associated with each other and they load highly on a single factor to represent a single construct (Hair *et al.,* 2006). Uni-dimensionality of a measurement model is achieved when the factor loading of items for the respective latent construct is 0.5 or more (Zainuddin, 2012). If the value is less than 0.5, then the process such as 'the item deletion process' or setting 'the free parameter estimate process' is conducted.

d) Validity

Validity of a measurement model is the ability to measure what is supposed to measure for a construct in a measurement model (Zainuddin, 2012). It includes convergent validity, construct validity and discriminant validity. Convergent validity assesses the degree to which two measures of the same concept are correlated (Hair *et al.,* 2006, p.137). In other words, all the items that indicate a specific construct should share a high proportion of variance in common. It could be verified using factor loading value as a high value indicates that the item converge on the same latent factor. To get a high convergent validity, factor loadings should be statistically significant and the value should be 0.5 or higher. The best is more than 0.7 as the square of standardised factor loading represents how much variation in an item is explained by the latent factor. Or, it could also be verified through average variance extracted (AVE), and to achieve adequate convergence the value of AVE should be 0.5 or higher (Zainuddin, 2012). This high correlation value indicated that the scale is measuring the intended concept. Construct validity is the extent to which a set of items actually reflect the theoretical latent construct those items are designed to measure (Hair *et al.,* 2006). It could be verified through the fitness

indices value (Zainuddin, 2012). Discriminant validity means that individual measured items should represent only one latent construct (Hair et al., 2012, p. 778). It could be achieved when the measurement model is free from redundant items. This is checked from the modification indices value and then followed by the deletion of item or by setting the correlated pair of items as 'free parameter estimates'. And, to look for distinctiveness between constructs, the correlation between constructs should be less than 0.85 (Kline, 2005). The requirements of all the validity are shown in Table 5.13.

Table 5.13: Requirement for the validity of the measurement model

Type of validity	Requirement
Convergent validity	Average Variance Extracted (AVE) ≥ 0.5 $AVE = \Sigma K^2 / n$ (K=factor loading of every item; n=number of items in a model)
Construct validity	Fitness index is achieved as: GFI ≥ 0.90; CFI ≥ 0.90; RMSEA ≤ 0.08; Chisq/df < 5.0
Discriminant validity	Free from redundant items Correlation coefficient between each pair of latent construct ≤ 0.85

e) Reliability

Reliability of a measurement model is a measure of the degree to which a set of indicators of a latent construct is internally consistent in their measurements (Hair *et al.,* 2006). A reliable model does not guarantee that the model is valid. However, reliability could be an indicator of convergent validity (Hair *et al.,* 2006). In this study, three types of reliability were employed. They are internal reliability, construct reliability (CR) and average variance extracted (AVE). Internal reliability refers to the degree to which all the items are measuring the same underlying construct (Pallant, 2007). It is measured by Cronbach's alpha coefficient and it is quite sensitive to the number of items measured. Construct reliability (CR) is intended to determine the consistency of the items representing a latent construct (Zainuddin, 2012). To achieve CR, a value of square of total standardised loading divided by the sum of a square of total standardised loading and measurement error (Said *et al.,* 2011) and the value should be 0.6 or more to represent the existence of internal consistency (Hair *et al.,* 2006). Average variance extracted (AVE) refers to the average percentage of variation explained by the items in a construct (Zainuddin, 2012). A value of 0.5 or higher indicates the items share a high proportion of

variance in common whereas a value less than 0.5 indicates that on average, more error remains in the items than variance explained by the latent factor structure imposed on the measure (Hair *et al.,* 2006). The requirements of all the reliability are shown in Table 5.14.

Table 5.14: Requirement for the reliability of the measurement model

Type of reliability	Requirement
Internal reliability	Cronbach Alpha ≥ 0.7
Construct reliability (CR)	CR ≥ 0.6 $CR = (\Sigma K)^2 / [(\Sigma K)^2 + (\Sigma 1-K^2)]$ (K=factor loading of every item)
Average Variance Extracted (AVE)	AVE ≥ 0.5

Next, the development of each measurement model was discussed. Pooled latent constructs (measurement model) are formed. Few constructs were combined simultaneously as there were some constructs with less than four items. Three measurement models were considered – input, process and product dimensions of evaluation. Most models were modelled as a second-order structure following the underlying theory, when there was a higher level factor which was considered accountable for the lower order factors (Byrne, 2010) and then, they were assessed for uni-dimensionality, validity and reliability of the models.

5.8.1 Input Dimension

Input dimension of evaluation consists of three constructs – material and personal needs in SBA ('mat'), appropriateness of personnel's qualification ('appr') and suitability of physical infrastructure and ICT ('suit'). A total of eight items were used to measure the three latent constructs – three items for 'mat', two items for 'appr' and three items for 'suit'. This second-order model with input dimension, a single second-order factor was hypothesised as accounting for all variance and covariance related to the first-order factors as shown in Figure 5.2.

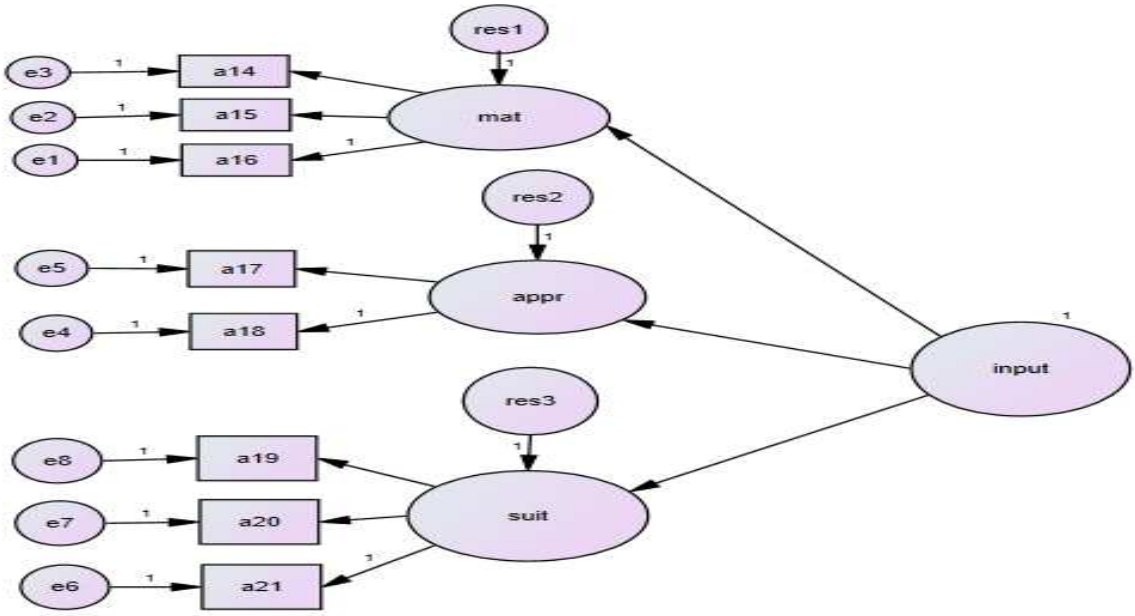


Figure 5.2: Hypothesised second-order model of factorial structure for the input dimension

Several aspects of this second-order model were:

i) the second-order factor loadings were to be freely-estimated;
ii) neither factor co-variances nor variances were to be estimated; and
iii) a residual error term was associated with each of the first-order factors because the prediction of each of the first-order factor from the second-order factor was presumed not to be without error. Each of the first-order factor was now a dependent variable in the model.

This CFA model hypothesized:

i) responses to the input dimension can be explained by three first-order factors and one second-order factor;
ii) each item had a nonzero loading on the targeted first-order factor and zero loadings on the other two non-targeted factors;
iii) error terms associated with each item were uncorrelated; and
iv) co-variation among the three first-order factors was explained fully by their regression on the second-order factor.

The hypothesized model identification status was ascertained by checking the number of degrees of freedom associated with the model. Referring to the hypothesised model in Figure 5.2, there were 36 pieces of information contained in the sample variance-covariance matrix and 19 parameters to be

estimated, thereby leaving 17 degrees of freedom as shown in Table 5.15. Thus, it was concluded that this model was identified with 17 degrees of freedom. The model was over-identified but with hierarchical models, the higher-order structure would be just-identified.

Table 5.15: AMOS Output for the hypothesised model for summary statistics

Computation of degrees of freedom	
Number of distinct sample moments	36
Number of distinct parameters to be estimated	19
Degrees of freedom	17
Result	
Chi-square	157.756
Degrees of freedom	17
Probability level	0.000

In reviewing the goodness-of-fit statistics, it showed that X^2/df=9.280; GFI=0.952; AGFI=0.898; NFI=0.928; CFI=0.935; TLI=0.892 and RMSEA=0.103. This measurement model provided a poor fit. However, in order to resolve just-identification issue in this second-order model, equality constraints were placed on particular parameters (appropriate residual terms) at the upper level to yield estimates that were more accurate (Arbuckle and Wothke, 1999). This additional constraints technique could increase degrees of freedom greatly but with a small increase in the chi-square statistic. The CRDIFF (critical ratio difference) method was conducted. Then, estimates were calculated.

Table 5.16 showed the critical ratio difference between the two parameters. The two prime candidates for the imposition of equality constraints were the higher order residuals related to the second ('appr') and the third construct ('suit') as their estimated values were non-significant (1.94<1.96). If the value was less than 1.96, the hypothesis that the two regression weights were equal in the population was not rejected at the 0.05 level (Arbuckle and Wothke, 1999). So, it seemed reasonable to constrain variances of the residuals associated with second and third construct to be equal. And, the higher order level of the model would be over-identified with one degree of freedom. The variance would be estimated for res2, and the same value held constant for res3. The degrees of freedom for the model as a whole should now increase from 17 to 18. This caused the hypothesided model to be assigned with these

equality constraints to the two factor residuals associated with the first-order factors as in Figure 5.3.

Table 5.16: Critical ratios for differences between parameters

	Par_9	Par_10	Par_11
Par_9	0.000		
Par_10	-2.406	0.000	
Par_11	-5.433	1.940	0.000

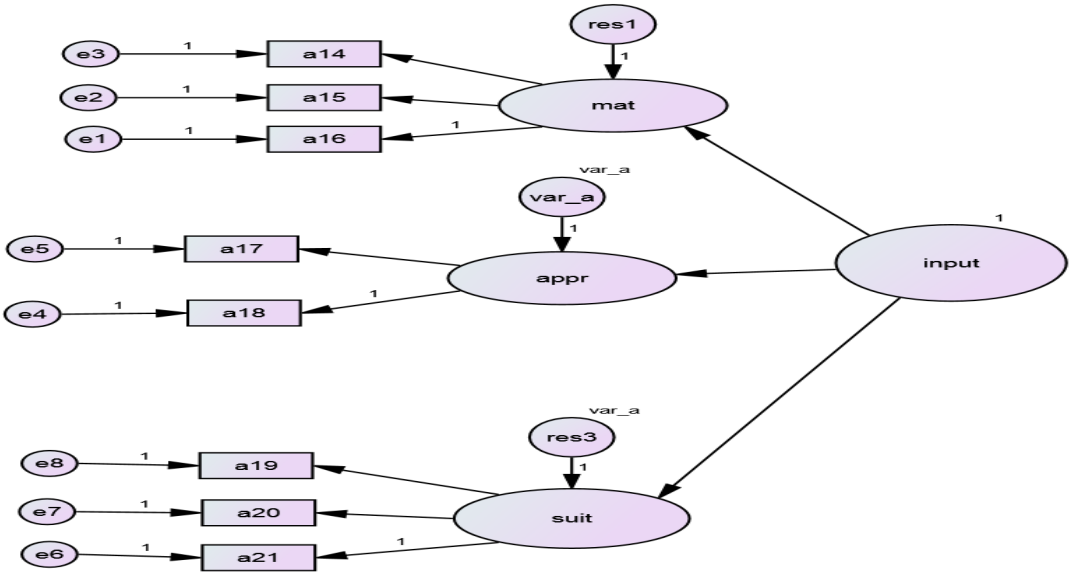


Figure 5.3: Hypothesised second-order model with residual variances for second and third construct constrained equal

Analyses were based on this re-specified model shown in Figure 5.3. It showed that X^2/df=9.931; GFI=0.946; AGFI=0.892; NFI=0.918; CFI=0.925; TLI=0.884 and RMSEA=0.107. So, these values showed that the hypothesised model did not really fit the data very well. But then, looked at the values of estimates. As shown in Table 5.17, all estimates were found to have critical ratio values more than 1.96 for a regression weight, thereby indicating their statistical significance at the 0.05 levels.

Table 5.17: AMOS Output for Hypothesized Model: Unstandardized Parameter Estimates

			Estimate	S.E.	C.R.	P	Label
mat	<---	input	0.139	0.030	4.609	***	par_7
appr	<---	input	0.784	0.041	19.027	***	par_8
suit	<---	input	0.384	0.039	9.961	***	par_9
a16	<---	mat	1.000				
a15	<---	mat	0.768	0.061	12.578	***	par_2
a14	<---	mat	0.748	0.057	13.072	***	par_3
a18	<---	appr	1.000				
a17	<---	appr	0.580	0.050	11.688	***	par_4
a21	<---	suit	1.000				
a20	<---	suit	1.829	0.152	11.994	***	par_5
a19	<---	suit	1.971	0.164	12.000	***	par_6

Since the value of the standardized estimate of 'mat' construct was so low (0.201), 'mat' construct was deleted from the model. 'Mat' refers to the material and personal needs in SBA which includes a complete assessment document, a teaching assistant and training of the personnel. In this case, the very low value of the estimate might be due to data was from categorical indicators and they were skewed. Looking at the distributions of responses from the item a14, a15 and a16, it clearly showed that each item was strongly skewed although the multivariate normality for this construct showed a value of 23.910 which was quite acceptable. Then, the model was re-specified and run. With the new model, as shown in Figure 5.4, an examination of the unstandardised solution revealed all estimates to be both reasonable and statistically significant; all standard errors appeared also to be in good order and all variances were positive. All estimates of the standardised solution were more than 0.5. The goodness-of-fit statistics showed that X^2/df=9.038; GFI=0.982; AGFI=0.934; NFI=0.974; CFI=0.977; TLI=0.943 and RMSEA=0.102. This model was still not fit.

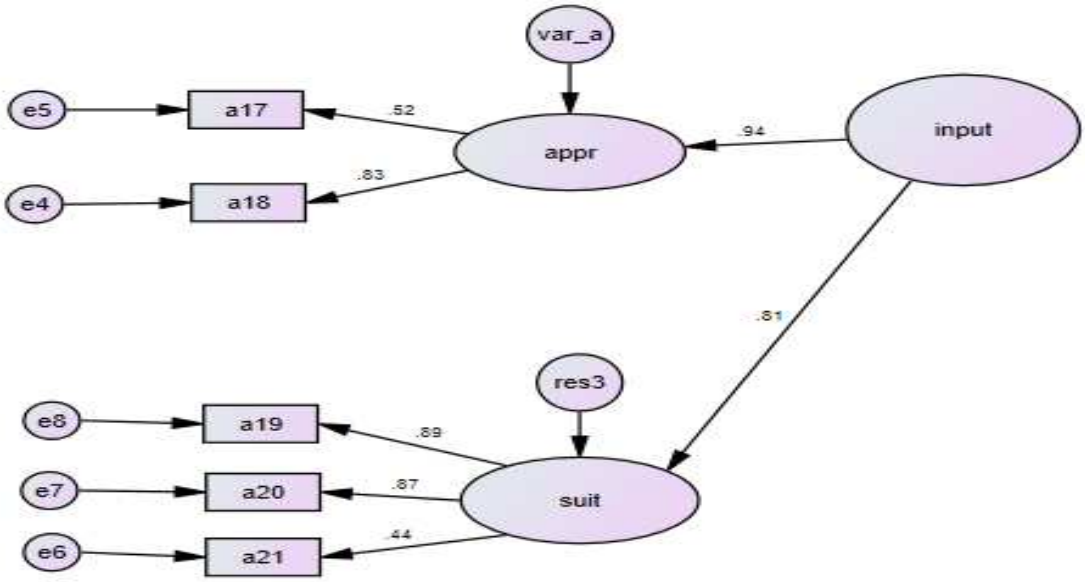


Figure 5.4: Hypothesised second-order model

As a consequence, modification indices (MIs) were examined in order to correlate the measurement error. But Arbuckle and Wothke (1999, p.153) asserted that a modification must only be considered if it makes a theoretical or a common sense to avoid producing incorrect models or models with absurd parameter estimates. MIs could improve any model by increasing the number of parameters and yet, decreasing the chi-square statistic. In reviewing the MIs related to the co-variances, there was one value of estimates which was 13.68. This related to co-variation between the error terms associated with item a21 and item a19. This value was a conservative estimate of the decrease in chi-square if e6 and e8 were allowed to correlate. If the additional parameter were indeed added, then the degrees of freedom associated with the new model would be one less. The new chi-square statistic would have 3 (=4-1) degrees of freedom, and would be no greater than 22.471(*i.e.*, 36.151 – 13.68). The actual decrease of the chi-square statistics might be much larger than 13.68. And, as indicated by the reported parameter change statistics, incorporation of the parameter into the model would result in AMOS to estimate that the covariance between e6 and e8 would be -0.084. The overlap between item a19 ('The physical equipment in classroom is suitable to conduct SBA activities') and item a21 ('The ICT hardware is suitable to conduct SBA activities') was when teachers might feel that ICT hardware can also be considered as physical equipment. This obvious overlap of item content

suggested that the freely estimated parameter should be added to the model as shown in Figure 5.5.

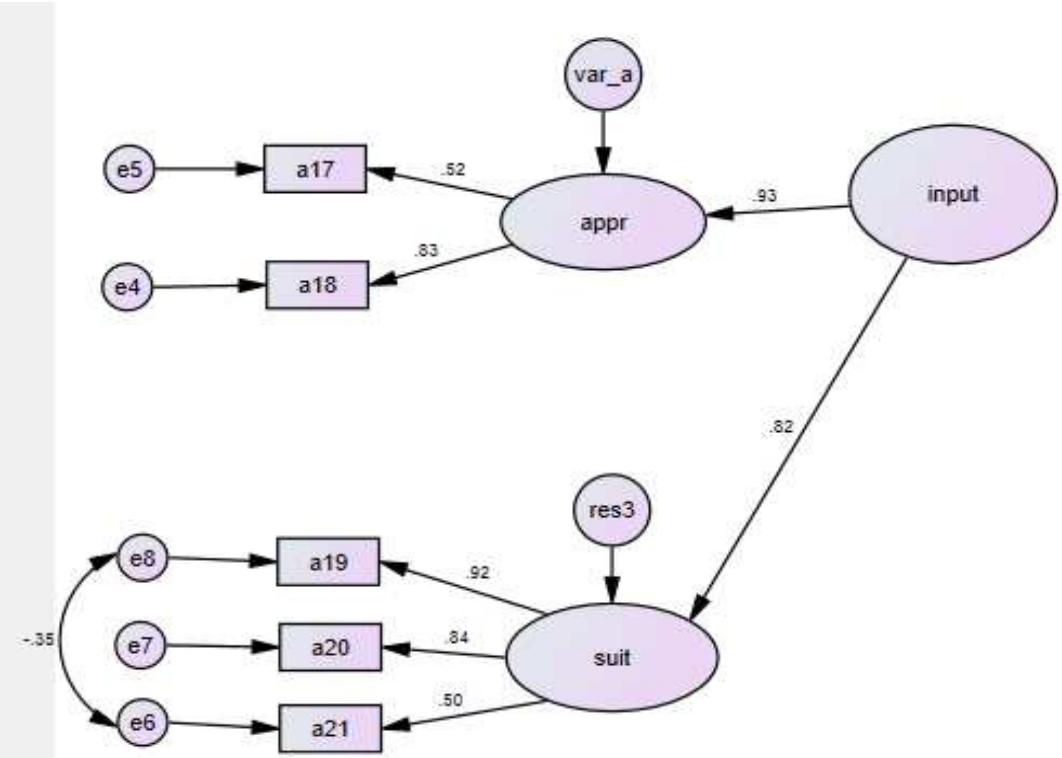


Figure 5.5: The final re-specified model of factorial structure for input

The incorporation of the error covariance or in other words, by allowing e6 and e8 to be correlated had made a substantially large improvement to model fit. The magnitude of the factor loadings was substantially significant. For this final measurement model, it showed that X^2/df=4.500; GFI=0.993; AGFI=0.966; NFI=0.990; CFI=0.992; TLI=0.975 and RMSEA=0.067 indicating a good fit. The fit indices indicated that this two-factor second-order model schematically portrayed in Figure 5.5 was the most optimal model representing input dimension of evaluation. In other words, the hypothesized three-factor CFA model did not fit the sample data, instead it was a two-factor model. In addition, Byrne (2010) asserted that assessment of model adequacy should be based on theoretical, statistical and practical considerations. Looking at Table 5.18, the value of skewness indicated that every item was approximately normally distributed. The value of kurtosis was also acceptable and the value of multivariate kurtosis which was lower than 50.0 indicated multivariate normality distribution of the data set. Finally, the issues of uni-dimensionality, validity, and reliability were addressed in Table 5.19.

Table 5.18: Assessment of normality

Variable	min	max	skew	c.r.	kurtosis	c.r.
a19	1.000	5.000	-0.348	-3.955	-0.785	-4.466
a20	1.000	5.000	-0.676	-7.686	-0.239	-1.359
a21	1.000	5.000	-0.613	-6.969	-0.477	-2.710
a17	1.000	5.000	-0.638	-7.260	0.208	1.182
a18	1.000	5.000	-0.402	-4.574	-0.445	-2.532
Multivariate					9.059	15.081

Table 5.19: 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			

5.8.2 Process Dimension

Process dimension consists of twelve major constructs – belief, feeling, readiness, understanding, skills, IHT, administration, moderation, monitoring, challenges, role and importance of SBA. A total of fifty-two items are involved to measure all the twelve constructs. Process dimension, a single second-order factor was hypothesised as accounting for all variance and covariance related to the twelve first-order factors. However, when the measurement model was run, it was not fit. So, the PCA technique was conducted. Specific factors such as 2, 3 and 4 were forced on the data. Process was divided into four main groups – Group 1 (belief, feeling, readiness, understanding, moderation and importance of SBA), Group 2 (challenge), Group 3 (IHT, admin, monitoring) and Group 4 (role of SBA and skills). But when all the four measurement models went through SEM, most models were not fit. The researcher then modified the models based on theory. For Group 1, theoretically, belief, feeling and readiness relate to attitude so they were grouped together. Then, understanding and skills were added, as according to Heritage (2007a), those were the main components in implementing formative assessments

successfully. Group 2 (challenge) was maintained. For Group 3, IHT and admin were deleted due to low factor loadings but moderation was added to monitoring as they both are processes that contribute to QA, theoretically. Lastly, for Group 4, theoretically, role and importance of SBA are constructs for school improvement processes. So, the distribution of constructs to each 2nd-order construct was based on both – theory driven and data driven. Finally, four main measurement models were then produced:

i) Process1 – attitude (belief, feeling, readiness), understanding and skills.
ii) Process2 – moderation and monitoring process.
iii) Process3 – school improvement process (role and importance of SBA).
iv) Challenges – lack of confidence, problems getting resources, school climate, insufficient financial resources, extra workloads, no recognition, no integrity, students may memorise and insufficient numbers of teachers.

In the case of process dimension constructs, only the final model of each measurement model was presented to avoid having lengthy explanation.

i) The measurement model - Process1

Process1 dimension of evaluation consists of three 1st-order constructs – attitude (belief, feeling and readiness), understanding and skills. The final model was shown in Figure 5.6. The incorporation of the covariance between error terms had made a substantially large improvement to model fit. This evidence of misspecification associated with the pairing of error terms associated with two items by determining the MIs related to the co-variances. It was believed that to all the six cases of the addition of freely estimated parameters were due to a high degree of overlap in the item content. For example, item a1 asks whether 'Using SBA in teaching is efficient in improving students' learning' while item a2 asks whether 'Using SBA in teaching can help students to become more independent learners'. Both items related to improving students.

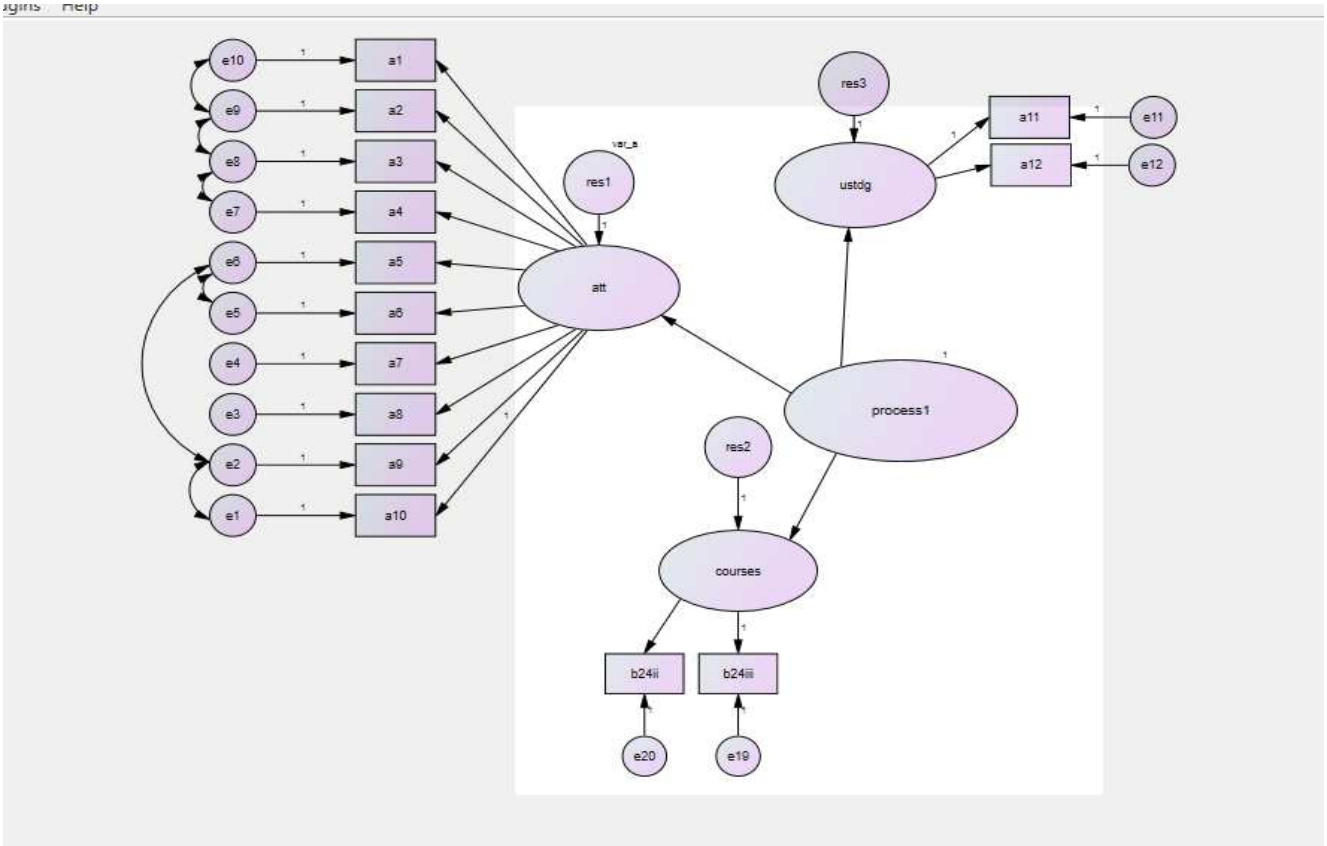


Figure 5.6: The final re-specified model of factorial structure for Process1

First was to look at the standardized regression weight shown in Table 5.20. All factor loadings were more than 0.5. Multivariate normality was 73.046 which was quite larger than 50.0 as shown in Table 5.21.

Table 5.20: Standardize Regression Weight

			Estimate
ustdg	<---	process1	0.962
att	<---	process1	0.998
courses	<---	process1	0.614
a10	<---	att	0.615
a9	<---	att	0.630
a8	<---	att	0.501
a7	<---	att	0.813
a6	<---	att	0.753
a5	<---	att	0.582
a4	<---	att	0.787
a3	<---	att	0.763
a2	<---	att	0.735
a1	<---	att	0.768
a11	<---	ustdg	0.870
a12	<---	ustdg	0.601
b24three	<---	courses	0.849

			Estimate
b24two	<---	courses	0.810

Table 5.21: Assessment of normality

Variable	min	max	skew	c.r.	kurtosis	c.r.
b24two	1.000	5.000	-1.353	-15.388	1.830	10.409
b24three	1.000	5.000	-1.055	-12.002	1.007	5.726
a12	1.000	5.000	-0.971	-11.040	1.295	7.365
a11	1.000	5.000	-0.662	-7.528	0.007	0.037
a1	1.000	5.000	-0.614	-6.986	-0.363	-2.065
a2	1.000	5.000	-0.548	-6.233	-0.399	-2.266
a3	1.000	5.000	-0.756	-8.597	0.180	1.023
a4	1.000	5.000	-0.839	-9.540	0.560	3.183
a5	1.000	5.000	0.247	2.806	-0.975	-5.543
a6	1.000	5.000	-0.585	-6.655	-0.272	-1.547
a7	1.000	5.000	-0.583	-6.627	-0.228	-1.296
a8	1.000	5.000	-0.911	-10.363	0.907	5.158
a9	1.000	5.000	-0.424	-4.821	-0.413	-2.350
a10	1.000	5.000	-0.589	-6.699	0.349	1.982
Multivariate					73.046	48.068

For this final re-specified model, X^2/df=5.322; GFI=0.936; AGFI=0.902; NFI=0.944; CFI=0.954; TLI=0.939 and RMSEA=0.075. The fit indices indicated that this three-factor second-order model schematically portrayed in Figure 5.6 was the most optimal model representing Process1 with the number of items for each construct were maintained. Finally, the issues of uni-dimensionality, validity and reliability were addressed in Table 5.22.

Table 5.22: The CFA results reporting for the measurement model

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 a2 a3 a4 a5 a6 a7 a8 a9	0.768 0.735 0.763 0.787 0.582 0.753 0.813 0.501 0.630	0.908	0.907	0.498

	a10	0.615			
Understanding	a11 a12	0.870 0.601	0.679	0.673	0.429
Skills	b24ii b24iii	0.810 0.849	0.813	0.902	0.698

ii) The measurement model – Process2

Process2 dimension of evaluation consists of two 1st-order constructs – moderation and monitoring process with both consists of three items. Figure 5.7 showed the final re-specified model.

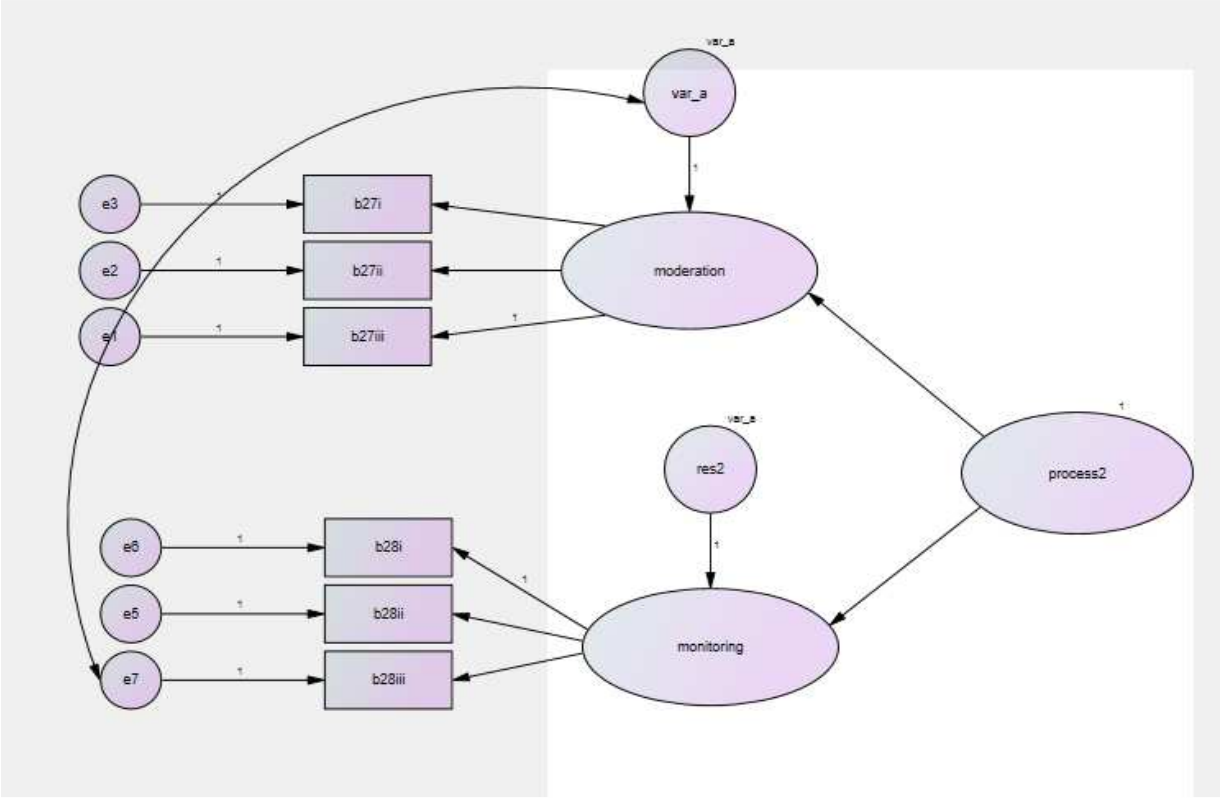


Figure 5.7: The final re-specified model of factorial structure for process2 dimension

All factor loadings were more than 0.5 as shown in Table 5.23. Multivariate kurtosis was 65.6 and there was one item (b24iii) with quite a high kurtosis value (2.942) as shown in Table 5.24.

Table 5.23: Standardized Regression Weights

			Estimate
moderation	<---	process2	0.623
monitoring	<---	process2	0.581
b27three	<---	moderation	0.891
b27two	<---	moderation	0.942

			Estimate
b27one	<---	moderation	0.827
b28two	<---	monitoring	0.734
b28one	<---	monitoring	0.815
b28three	<---	monitoring	0.601

Table 5.24: Assessment of normality

Variable	min	max	skew	c.r.	kurtosis	c.r.
b28three	1.000	5.000	-1.283	-14.594	2.942	16.728
b28one	1.000	5.000	-0.891	-10.138	0.785	4.466
b28two	1.000	5.000	-0.752	-8.556	0.158	0.897
b27one	1.000	5.000	-1.247	-14.176	1.915	10.892
b27two	1.000	5.000	-1.243	-14.141	2.528	14.372
b27three	1.000	5.000	-1.242	-14.124	2.480	14.100
Multivariate					65.600	93.255

The incorporation of the covariance between error terms had made a substantially large improvement to model fit. This evidence of misspecification associated with the pairing of error terms associated with two items by determining the MIs related to the co-variances. It was believed that the addition of freely estimated parameters was due to a high degree of overlap in the item content. Item b28iii which says 'evidence of centre assessment is kept in a systematic way in a safe place before it is returned to the students' is also suitable to load not only to the monitoring process but to the moderation process as well. The word 'evidence' might be interpreted by teachers as belong to both processes, moderation and monitoring. In reviewing the goodness-of-fit statistics, it showed that X^2/df=2.873; GFI=0.991; AGFI=0.974; NFI=0.992; CFI=0.995; TLI=0.989 and RMSEA=0.049.The fit indices indicated that this two-factor second-order model schematically portrayed in Figure 5.7 was the most optimal model representing process2 dimension of evaluation with the number of items for each construct was maintained. Finally, the issues of uni-dimensionality, validity and reliability were addressed in Table 5.25.

Table 5.25: The CFA results reporting for the measurement model

2nd order factor	1st order factor	Standardized Factor Loading			
Process2 dimension	Moderation	0.623			
	Monitoring	0.581			
1st order factor	Item	Standardized	CR Alpha	CR	AVE

		Factor loading (>0.5)	(>0.7)	(>0.6)	(>0.5)
Moderation	b27i b27ii b27iii	0.827 0.942 0.891	0.915	0.831	0.788
Monitoring	b28i b28ii b28iii	0.815 0.734 0.601	0.758	0.763	0.521

iii) Process3 – school improvement

Process3 dimension of evaluation consists of two 1st-order constructs – role of SBA and importance of SBA consisting of four and eight items, respectively. Theoretically, items under role of SBA ('role') and importance of SBA ('crucial') represented school improvement process involved in this study. Figure 5.8 showed the final re-specified model.

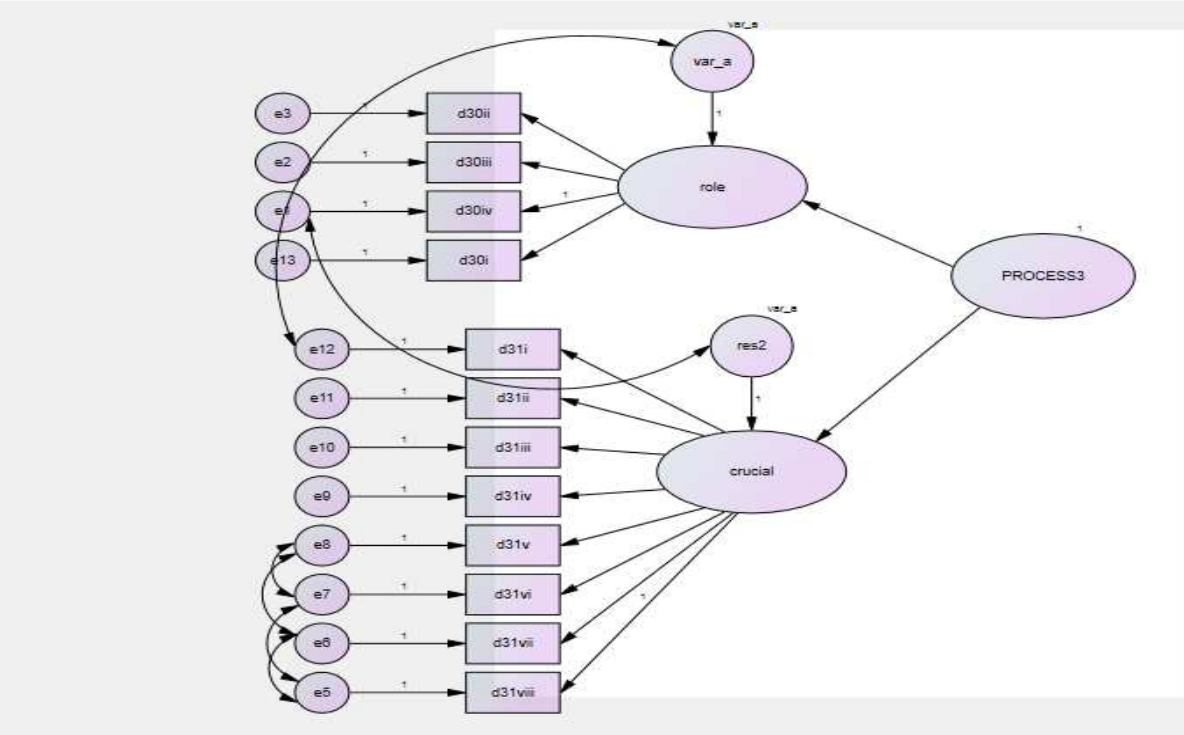


Figure 5.8: The final re-specified model of factorial structure for process3 dimension

The standardized regression weight was shown in Table 5.26. All factor loadings were more than 0.5. Assessment of normality was shown in Table 5.27.

Table 5.26: Standardized Regression Weights

			Estimate
role	<---	PROCESS3	0.733
Importance	<---	PROCESS3	0.774
d30four	<---	role	0.720
d30three	<---	role	0.914
d30two	<---	role	0.875
d31eight	<---	Importance	0.789
d31seven	<---	Importance	0.622
d31six	<---	Importance	0.829
d31five	<---	Importance	0.787
d31four	<---	Importance	0.901
d31three	<---	Importance	0.909
d31two	<---	Importance	0.883
d31one	<---	Importance	0.815
d30one	<---	role	0.584

Table 5.27: Assessment of normality

Variable	min	max	skew	c.r.	kurtosis	c.r.
d30one	1.000	5.000	0.079	0.903	-0.727	-4.134
d31one	1.000	5.000	-0.593	-6.745	-0.304	-1.726
d31two	1.000	5.000	-0.812	-9.239	0.016	0.091
d31three	1.000	5.000	-0.841	-9.567	0.239	1.360
d31four	1.000	5.000	-0.851	-9.679	0.346	1.965
d31five	1.000	5.000	-0.658	-7.486	-0.292	-1.661
d31six	1.000	5.000	-0.672	-7.647	-0.066	-0.376
d31seven	1.000	5.000	-0.546	-6.206	-0.235	-1.338
d31eight	1.000	5.000	-0.808	-9.188	0.151	0.857
d30two	1.000	5.000	-0.295	-3.357	-0.598	-3.398
d30three	1.000	5.000	-0.044	-0.498	-0.605	-3.441
d30four	1.000	5.000	-0.544	-6.182	-0.310	-1.761
Multivariate					94.815	72.046

Due to overlap of item content, error covariance parameters were included to improve the fitness of the model. In addition, modifying this model also involved an addition of the error covariance between items and constructs. The value of MIs was like telling us that this item should additionally load on to another construct due to the essence of their item content. Item d30iv which says 'SBA provides new learning opportunities for students' was suitable to load on 'importance of SBA', whereas item d31i which says 'SBA helps students improve generic skills' was suitable to load on 'role of SBA'. For this final model, it showed that X^2/df=5.979; GFI=0.944; AGFI=0.907; NFI=0.964; CFI=0.970; TLI=0.958 and RMSEA=0.080.The fit indices indicated that this two-factor second-order model schematically portrayed in Figure 5.8 was the most optimal model representing process3 with the number of items for each

construct was maintained. Finally, the issues of uni-dimensionality, validity and reliability were addressed in Table 5.28.

Table 5.28: The CFA results reporting for the measurement model

2nd order factor	1st order factor	Standardized Factor Loading			
Process3 dimension	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 d30ii d30iii d30iv	0.584 0.875 0.914 0.720	0.849	0.861	0.615
Importance of SBA	d31i d31ii d31iii d31iv d31v d31vi d31vii d31viii	0.815 0.883 0.909 0.901 0.787 0.829 0.622 0.789	0.945	0.943	0.675

iv) Challenges

Challenges are 1st-order construct which originally consists of 12 items. Figure 5.9 showed the final re-specified model. Three items were deleted due to low factor loadings.

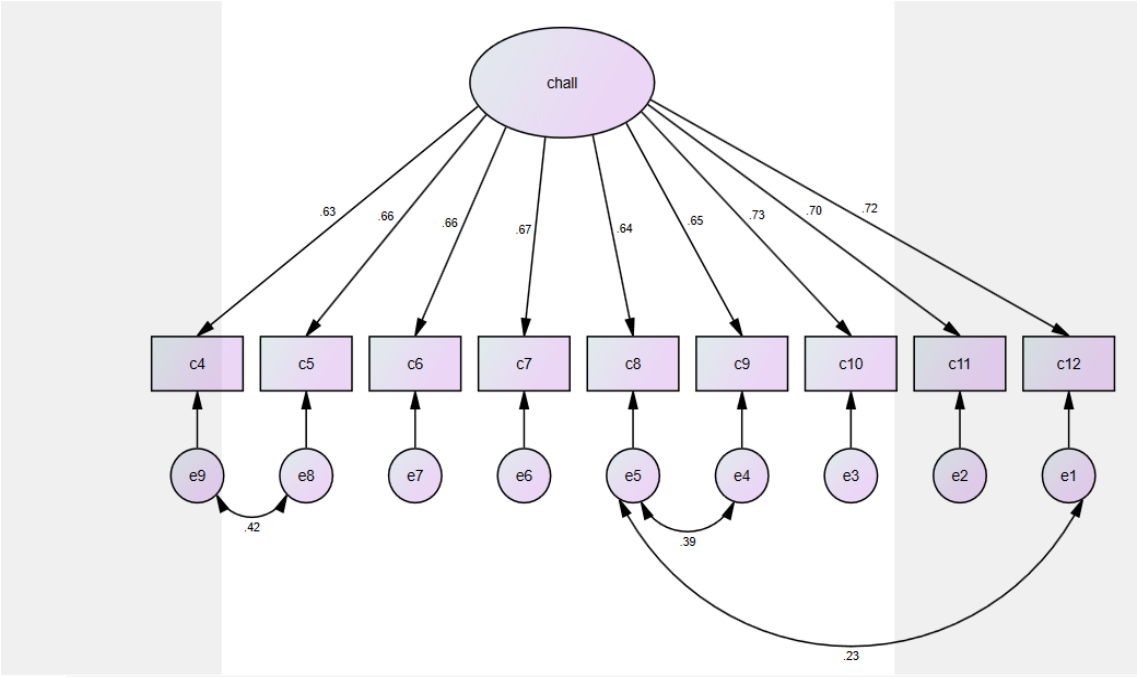


Figure 5.9: The final re-specified model of factorial structure for challenges

The standardised regression weight was shown in Table 5.29. All factor loadings were above 0.5, so no more item was deleted. Assessment normality was shown in Table 5.30.

Table 5.29: Standardised Regression Weights

			Estimate
c12	<--	Challenge	0.722
c11	<--	Challenge	0.703
c10	<--	Challenge	0.726
c9	<--	Challenge	0.655
c8	<--	Challenge	0.641
c7	<--	Challenge	0.668
c6	<--	Challenge	0.658
c5	<--	Challenge	0.662
c4	<-	Challenge	0.630

Table 5.30: Assessment of normality

Variable	min	max	skew	c.r.	kurtosis	c.r.
c4	1.000	5.000	-0.193	-2.191	-0.653	-3.715
c5	1.000	5.000	-0.494	-5.618	-0.603	-3.431
c6	1.000	5.000	0.105	1.195	-0.793	-4.508
c7	1.000	5.000	-0.347	-3.945	-0.814	-4.627
c8	1.000	5.000	-0.896	-10.193	0.087	0.493
c9	1.000	5.000	-0.610	-6.942	-0.135	-0.769
c10	1.000	5.000	-0.393	-4.469	-0.507	-2.881
c11	1.000	5.000	-0.432	-4.915	-0.546	-3.103

Variable	min	max	skew	c.r.	kurtosis	c.r.
c12	1.000	5.000	-0.364	-4.135	-0.738	-4.198
Multivariate					32.646	32.315

When MIs were examined, there were three cases where two items had overlap of item contents; between items c4 and c5, items c8 and c9 and items c8 and c12. So, this problem was solved by setting them to be 'free parameter estimate' by connecting them together. It was considered appropriate to re-specify the model with this parameter freely estimated due to the high MIs values and the obvious overlap of item content. For example, item c4 asks if the respondent feels lack of confidence to conduct SBA as a challenge and item c5 asks if the respondents feels problem to get related resources on SBA as a challenge. Teachers might feel lack of confidence when there were insufficient resources. Similarly with item c8 and item c12. Teachers might perceive 'extra workload' as similar to 'insufficient teachers' because their burden would increase when there was not enough teachers in school. So, it was clear that there appeared to be an overlap of content between these two items. Then, the model was re-specified. It showed that X^2/df=6.411; GFI=0.956; AGFI=0.918; NFI=0.951; CFI=0.959; TLI=0.938 and RMSEA=0.084. The fit indices indicated that this 1st-order model schematically portrayed in Figure 5.9 was the most optimal model representing challenges as one of the process dimension of evaluation. The issues of uni-dimensionality, validity and reliability were addressed in Table 5.31.

Table 5.31: The CFA results reporting for the measurement model

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			

5.8.3 Product Dimension

Product dimension includes three constructs – students' attitude towards SBA ('att'), students' knowledge in SBA ('know') and motivational source towards learning ('mot'). A total of eight items were involved to measure the three constructs. Product dimension, a single second-order factor was hypothesized as accounting for all variance and covariance related to the first-order factors as shown in Figure 5.10.

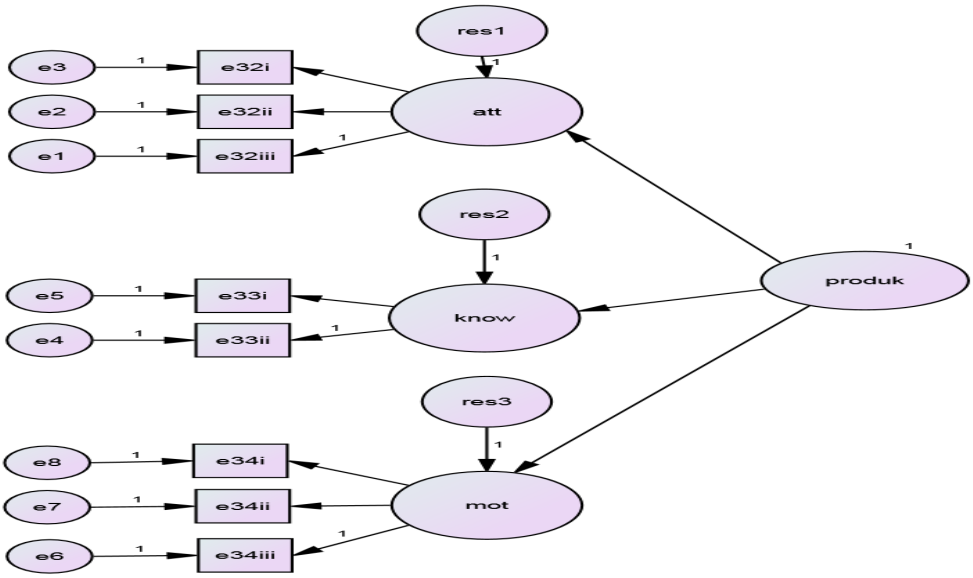


Figure 5.10: Hypothesised second-order model of factorial structure for the product dimension.

Referring to Table 5.32, it can be concluded that this model was identified with 17 degrees of freedom. The model was over-identified but with hierarchical models, the higher-order structure would be just-identified.

Table 5.32: AMOS Output for the hypothesised model for summary statistics

Computation of degrees of freedom	
Number of distinct sample moments	36
Number of distinct parameters to be estimated	19
Degrees of freedom	17
Results	
Minimum was achieved	
Chi-square	138.876
Degrees of freedom	17
Probability level	0.000

It showed that X^2/df=8.169; GFI=0.960; AGFI=0.915; NFI=0.971; CFI=0.974; TLI=0.958 and RMSEA=0.096. Then, to resolve just-identification issue in the second-order model, equality constraints were placed on particular parameters (appropriate residual terms) at the upper level to yield estimates that were approximately equal. The CRDIFF (critical ratio difference) method was conducted. Then, estimates were calculated. And, Table 5.33 showed the critical ratio difference between two parameters. The two prime candidates for the imposition of equality constraints were the higher order residuals related to the first and the third construct as their estimated value was non-significant (-2.911<1.96). And, the higher order level of the model would be over-identified with one degree of freedom.

Table 5.33: Critical ratios for differences between parameters

	Par_9	Par_10	Par_11
Par_9	0.000		
Par_10	4.946	0.000	
Par_11	-2.911	-6.060	0.000

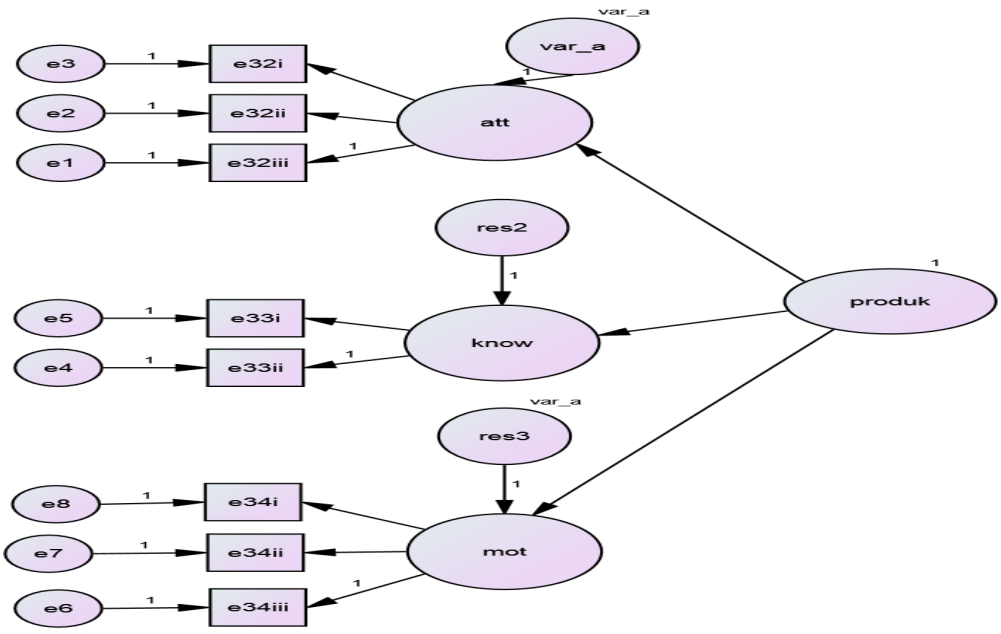


Figure 5.11: Hypothesised second-order model of factorial structure for the product dimension.

Analyses were based on this re-specified model shown in Figure 5.11. Two differences were acknowledged; i) there were 18 degrees of freedom rather

than 17 and ii) there were labeled parameters (the var_a assigned to res1 and res3). It showed that X^2/df=8.221; GFI=0.955; AGFI=0.910; NFI=0.969; CFI=0.973; TLI=0.957 and RMSEA =0.097. So, the goodness-of-fit statistics showed that the hypothesised model did not really fit the data very well. As a consequence, we examined the modification indices as shown in Table 5.34. Looking at the resulting MIs, it was observed that the error covariance related to item e34iii) ('SBA helps students to understand more on their strength and weakness') and the construct 'students' knowledge of SBA' remained a strongly mis-specified parameter in the model, with the estimated parameter change statistic suggesting that, if this parameter was incorporated into the model, it would result in an estimated value of approximately 0.069 which was quite small. The error covariance between this item, e34iii) and the construct 'students' knowledge of SBA' suggested redundancy due to the content, which made item e34iii) was suitable to load on the construct. Given the strength of the MIs and obvious overlap of item content, the error covariance parameter was suggested to be included in the model. The same reason went to the error covariance parameter between item e34iii) and another construct, which was 'students' attitude towards SBA'. Figure 5.12 showed the re-specified model.

Table 5.34: AMOS Output for Hypothesised Model: Modification Indices

Covariances

			M.I.	Par Change
e8	<-->	res1	10.272	0.034
e7	<-->	res2	25.120	-0.079
e7	<-->	res1	20.115	0.051
e6	<-->	res2	29.150	0.069
e6	<-->	res1	22.121	-0.043

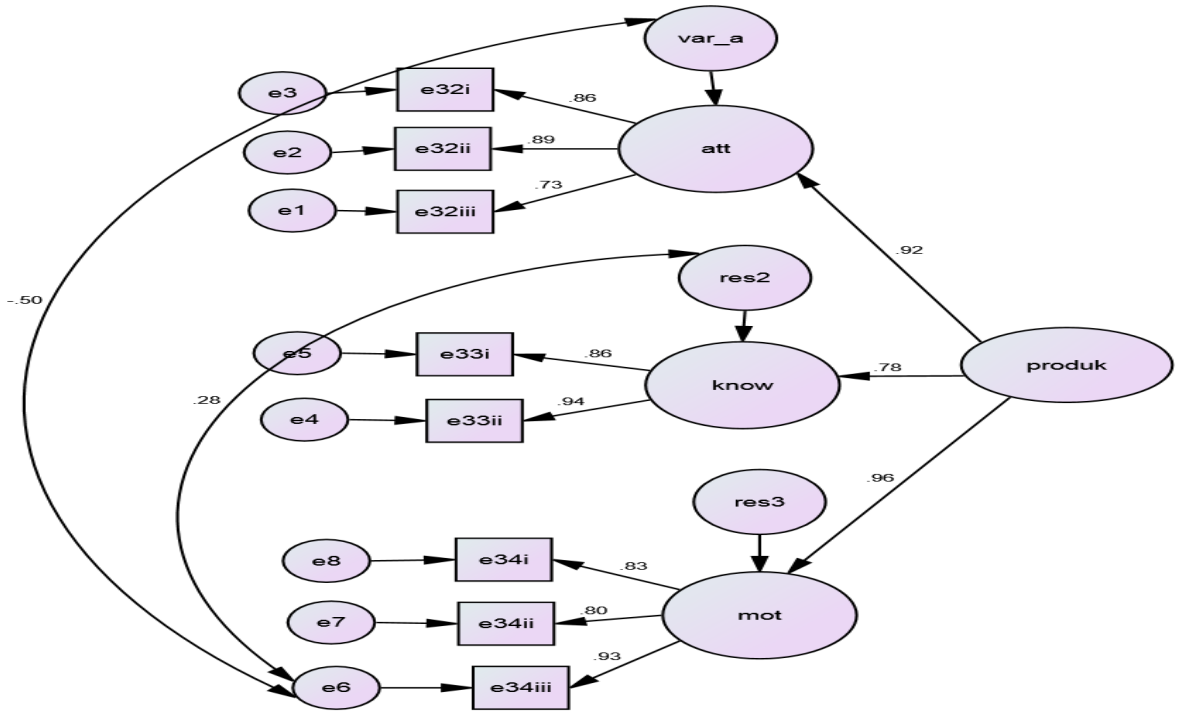


Figure 5.12: Re-specified model of factorial structure for product dimension

As shown in Table 5.35, all estimates were found to have critical ratio values more than 1.96, thereby indicating their statistical significance.

Table 5.35: AMOS Output for Hypothesised Model: Unstandardized Parameter Estimates

			Estimate	S.E.	C.R.	P	Label
att	<---	produk	0.684	0.033	20.614	***	par_7
know	<---	produk	0.760	0.033	22.709	***	par_8
mot	<---	produk	0.938	0.033	28.737	***	par_9
e32three	<---	att	1.000				
e32two	<---	att	1.234	0.051	24.229	***	par_2
e32one	<---	att	1.132	0.049	23.341	***	par_3
e33two	<---	know	1.000				
e33one	<---	know	0.906	0.028	31.803	***	par_4
e34three	<---	mot	1.000				
e34two	<---	mot	0.807	0.029	27.465	***	par_5
e34one	<---	mot	0.867	0.030	29.118	***	par_6

With the new model, all factor loadings exceeded the required value of 0.5 and it showed that X^2/df=5.821; GFI=0.972; AGFI=0.937; NFI=0.980; CFI=0.984; TLI=0.971 and RMSEA=0.079. This second-order model shown in Figure 5.12 was the most optimal model representing product dimension of evaluation. All eight items were maintained for the construct and their standardized factor

loadings were quite high. The value of kurtosis was also acceptable and the value of multivariate kurtosis was 37.499 which was lower than 50.0 indicated multivariate normality distribution of the data set was assumed. Finally, the issues of uni-dimensionality, validity and reliability were addressed in Table 5.36.

Table 5.36: The CFA results reporting for the measurement model

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			

Out of the 6 measurement models (input, Process1, Process2, Process3, challenge and product) formed in the study, only 3 models were satisfied with the multivariate normality distribution assumption suggested by Zainuddin (2012). So, in order to improve the multivariate normality of the measurement models, few extreme outliers were deleted. Out of 776 samples, 9 cases were deleted and it ended with 767 samples altogether. After cases were deleted, the assumption of multivariate normality had improved but still not satisfied, so the researcher deleted some of the items with low standardized estimate value. The decision was made not simply for the reason that the items possessed the lowest loading but also after careful analysis of practicality based on experience and literature relating to the overlap of its content with other items. Table 5.37 listed the characteristics of the final six measurement models.

i) The fitness indices value were the most well-fitting compared to previous models;

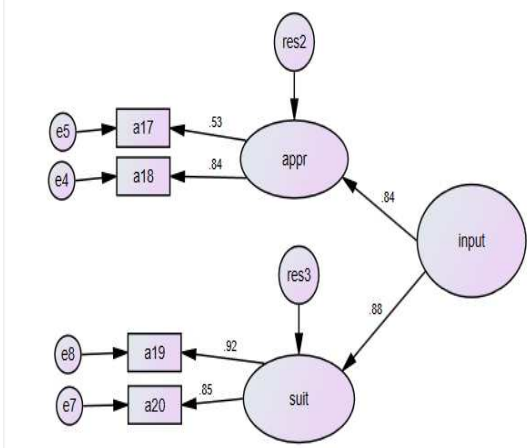
ii) All unstandardized estimates were statistically significant given critical values more than 1.96.
iii) All standard errors were in good order.
iv) All standardized estimates were above moderate strength.
v) The multivariate kurtosis value had improved and had achieved the required level.

All the six measurement models were shown in Figure 5.13. The researcher decided to stop at this point and continued with modelling all the constructs for further analysis. Finally, Table 5.38 listed out all the variables in the initial and final models.

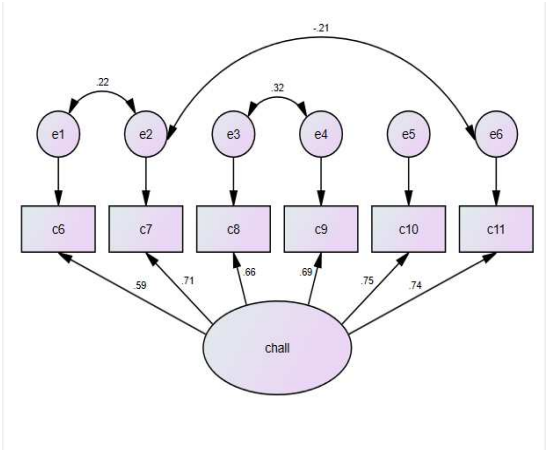
Table 5.37: Final characteristics of the measurement models

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	

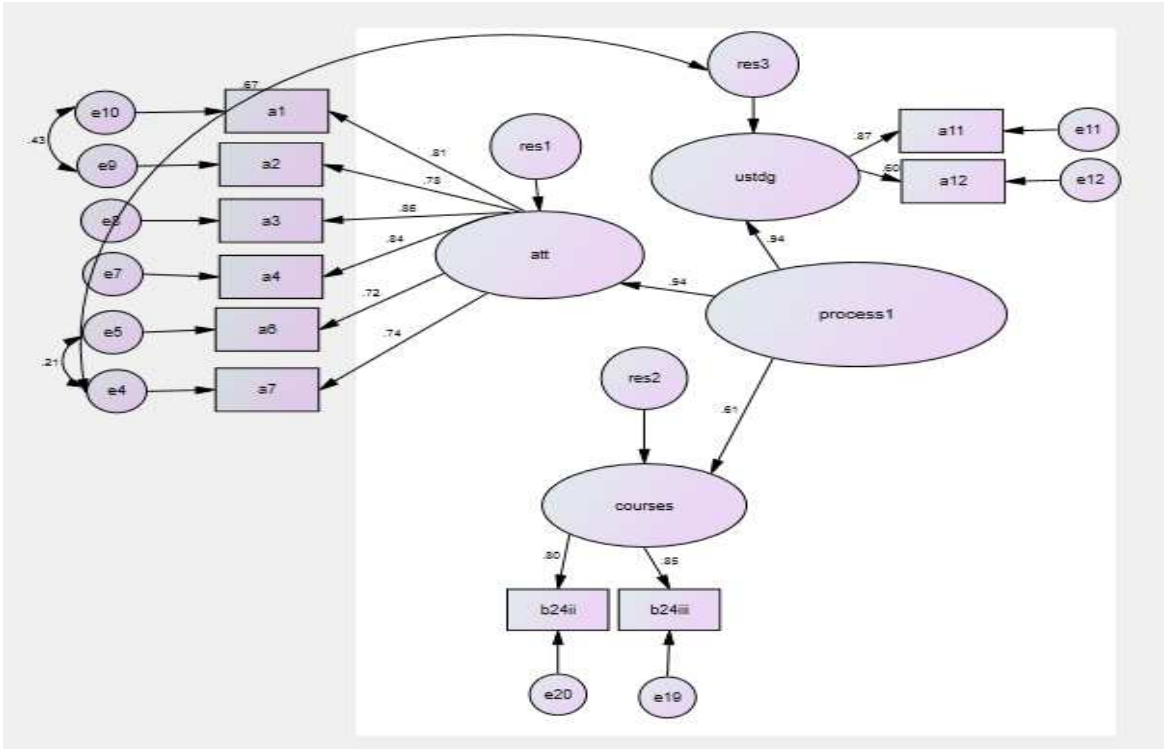
a) Input dimension



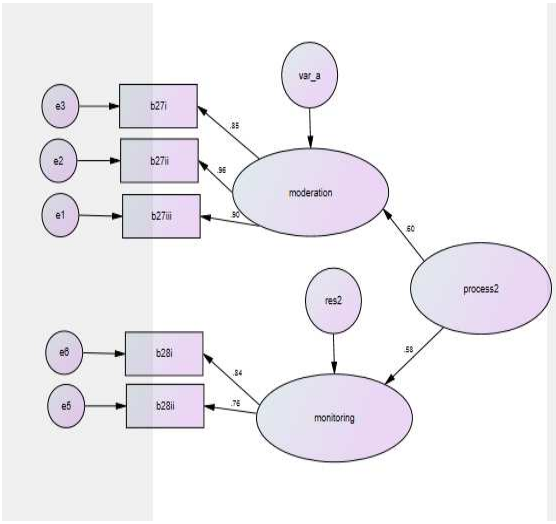
b) Challenges dimension



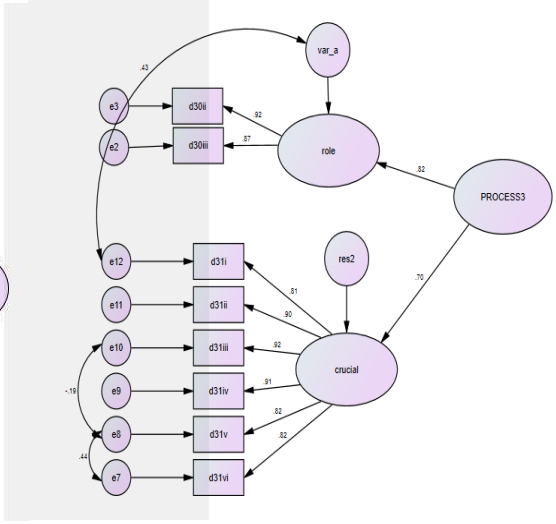
c) Process1 dimension



d) Process2 dimension



e) Process3 dimension



f) Product dimension

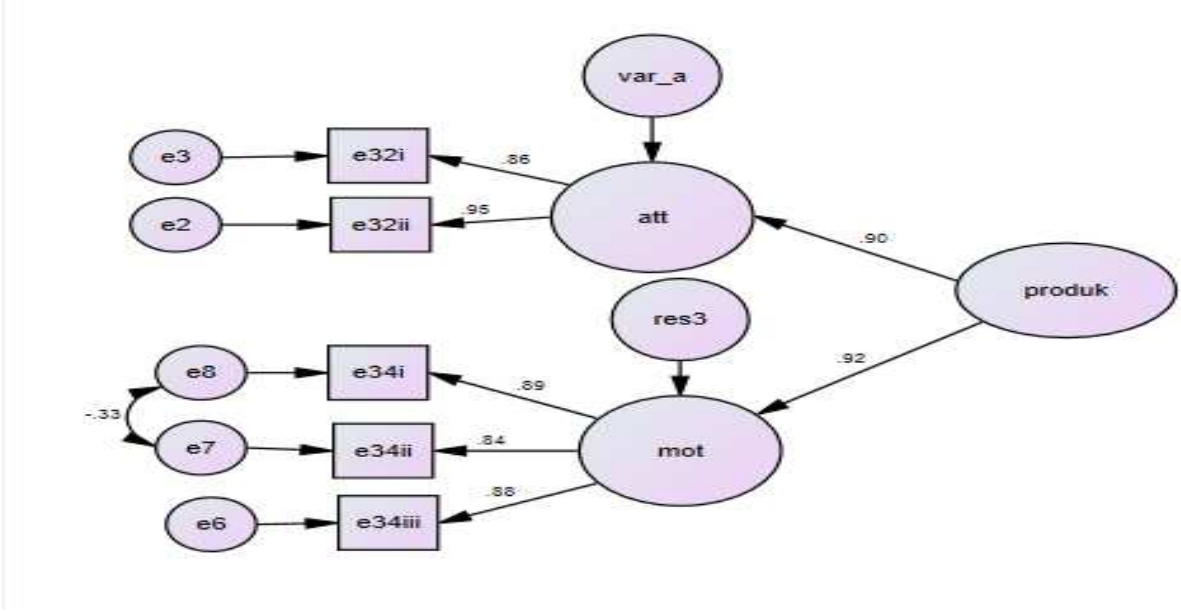


Figure 5.13: The measurement models – input, process and product dimension.

Table 5.38: Variables of the initial and final models

2nd order construct	1st order construct	Initial model	Final measurement model
INPUT	Material	a14, a15, a16	(none)
	Appropriateness	a17, a18	a17, a18

	Suitability	a19, a20, a21	a19, a20
Process1	Attitude	a1, a2, a3, a4, a5, a6, a7, a8, a9, a10	a1, a2, a3, a4, a6, a7
	Understanding	a11, a12, a13	a11, a12
	Courses (skills)	b24i, b24ii, b24iii, b24iv	b24ii, b24iii
Process2	Moderation	b27i, b27ii, b27iii	b27i, b27ii, b27iii
	Monitoring	b28i, b28ii, b28iii	b28i, b28ii
	IHT	b25i, b25ii, b25iii	(none)
	Admininstration	b26i, b26ii	(none)
Process3	Role of SBA	d30i, d30ii, d30iii, d30iv	d30ii, d30iii
	Importance of SBA	d31i, d31ii, d31iii, d31iv, d31v, d31vi, d31vii, d31viii	d31i, d31ii, d31iii, d31iv, d31v, d31vi
	Challenges	c1, c2, c3, c4, c5, c6, c7, c8, c9, c10, c11, c12	c6, c7, c8, c9, c10, c11
PRODUCT	Attitude	e32i, e32ii, e32iii	e32i, e32ii
	Knowledge	e33i, e33ii	(none)
	Motivation	e34i, e34ii, e34iii	e34i, e34ii, e34iii
TOTAL ITEM		68 items	38 items

Then, the final measurement model with all the constructs was formed. Table 5.39 showed the correlation between constructs. The correlation between Process3 and Process1 was equal to 0.923 and the correlation between Process3 and Product was equal to 0.939. Process3 was then deleted from the model as Kline (2005) suggested that the correlation between constructs should be less than 0.85 to avoid multicollinearity problem.

Table 5.39: Correlation between constructs

Factor correlations			Estimate
INPUT	<-->	Process1	0.765
INPUT	<-->	CHALL	-0.398
INPUT	<-->	Process2	0.601
INPUT	<-->	Process3	0.782
INPUT	<-->	PRODUCT	0.616
CHALL	<-->	Process1	-0.368
CHALL	<-->	Process2	-0.281
CHALL	<-->	Process3	-0.474
CHALL	<-->	PRODUCT	-0.532
Process2	<-->	Process3	0.807
Process2	<-->	PRODUCT	0.670
Process1	<-->	Process2	0.856
PRODUCT	<-->	Process3	0.939
Process1	<-->	Process3	0.923
Process1	<-->	PRODUCT	0.696

A hypothesized confirmatory measurement model was constructed in order to answer the research questions of the study. The answers for Research Question 4, 5 and 6 were presented.

RQ4. Is the measurement scale for input dimension construct valid?
This validated measurement model can be explained by two factors.

RQ4.1: Can the evaluation in input dimension be explained by the following three factors: 'material', 'appropriateness' and 'suitability'?
No. Only two factors – 'appropriateness' and 'suitability'. One factor – 'material' is deleted from the model. Only two factors loaded significantly on the input dimension construct. This means that a measurement model for input dimension of evaluation could be explained by two factors, namely 'appropriateness' and 'suitability'.

RQ4.2: Does each indicator have a non-zero loading on the 1st-order factor it was designed to measure?
Yes.

RQ4.3: Does each indicator have a zero loadings on the other 1st-order factor?
Yes.

RQ4.4: Are the error terms uncorrelated?
Yes.

RQ5. Is the measurement scale for process dimension construct valid?
There were four measurement models; Process1 is validated with three factors, Process2 is validated with two factors, Process3 is validated with two factors and Challenge is validated with six indicators.

RQ5.1: Can the evaluation in process dimension be explained by the following 12 factors?
Out of four measurement models, Process3 is deleted. Process1 can be explained by three factors – attitude, understanding and courses; Process2 can be explained by two factors – moderation and monitoring and; challenges can be explained by six indicators.

RQ5.2: Does each indicator have a non-zero loading on the hypothesized or targeted factor?
Yes.

RQ5.3: Does each indicator have a zero loading on the other non-targeted factor?

Yes.

RQ5.4: Are the error terms uncorrelated?

Six set of measurement errors are correlated.

RQ6. Is the measurement scale for product dimension construct valid?

The validated measurement model can be explained by two factors with one set of correlated error terms.

RQ6.1: Can the evaluation in input dimension be explained by the following three factors: attitude, knowledge and motivation?

No, only two factors load significantly on the product dimension construct. This means that a measurement model for product dimension of evaluation can be explained by the two factors – attitude and motivation.

RQ6.2: Does each indicator have a non-zero loading on the 1st-order factor it was designed to measure?

Yes.

RQ6.3: Does each indicator have a zero loadings on the other 1st-order factors?

Yes.

RQ6.4: Are the error terms uncorrelated?

All the error terms associated with each item are uncorrelated except for one measurement error (*e7-e8*).

5.9 Analyses and Results of Structural Model: Stage 2

Following the first stage of assessing the measurement model was the second stage, analysing the SEM. The structural models were evaluated by comparing several models to ensure that the model fits the data. Then, the hypotheses testing were carried out. The hypotheses testing were separated into two main parts. The first part was to test the interrelationships of the main constructs and the second part was to test the moderating effect of school type and school category.

In this section, another nine hypotheses were tested to answer the research questions developed in Chapter 1. Since process dimension was split into four

constructs, the hypothesis related to process dimension was also split into four hypotheses. So, there were 15 hypotheses altogether as shown in Table 5.40.

Table 5.40: List of Hypotheses

Hypotheses	
H_{13}	Process dimension is associated with product dimension
	a) Process1 is positively associated with product b) Process2 is positively associated with product c) Process3 is positively associated with product d) Challenges is negatively associated with product
H_{14}	Input dimension is associated with process dimension
	a) Input is positively associated with process1 b) Input is positively associated with process2 c) Input is positively associated with process3 d) Input is negatively associated with challenges
H_{15}	Input dimension is positively associated with product dimension
H_{16}	School type moderates the relationship between input and product
H_{17}	School type moderates the relationship between process and product
H_{18}	School type moderates the relationship between input and process
H_{19}	School category moderates the relationship between input and product
H_{20}	School category moderates the relationship between process and product
H_{21}	School category moderates the relationship between input and process

In order to determine the best-fit structural model, another fit index under predictive fit indices, the Expected Cross-Validation Index (ECVI) was included. This fit index is seen suitable in comparing several models in order to look for the best-fit model (Byrne, 2010). Previously, we had seen that few indices had been used, such as X^2/df, GFI, AGFI, NFI, CFI, TLI and RMSEA value with certain benchmark for each index. ECVI is a goodness-of-fit index which measures "the discrepancy between the fitted covariance matrix in the analysed sample, and the expected covariance matrix that would be obtained in another sample of equivalent size" (Byrne, 2010, p. 82). In applying this index, its values were compared between several models, and the one with the smallest ECVI value was considered the best-fit model as it was the one most likely to replicate (Kline, 2005). However, ECVI value has no benchmark as it could take any value.

5.9.1 The Hypothesized Model - The Structural Model 1

In general, structural model was assembled for further analysis from the measurement models based on the theoretical interrelationships among the constructs or grounded in empirical research (Byrne, 2010). As was determined

in previous study by Stufflebeam (1971a), it was therefore hypothesised that input dimension of SBA was positively related to product dimension. Input was also related to process which consisted of four main latent constructs – challenges, Process1, Process2 and Process3. Input was hypothesized to be negatively related to challenges faced by teachers, while input was also hypothesized to be positively related to Process1, Process2 and Process3. Similarly, challenges were negatively related to product dimension and all the other three process dimensions - Process1, Process2 and Process3 were positively related to product dimension. School type and school category were added as moderating variables in the model.

The hypothesised structural Model 1 shown in Figure 5.14 was tested assuming the above relationship. There were 9 hypothesized causal paths altogether. Initially, when the survey was distributed to the respondents, there were 68 observed variables altogether, which came from the input, process and product dimensions. After going through procedures to gain the best measurement models, the number of observed variables was reduced. The first structural model was formed with 38 observed variables as shown in Figure 5.15. The summary notes of the model were reviewed in Table 5.41.

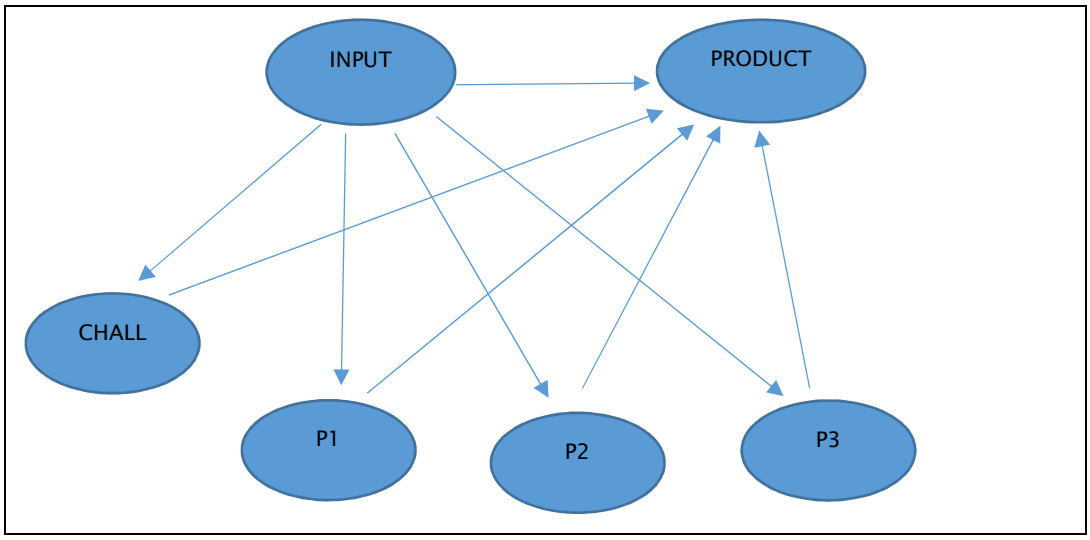


Figure 5.14: The hypothesised structural Model 1

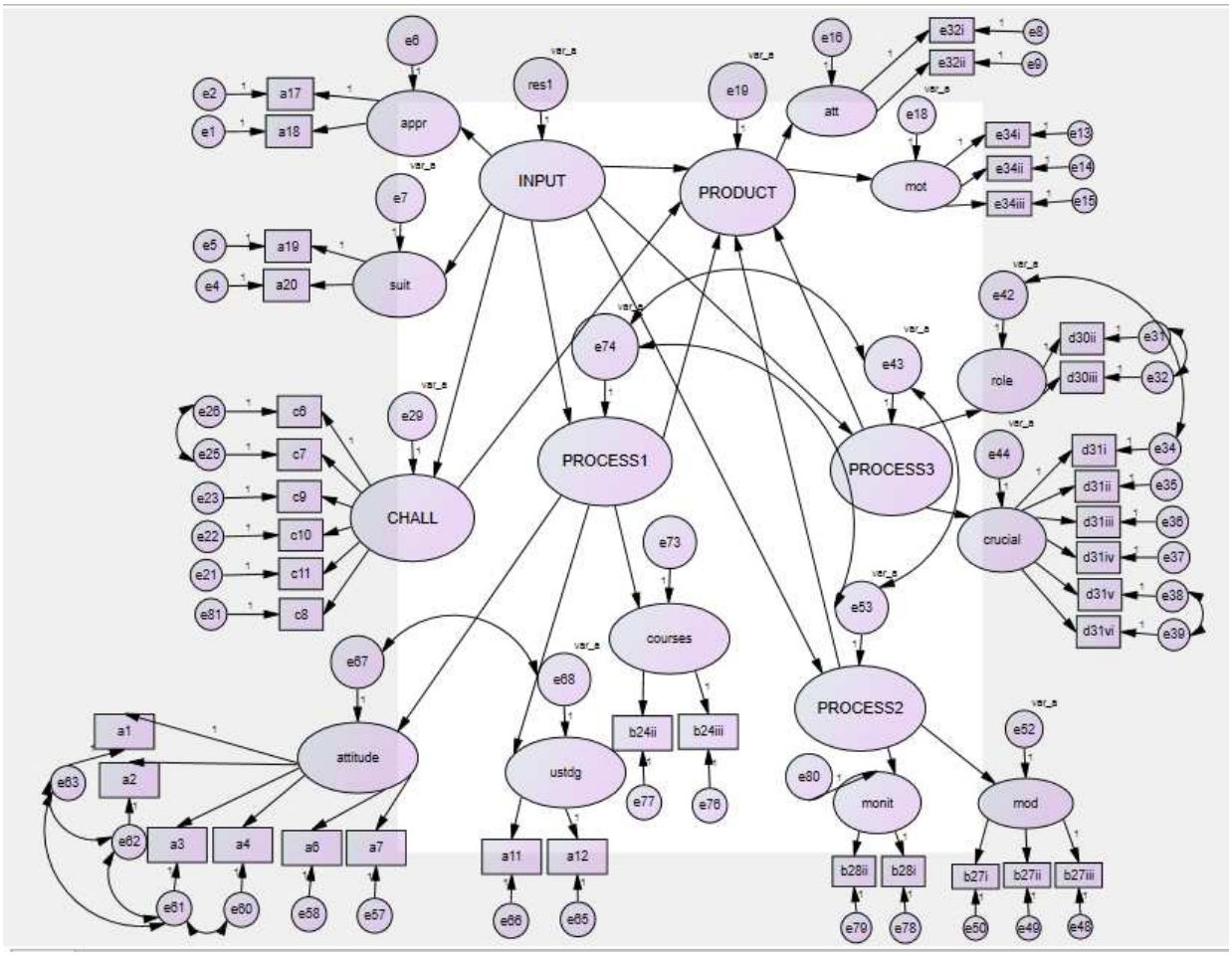


Figure 5.15: The hypothesised model – structural Model 1

Table 5.41: AMOS Output for the hypothesized Model 1

Computation of degrees of freedom	
Number of distinct sample moments	741
Number of distinct parameters to be estimated	102
Degrees of freedom	639
Result	
Chi-square	2692.337
Degrees of freedom	639
Probability level	0.000

First was to evaluate model fit to determine how well the structural Model 1 explained the data. It showed that the overall X^2 value was 2692.337 and the degrees of freedom was 639. The indices of fit showed X^2/df=4.213; GFI=0.836; AGFI=0.810; NFI=0.884; CFI=0.909; TLI=0.899 and RMSEA=0.065. Since this model was compared to other few models, ECVI value was also included (ECVI= 3.781). The examination of goodness-of-fit indicated that the hypothesized model did not fit the data very well. Process3 was deleted from

the model since the value of intercorrelation between Process3 and Process1 (r=+0.923) and between Process3 and Product (r=+0.944) were very high, thus rejected discriminant validity. Then, the structural parameter estimates were reviewed. After Process3 was deleted, there were 3 paths which were non-significant (PRODUCT <- P1; PRODUCT <- INPUT and PRODUCT <- P2). One path was deleted one at a time as suggested by Byrne (2010) to check for the best solution. Two paths were deleted; a path from Process2 to Product (PRODUCT <- PROCESS2; CR=1.212) and from Input to Product (PRODUCT <- INPUT; CR=1.091). Then, a new model was produced, which was the structural Model 2.

5.9.2 The Hypothesized Model - The Structural Model 2

Estimation of the structural Model 2 yielded an overall X^2 value of 1652.327 and degree of freedom of 388. Model fit statistics were as follows: X^2/df=4.259; GFI=0.870; AGFI=0.845; NFI=0.895; CFI=0.917; TLI=0.907; RMSEA=0.065 and ECVI=2.358. The deletion of construct can change the number of sample moments and degree of freedom, and then cause a big difference in X^2 value (Byrne, 2010). So, the deletion of construct and two non-significant paths had improved the fitness of the model. ECVI had also improved from 3.781 in Model 1 to 2.358 in Model 2. All the five causal paths left were statistically significant.

5.9.3 The Hypothesized Model - The Structural Model 3

To further improve the fitness of the model, the MIs were reviewed. Because the researcher was interested solely in the causal paths of the model so only MIs associated with the regression path were detected. Only one value of MI for the regression weights and it was less than 10.00, showing little to be gained if specifying cross-loading (Byrne, 2010), as shown in Table 5.42. Then, one construct, 'monitoring process' and one item, item c6 ('school climate seems to hinder SBA implementation') from the 'challenges' construct were deleted due to their low estimates parameter value and the overlap of item content. The overlap of item c6 and item c7 ('insufficient financial resources') was when teachers felt that factors such as 'school climate' and 'insufficient financial resources' were similar in challenging SBA implementation.

Table 5.42: AMOS Output for Model 3: Modification Indices

Regression weights	MI	Par Change
Process1 ← Challenge	5.371	-0.109

5.9.4 The Hypothesized Model - The Structural Model 4

Next, item a12 ('SBA is conducted during the teaching and learning process') was deleted due to low parameter estimates. Looking at the data for item a12 descriptively, it was skewed negatively. Then, it followed by the deletion of 'understanding' construct as there was only one item left, which was item a11 ('educational assessment transformation through SBA leads to changes from examination-oriented assessment to an assessment which is more integrated'). Looking at item a11, theoretically the content suited 'attitude' construct, so it was grouped together under 'attitude' construct. Since there is no firm rules and regulations of when to stop fitting a model and to incorporate a sufficient number of parameters in getting the best model representing the data (Byrne, 2010), and in an attempt to get the best parsimonious model, the researcher stopped at this point. The structural Model 4 looked fine. The hypothesised structural Model 4 was produced as shown in Figure 5.16. AMOS output was shown in Table 5.43.

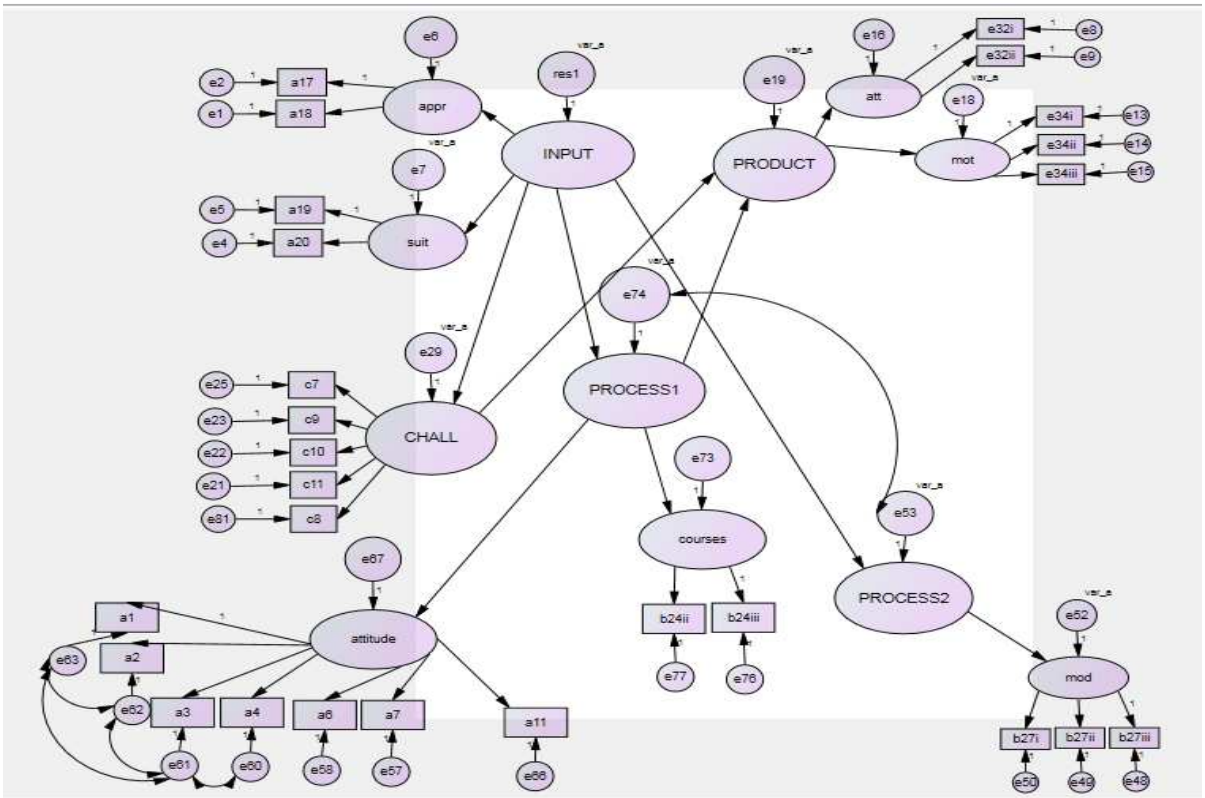


Figure 5.16: The hypothesised model – structural Model 4

Table 5.43: AMOS Output for the hypothesised Model 4

Computation of degrees of freedom	
Number of distinct sample moments Number of distinct parameters to be estimated Degrees of freedom	351 67 284
Result	
Chi-square Degrees of freedom Probability level	1066.796 284 0.000
Variables	
Number of variables in your model	76
Number of observed variables	26
Number of unobserved variables	50
Number of exogenous variables	38
Number of endogenous variables	38

The ML estimation of this structural Model 4 (final model) yielded several noteworthy result:

1) The overall X^2 value was 1066.796 with 284 degrees of freedom. The chi-square test of overall model fit resulted with a statistically significant discrepancy between the model and the data (X^2=1066.796, p=0.000; CMIN/df=3.756)
2) The other overall fit statistics were as follows: GFI=0.901; AGFI=0.877; NFI=0.924; CFI=0.943; TLI=0.935 and RMSEA=0.060 and the ECVI was 1.568.
3) All of the unstandardized estimates for the five structural parameter paths were statistically significant as indicated by the critical values and their related p-values as shown in Table 5.44. All values of standardized estimates were sound and all were less than 1.00. The value of standardized estimate should be above 0.30, or at least 0.20 for it to be meaningful (Chin, 1998).

Table 5.44: AMOS Output for Model 4 for unstandardized and standardized estimates

			Unstandardized Estimate	S.E.	C.R.	P	Label
PROCESS1	<--	INPUT	1.555	0.149	10.424	***	par_26
CHALL	<--	INPUT	-0.469	0.050	-9.457	***	par_28
PRODUCT	<--	CHALL	-0.406	0.054	-7.535	***	par_29
PROCESS2	<--	INPUT	0.532	0.062	8.535	***	par_33
PRODUCT	<--	PROCESS1	0.515	0.051	10.059	***	par_36

			Standardized Estimate
PROCESS1	<---	INPUT	0.841
CHALL	<---	INPUT	-0.425
PRODUCT	<---	CHALL	-0.289
PROCESS2	<---	INPUT	0.470
PRODUCT	<---	PROCESS1	0.613

4) All error covariances and error variances were statistically significant.
5) The value of the squared multiple correlations were shown in Table 5.45 which represented the proportion of variance that was explained by the predictors of the construct (Byrne, 2010). Only values for the main constructs were shown. For example, 70.7 percent of the variance associated with process1 was accounted for by its predictor – input dimension.
6) This structural Model 4 (final model) consisted of 26 items whereas initially the structural Model 1 consisted of 38 items instead.

Table 5.45: Squared Multiple Correlations

CONSTRUCT	Estimate
INPUT	0.000
PROCESS1	0.707
CHALL	0.180
PROCESS2	0.221
PRODUCT	0.585

Model statistics shown in Table 5.46 compared all the four models involved. It clearly showed how the fit indices values had improved. ECVI also dropped, thereby indicating that Model 4 represented the best fit model to the data. This value of fitness index of the final proposed model was considered acceptable since improving the structural model fit by deleting constructs and non significant paths had to be balanced out with respect to substantive theory,

and it had to have a strong theoretical justification as suggested by Hooper *et al.* (2008). Modification and re-specification done to the structural models to improve fit have to be balanced with the purpose of conducting SEM itself, which was to test theory.

Table 5.46: Values of fit statistics of all the four hypothesised models

	Model 1	Model 2	Model 3	Model 4
X^2/df	4.213	4.259	3.979	3.756
CFI	0.909	0.911	0.935	0.943
RMSEA	0.065	0.065	0.062	0.060
ECVI	3.781	2.358	1.783	1.568
GFI	0.836	0.870	0.891	0.901
AGFI	0.810	0.845	0.867	0.877
Number of items	38	30	27	26
Multivariate kurtosis	410.989	410.989	289.472	174.396

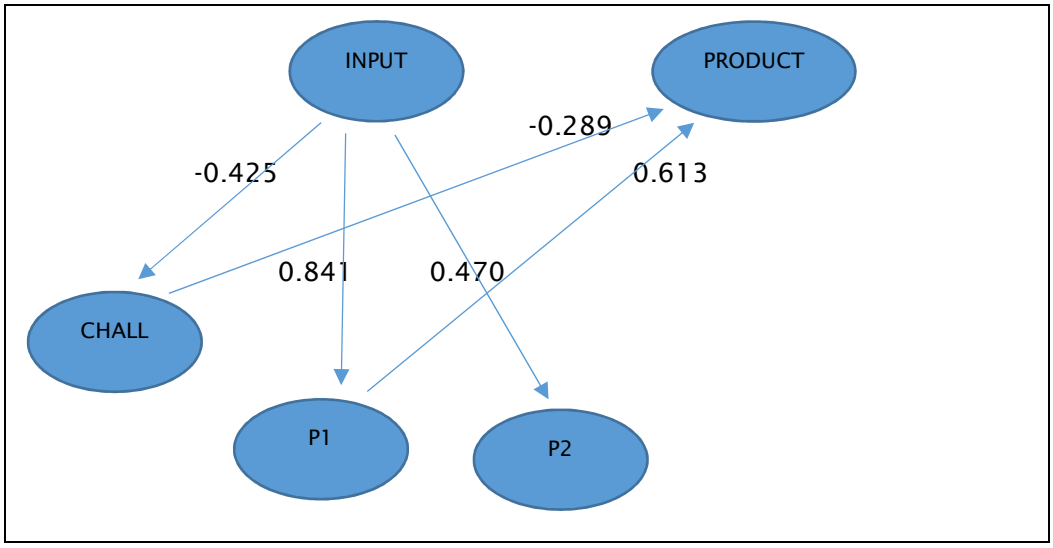


Figure 5.17: The final causal path with the standardised estimates

In summarizing the final model (as shown in Figure 5.17), i) out of nine causal paths specified in the revised hypothesized model, five paths were found to be statistically significant. These path reflected the impact of input on Process1, Process2 and challenges, challenges and Process1 on product; ii) four hypothesized paths (Input→Product; Process2→Product; Process3→Product and Input→Process3) were deleted from the model.

An issue concerning the final model was that, while CFI and RMSEA value showed a good fit, this was not the case for GFI and AGFI value. According to Hooper *et al.* (2008), CFI and RMSEA values should be reported as fit indices for they were the most insensitive to sample size, model misspecification and parameter estimates. Even, CFI has relative insensitivity to model complexity. However, GFI has limitation. Its value tends to have a downward value bias if the model has a large number of degree of freedom compared to the number of sample size. The final model in this study had a degree of freedom equals 284 and the sample size was 767, so that might be the possible reason why GFI value was low. And AGFI value was also reduced with more complex models.

5.10 Moderating Effect of Observed Variables

A moderating variable is a variable that moderates the relationship between an independent variable, X and dependent variable, Y (Zainuddin, 2012). In other words, it moderates the strength of causal effects from X to Y; either to enhance or reduce the effects of X on Y as illustrated in Figure 5.18. In this study, school type (urban-rural) and school category (secondary-primary) were hypothesised as moderator variables.

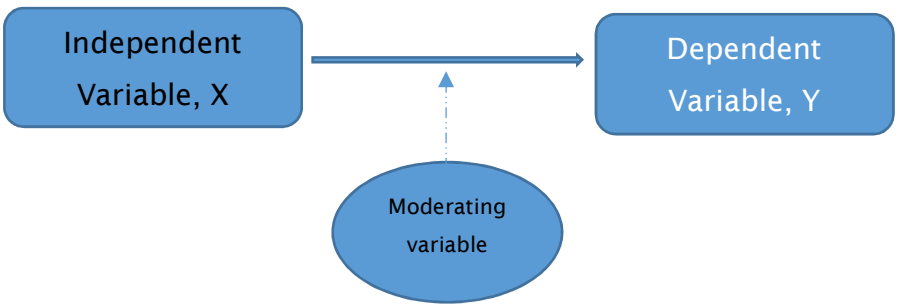


Figure 5.18: Illustration of the moderator variable

AMOS is used to test the moderating effect of the moderator variables on the relationship between variables. Or, in particular, in analysing the moderator for latent constructs, the multi-group CFA method was used to avoid model convergence problem or standard errors distortion problem (Zainuddin, 2012).

5.10.1 Moderating Effect of School Type

There were 380 urban school teachers and 396 rural school teachers from the sample. From the full structural model that had been finalized, the moderation effect of urban and rural was examined to all the significant paths. Figure 5.19 showed the output for urban group and Figure 5.20 for rural group.

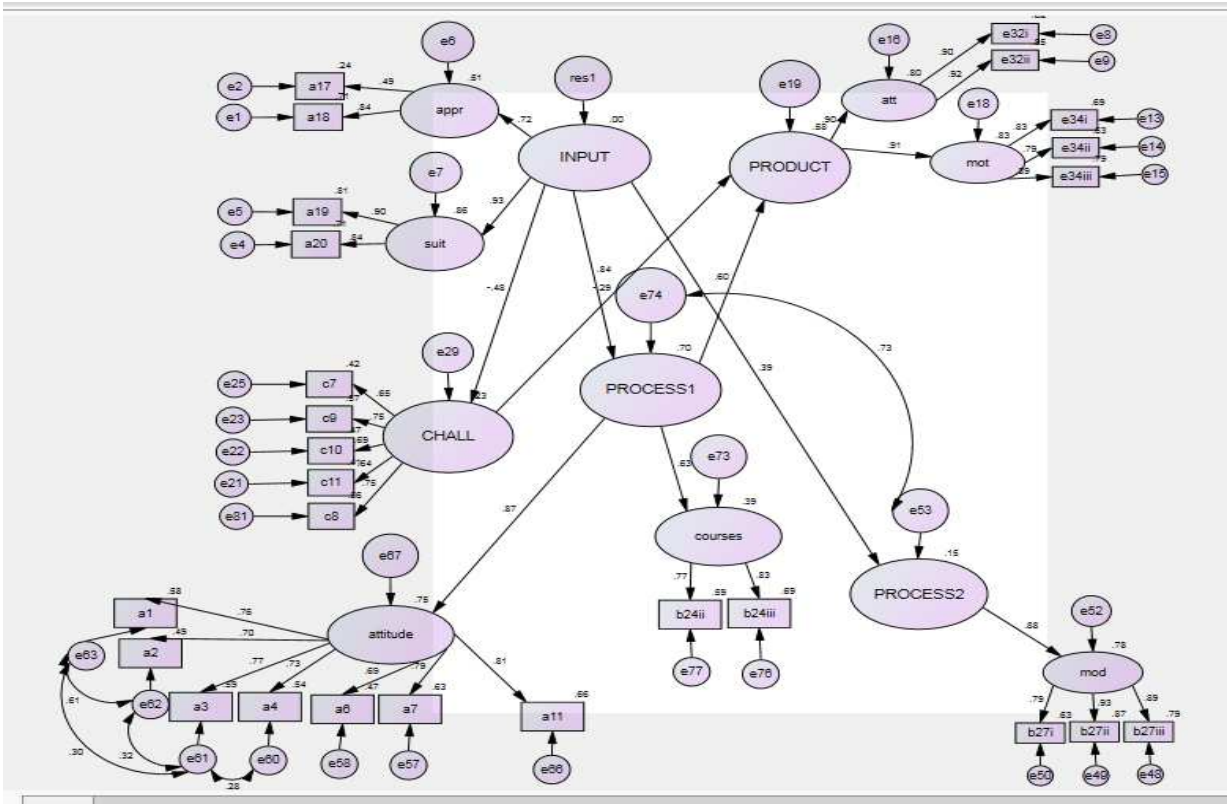


Figure 5.19: The hypothesised moderating model for Urban school

The ML estimation of this urban school model yielded several noteworthy result:

1) The overall X^2 value, with 284 degrees of freedom, was 775.314. The chi-square test of overall model fit resulted with a statistically significant discrepancy between the model and the data (X^2=775.314, p=0.000; CMIN/df=2.730).
2) The other overall fit statistics were as follows: GFI=0.860; AGFI=0.828; NFI=0.880; CFI=0.920; TLI=0.909 and RMSEA=0.068 and the ECVI was 2.425.
3) All of the unstandardised estimates for the five structural parameter paths were statistically significant as indicated by the critical values and

their related p-values as shown in Table 5.47. All values of standardized estimates were sound and all were less than 1.00.

Table 5.47: AMOS Output for urban model for unstandardised and standardised estimates

			Unstandardised Estimate	S.E.	C.R.	P	Label
PROCESS1	<---	INPUT	1.529	0.230	6.645	***	par_26
CHALL	<---	INPUT	-0.553	0.076	-7.271	***	par_28
PRODUCT	<---	CHALL	-0.385	0.078	-4.923	***	par_29
PROCESS2	<---	INPUT	0.427	0.081	5.244	***	par_33
PRODUCT	<---	PROCESS1	0.502	0.077	6.538	***	par_36

CONSTRUCT			Standardised Estimate
PROCESS1	<---	INPUT	0.837
CHALL	<---	INPUT	-0.484
PRODUCT	<---	CHALL	-0.286
PROCESS2	<---	INPUT	0.392
PRODUCT	<---	PROCESS1	0.597

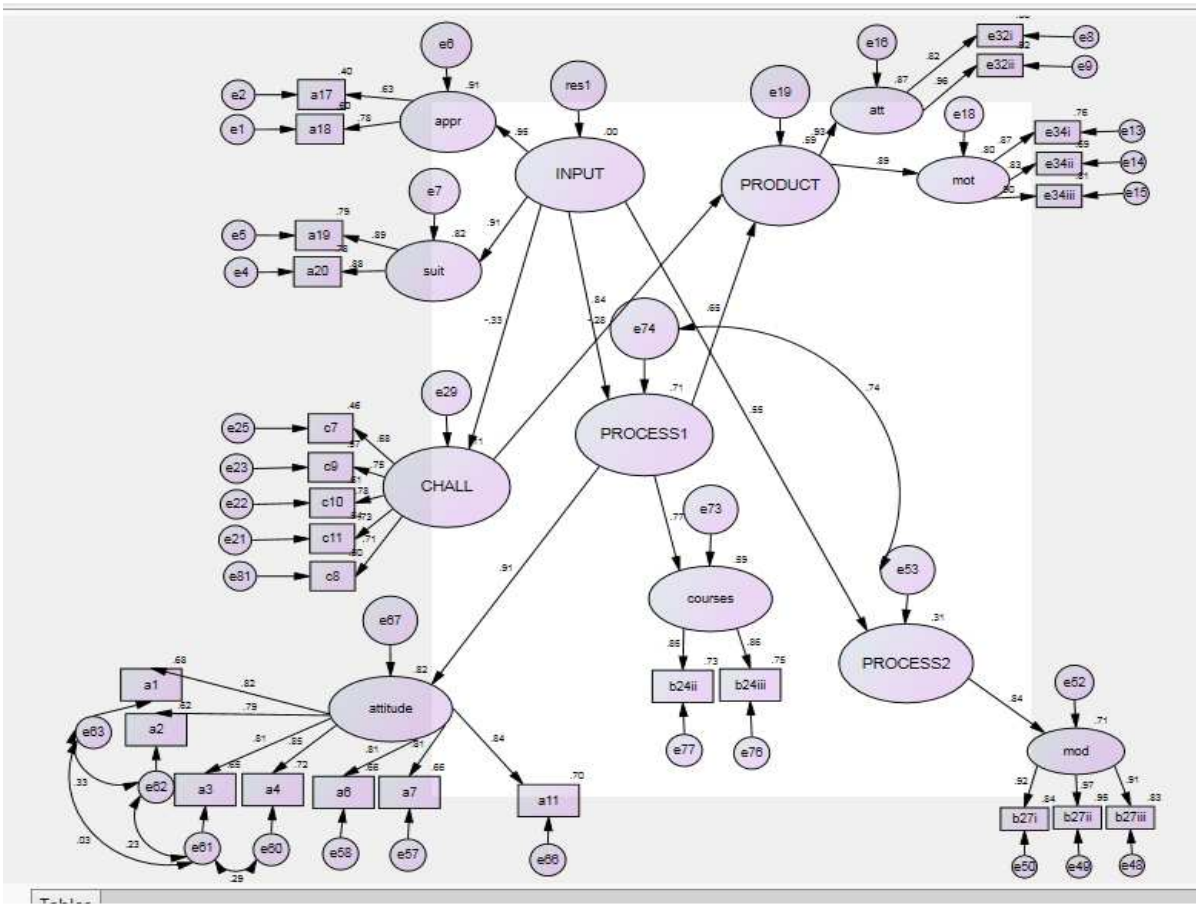


Figure 5.20: The hypothesised moderating model for Rural school

The ML estimation of this rural school model yielded several noteworthy result:

1) The overall X^2 value, with 284 degrees of freedom, was 919.058. The chi-square test of overall model fit resulted with a statistically significant discrepancy between the model and the data (X^2=919.058, p=0.000; CMIN/df=3.236).
2) The other overall fit statistics were as follows: GFI=0.843; AGFI=0.806; NFI=0.891; CFI=0.921; TLI=0.910 and RMSEA=0.076 and the ECVI was 2.700.
3) All of the unstandardised estimates for the five structural parameter paths were statistically significant as indicated by the critical values and their related p-values as shown in Table 5.48. All values of standardized estimates were sound and all were less than 1.00.

Table 5.48: AMOS Output for rural model for unstandardised and standardised estimates

CONSTRUCT			Unstandardised Estimate	S.E.	C.R.	P	Label
PROCESS1	<---	INPUT	1.570	.188	8.359	***	par_26
CHALL	<---	INPUT	-.345	.064	-5.381	***	par_28
PRODUCT	<---	CHALL	-.414	.074	-5.611	***	par_29
PROCESS2	<---	INPUT	0.666	.099	6.723	***	par_33
PRODUCT	<---	PROCESS1	0.545	.069	7.950	***	par_36

CONSTRUCT			Standardised Estimate
PROCESS1	<---	INPUT	0.843
CHALL	<---	INPUT	-0.326
PRODUCT	<---	CHALL	-0.279
PROCESS2	<---	INPUT	0.554
PRODUCT	<---	PROCESS1	0.646

The comparison was made between models from urban and rural schools and also the final model (Model 4) to see the difference between them.

Table 5.49: The moderation test using urban and rural schools

	MODEL 4 (FINAL MODEL)	URBAN	RURAL
X^2/df	3.756	2.730	3.236
CFI	0.943	0.920	0.921
RMSEA	0.060	0.068	0.076
ECVI	1.568	2.425	2.700
INPUT→P1	0.841	0.837	0.843
INPUT → P2	0.470	0.392	0.554

INPUT→CHALL	-0.425	-0.484	-0.326
CHALL→PRODUCT	-0.289	-0.286	-0.279
P1→PRODUCT	0.613	0.597	0.646

From Table 5.49, urban had a slightly better model fit compared to rural. However, since the value of estimates was similar, it could be concluded that the moderator variable which was the school type (urban-rural) did not moderate the causal effects of variables.

5.10.2 Moderating Effect of School Category

There were 350 secondary school teachers and 426 primary school teachers from the sample. From the full structural model that had been finalized, the moderation effect of secondary and primary were examined to all the significant paths. Figure 5.21 showed the output for secondary group and Figure 5.22 for primary group.

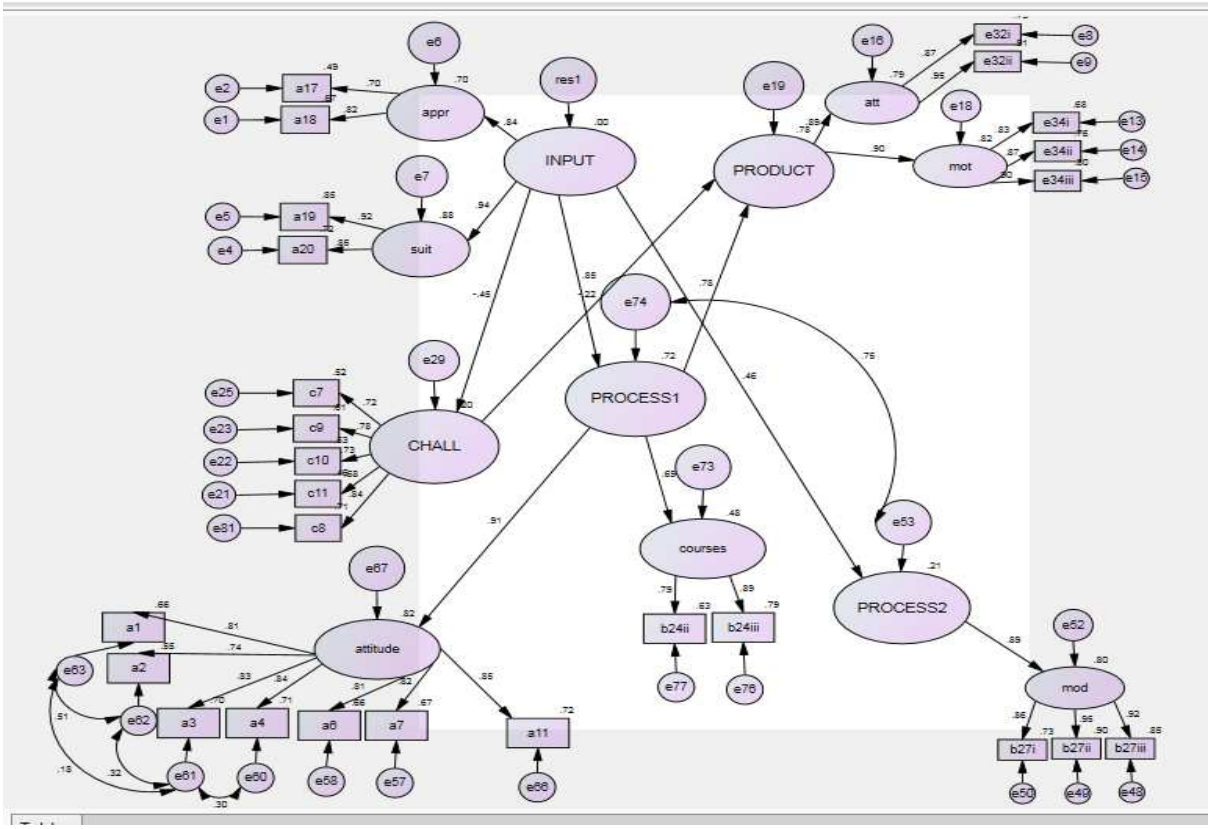


Figure 5.21: The hypothesised moderating model for Secondary school

The ML estimation of this secondary school model yielded several noteworthy results:

1) The overall X^2 value, with 284 degrees of freedom, is 923.495. The chi-square test of overall model fit resulted with a statistically significant discrepancy between the model and the data (X^2=923.495, p=0.000; CMIN/df=3.252).
2) The other overall fit statistics were as follows: GFI=0.855; AGFI=0.821; NFI=0.897; CFI=0.926; TLI=0.916 and RMSEA=0.074 and the ECVI were 2.542.
3) All of the unstandardized estimates for the five structural parameter paths were statistically significant as indicated by the critical values and their related p-values as shown in Table 5.50. All values of standardized estimates were sound and all were less than 1.00.

Table 5.50: AMOS Output for secondary model for unstandardised and standardised estimates

CONSTRUCT			Unstandardised Estimate	S.E.	C.R.	P	Label
PROCESS1	<--	INPUT	1.586	0.176	9.013	***	par26
CHALL	<--	INPUT	-0.506	0.066	-7.679	***	par28
PRODUCT	<--	CHALL	-0.420	0.086	-4.883	***	par29
PROCESS2	<--	INPUT	0.520	0.076	6.807	***	par33
PRODUCT	<--	PROCESS1	0.882	0.114	7.720	***	par36

CONSTRUCT			Standardised Estimate
PROCESS1	<--	INPUT	0.846
CHALL	<--	INPUT	-0.451
PRODUCT	<--	CHALL	-0.221
PROCESS2	<--	INPUT	0.461
PRODUCT	<--	PROCESS1	0.775

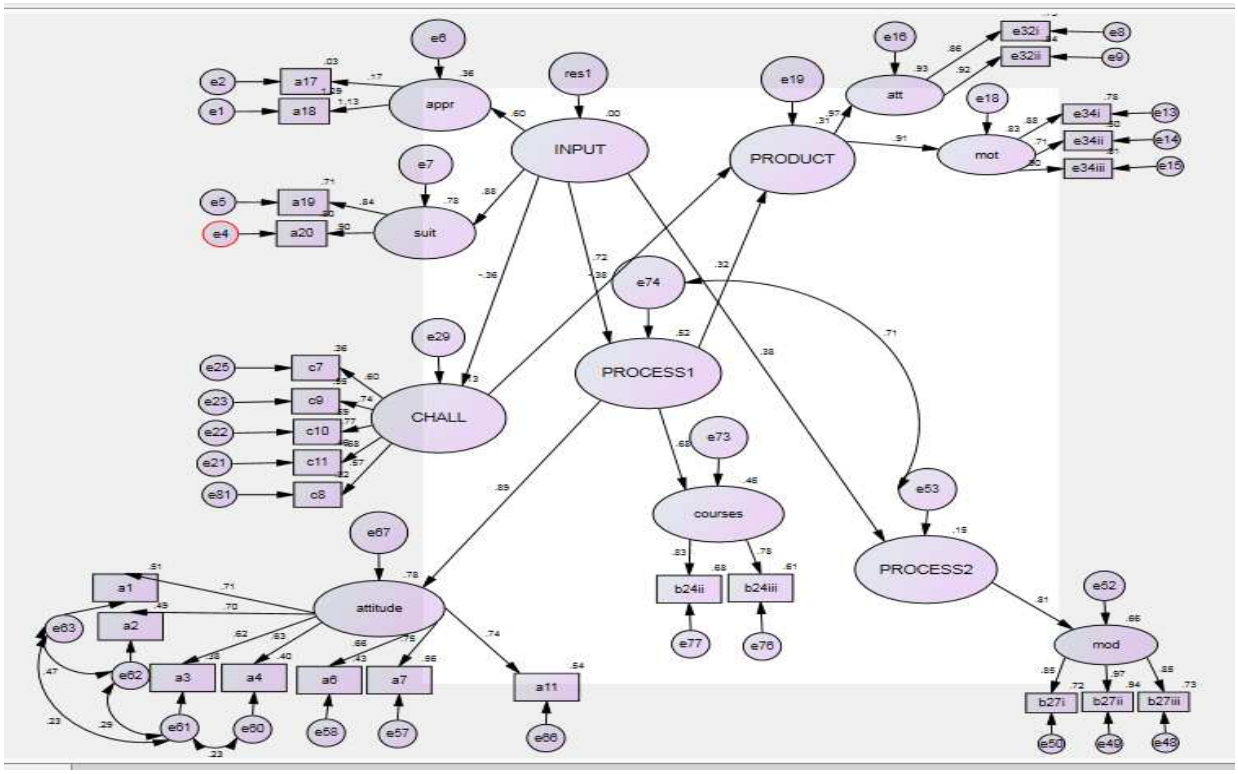


Figure 5.22: The hypothesised moderating model for primary schools

The ML estimation of this primary school model yielded several noteworthy results:

1) The overall X^2 value, with 284 degrees of freedom, was 1029.064. The chi-square test of overall model fit resulted with a statistically significant discrepancy between the model and the data (X^2=1029.064, p=0.000; CMIN/df=3.623).
2) The other overall fit statistics were as follows: GFI=0.811; AGFI=0.766; NFI=0.819; CFI=0.861; TLI=0.841 and RMSEA=0.087 and the ECVI were 3.333.
3) All of the unstandardized estimates for the five structural parameter paths were statistically significant as indicated by the critical values and their related p-values as shown in Table 5.51. All values of standardized estimates were sound and all were less than 1.00.

Table 5.51: AMOS Output for primary model for unstandardised and standardised estimates

			Unstandardised Estimate	S.E.	C.R.	P	Label
PROCESS1	<--	INPUT	1.050	0.155	6.786	***	par_26
CHALL	<--	INPUT	-0.386	0.074	-5.203	***	par_28

	Unstandardised Estimate	S.E.	C.R.	P	Label
PRODUCT <-- CHALL	-0.428	0.072	-5.965	***	par_29
PROCESS2 <-- INPUT	0.417	0.098	4.246	***	par_33
PRODUCT <-- PROCESS1	0.262	0.055	4.756	***	par_36

	Standardised Estimate
PROCESS1 <--- INPUT	0.724
CHALL <--- INPUT	-0.360
PRODUCT <--- CHALL	-0.382
PROCESS2 <--- INPUT	0.385
PRODUCT <--- PROCESS1	0.315

The comparison was made between models from secondary, primary school and final model (Model 4) to see the difference between them.

Table 5.52: The moderation test using secondary and primary schools

	MODEL 4 (FINAL MODEL)	SECONDARY	PRIMARY
X^2/df	3.756	3.252	3.623
CFI	0.943	0.926	0.861
RMSEA	0.060	0.074	0.087
ECVI	1.568	2.542	3.333
INPUT→P1	0.841	0.846	0.724
INPUT → P2	0.470	0.461	0.385
INPUT→CHALL	-0.425	-0.451	-0.360
CHALL→PRODUCT	-0.289	-0.221	-0.382
P1→PRODUCT	0.613	0.775	0.315

From Table 5.52, secondary had a better model fit compared to primary. In terms of relationship between dimensions; all the significant relationships were stronger in secondary schools except for challenge-product relationship which was stronger in primary schools.

5.11 Summary of Hypotheses Testing

In this research, 21 hypotheses were developed in order to answer the research questions as listed in Table 5.53. The main concerns were to test: i) the validity of constructs; ii) the interrelationships of constructs; and iii) the moderating effect of school type and school category.

Table 5.53: Summary of Results of Hypotheses Testing

Hypotheses	Results of the test
H_1: Evaluation in input dimension is explained by three factors	Rejected
H_2: Each indicator has a nonzero loading on the targeted factor	Fail to reject
H_3: Each indicator has a zero loading on the non-targeted factors	Fail to reject
H_4: The error terms are uncorrelated	Fail to reject
H_5: Evaluation in process dimension is explained by twelve factors	Rejected
H_6: Each indicator has a nonzero loading on the targeted factor	Fail to reject
H_7: Each indicator has a zero loading on the non-targeted factors	Fail to reject
H_8: The error terms are uncorrelated	Partially rejected
H_9: Evaluation in product dimension is explained by three factors	Rejected
H_{10}: Each indicator has a nonzero loading on the targeted factor	Fail to reject
H_{11}: Each indicator has a zero loading on the non-targeted factors	Fail to reject
H_{12}: The error terms are uncorrelated	Partially rejected
H_{13}: Process dimension is associated with product dimension	
i) P1 is positively associated with product	Fail to reject
ii) P2 is positively associated with product	Rejected
iii) P3 is positively associated with product	Rejected
iv) Challenges is negatively associated with product	Fail to reject
H_{14}: Input dimension is associated with process dimension	
i) Input is positively associated with P1	Fail to reject
ii) Input is positively associated with P2	Fail to reject
iii) Input is positively associated with P3	Rejected
iv) Input is negatively associated with challenges	Fail to reject
H_{15}: Input dimension is positively associated with product dimension	Rejected
H_{16}: School type moderates the relationship between input and product dimension	
Urban moderates the relationship between input and product	Rejected
Rural moderates the relationship between input and product	Rejected
H_{17}: School type moderates the relationship between process and product dimension	
i) Urban moderates the relationship between P1 and product	Rejected
Rural moderates the relationship between P1 and product	Rejected
ii) Urban moderates the relationship between P2 and product	Rejected
Rural moderates the relationship between P2 and product	Rejected
iii) Urban moderates the relationship between P3 and product	Rejected
Rural moderates the relationship between P3 and product	Rejected
iv) Urban moderates the relationship between challenges and product	Rejected
Rural moderates the relationship between challenges and product	Rejected
H_{18}: School type moderates the relationship between input and process dimension	
i) Urban moderates the relationship between input and P1	Rejected
Rural moderates the relationship between input and P1	Rejected
ii) Urban moderates the relationship between input and P2	Rejected
Rural moderates the relationship between input and P2	Rejected
iii) Urban moderates the relationship between input and P3	Rejected
Rural moderates the relationship between input and P3	Rejected
iv) Urban moderates the relationship between input and challenges	Rejected
Rural moderates the relationship between input and challenges	Rejected

H_{19}: School category moderates the relationship between input and product dimension	
Secondary moderates the relationship between input and product	Rejected
Primary moderates the relationship between input and product	Rejected
H_{20}: School category moderates the relationship between process and product dimension	
i) Secondary moderates the relationship between P1 and product	Fail to reject
Primary moderates the relationship between P1 and product	Fail to reject
ii) Secondary moderates the relationship between P2 and product	Rejected
Primary moderates the relationship between P2 and product	Rejected
iii) Secondary moderates the relationship between P3 and product	Rejected
Primary moderates the relationship between P3 and product	Rejected
iv) Secondary moderates the relationship between challenges and product	Fail to reject
Primary moderates the relationship between challenges and Product	Fail to reject
H_{21}: School category moderates the relationship between input and process dimension	
i) Secondary moderates the relationship between input and P1	Fail to reject
Primary moderates the relationship between input and P1	Fail to reject
ii) Secondary moderates the relationship between input and P2	Fail to reject
Primary moderates the relationship between input and P2	Fail to reject
iii) Secondary moderates the relationship between input and P3	Rejected
Primary moderates the relationship between input and P3	Rejected
iv) Secondary moderates the relationship between input and challenges	Fail to reject
Primary moderates the relationship between input and challenges	Fail to reject

5.12 Chapter Summary

This chapter presented the data analyses and results of this study in evaluating the implementation of SBA in Malaysia. SBA implementation is aimed to produce a holistic and balanced human capital in terms of their physical, emotional, spiritual and intellectual as aspirated in the Malaysian NPE. Its implementation might also contribute to the whole-school improvement process, thus improve student achievement. This study involved the analysis of data in two main phases. The first phase was the analysis of the preliminary data including screening, cleaning process and checking for missing data, outliers and normal distribution of the data. It followed with the presentation of the profile of the respondents. Then it moved on to answer the research questions proposed in Chapter 1. The respondents' perception on the input, process and product evaluation was presented descriptively. Next was the second phase and inferential statistics were involved. At this point, the aims were to test all the hypotheses developed earlier. There were two stages of SEM. The first stage of SEM was to assess the pooled measurement models

using CFA procedures by addressing the issue of uni-dimensionality, validity and reliability of the models. The second stage involved the analysis of the structural model after getting to know that all the three measurement models were operating adequately. But before that, the structural models had gone through few re-specifications in order to get the best fit model. After having achieved the best parsimonious structural model, then the hypotheses testing were carried out.

The results were detailed in the next chapter, in section 6.3 on summary of the findings. To simplify the term in the final discussion, Input was replaced with ‘qualification and equipment’, Process1 with ‘attitude and skills’, Process2 with ‘moderation process’ and Product with ‘students attitude and motivation’. Only ‘challenge’ term was maintained. The next chapter would be discussing details on this research finding together with the implications of the study theoretically, methodologically and practically. These were all done in order to achieve the objectives of the research. It then continued with the limitations of the study and also some suggestions for future research.

Chapter 6: Discussions, Implications and Conclusion

6.1 Introduction

The concept of assessing students' learning has gone through gradual changes, from using excessive dictation exercises, mechanical drills, tests and examinations to the use of formative assessments (Kennedy, 2013). The research in the area of this new form of assessment, formative assessment or assessment for learning is believed to be consistent (Wiliam, 2009). And with the support of the work by Paul Black and Dylan Wiliam in the UK and the US since 1998, formative assessment practises have been developed (Wiliam, 2011). The literature review showed both theories and models on assessment and programme evaluation have influenced the development of the new model of the assessment system in Malaysia. This in turn provides valuable framework for understanding the evaluation of SBA implementation in schools, to aid in identifying the strengths and weaknesses of its implementation.

This final chapter is divided into eight main sections. Following this section, the second section presents an overview of the study. The third section lists the summary of the findings. This section is followed by a detailed discussion of the findings. The fifth section provides implications of the results in terms of theoretical, methodological and practical components. The sixth and seventh sections state the limitations of the research and suggestions for future research. Lastly, there is a brief conclusion of this study.

6.2 An Overview of the Study

The concept of formative and summative assessment is developed from the attributes and components of various models such as the formative model (IowaCORE, 2009), the logic model (Wiliam, 2009) and the SCAP Model (Rust *et al.,* 2005) and learning theories such as behaviourism, Piaget's learning theory, constructivism, multiple intelligence and brain research, all of which are implicit to the assessment system in schools. Formative assessments used in conjunction with summative assessments are hoped to improve a variety of

learning styles, especially learning with deeper understanding and finally to improve the educational standards of students.

Previously, research related to SBA evaluation has focused on one dimension of evaluation only; for example, input dimension (Othman *et al.,* 2013 and Adnan *et al.,* 2009), process dimension (Othman *et al.,* 2013; Salmiah, 2013; Adediwura, 2012; Begum and Farooqui, 2008) or product dimension (Mansor *et al.,* 2013). However, recently attempts have been carried out to determine the relationship between the two dimensions of evaluation, thus linking process and product dimensions (Cheng *et al.,* 2011; Christina, 2009) or input and product (Koh and Velayutham, 2009). Yet, prior to the research work undertaken for this thesis, researchers have ignored the interrelationship between all the four dimensions of evaluation – the context, input, process and product in evaluating the SBA system implementation in schools.

This current study has focused on researching teachers' perception in a recent assessment innovation, the SBA system within the Malaysian education context. In Malaysia, there are 13 states and three federal territories. At the national level, the MES is responsible for the development of a national curriculum and procedures implemented in SBA. Assessments practised are public examinations and SBA. SBA is a holistic assessment which assesses cognitive, affective and psychomotor aspects of students in accordance to the NPE and Standard Curriculum. SBA consists of two main components: academic (SA and CA) and non-academic (PASCA and Psi). This study aimed to assess the context, input, process and product dimensions of the evaluation and to determine the interrelationship between the dimensions. The four general objectives of this study were to: i) identify the input, process and product of the system; ii) develop and validate an evaluation scale; iii) determine the relationships among the evaluation dimensions; and iv) look at the moderating effect of school type and school category on the relationship between the evaluation dimensions. In view of these objectives, a conceptual framework was developed from a comprehensive evaluation framework, 'Stufflebeam's CIPP Model'. After several model re-specifications, two constructs were chosen for input dimension – appropriateness of personnel's qualification and suitability of physical infrastructure and ICT; six main constructs for process – attitude, understanding, courses, moderation, monitoring and challenges and

two constructs for product – attitude towards SBA and motivation towards learning. Two constructs were chosen for context dimension – type of school (urban-rural) and category of school (primary-secondary). All of the three dimensions; - input, process and product were hypothesised to be associated with each other while context (school type and school category) was posited to moderate the relationship between dimensions.

The developed conceptual framework was used to test hypotheses in an attempt to answer the following research questions:

RQ1: What are the teachers' perceptions on the input dimension of the SBA system implementation?
RQ2: What are the teachers' perceptions on the process dimension of the SBA system implementation?
RQ3: What are the teachers' perceptions on the product dimension of the SBA system implementation?
RQ4. Is the measurement scale for input dimension construct valid?
RQ5. Is the measurement scale for process dimension construct valid?
RQ6. Is the measurement scale for product dimension construct valid?
RQ7. Does process dimension influence product dimension?
RQ8. Does input dimension influence process dimension?
RQ9. Does input dimension influence product dimension?
RQ10. Does school type (urban-rural) moderate the relationship between evaluation dimensions in SBA?
RQ11. Does school category (primary-secondary) moderate the relationship between evaluation dimensions in SBA?

6.3 Summary of the Findings

The purpose of this study was to evaluate the implementation of the SBA in schools in preparing a summative report of the effectiveness of the system, in order to guide decision makers and promote understanding of the involved phenomena. Overall, the findings provided evidence indicating a high level of agreement from teachers concerning the effectiveness of the system which might contribute to positive student learning outcomes. Looking at the descriptive analysis, the survey data demonstrated that the majority of

teachers were satisfied with input received from the governments, such as the ICT hardware and space provided in classroom and also the qualifications and skills they have gained but were less satisfied with the physical equipment they had to work with. In addition, they were in need of complete assessment documents, teacher assistants and training. Regarding process dimension, the implementation went well but could still be strengthened. Generally, teachers had a positive attitude towards SBA, understood the main goals of SBA and IHT provided and supported the QA processes of moderation and monitoring. Administration was not perceived as very helpful; and workload and time constraints seemed to be the biggest challenges. Yet SBA was perceived to contribute to the school improvement processes. And, looking at the outcome of the system implementation, teachers thought that they were satisfied with the students' attitude and knowledge but not really with the motivational level especially the students' reading interest.

The study was able to validate some components from the Input, Process and Product dimension. The Input measurement model was validated with two constructs ('personnel qualification' and 'physical infrastructure'). The Process measurement model was validated with six out of twelve constructs ('challenges', 'attitude', 'understanding', 'courses', 'moderation' and 'monitoring') and the Product measurement model was validated with two constructs ('students attitude' and 'motivation').

When tested with SEM, there was an interrelationship between the dimensions of evaluation: a significant positive relationship between Process and Product. In other words, as *process* dimension increases, *product* dimension increases. There was a high and significant positive relationship between 'attitude and skills' and 'students' attitude and motivation' whereas there was a moderate and significant negative relationship between challenges and 'students' attitude and motivation'. It means that, according to teachers, whenever teachers have a positive 'attitude' and gain 'skills', students are more likely to have an improved 'attitude and motivation' towards SBA. A negative correlation indicated that, lower levels of 'challenges' gained by teachers are related to higher students 'attitude and motivation'. There was a significant positive relationship between Input and Process. There was a high and significant positive relationship between 'qualification and equipment' and teachers

'attitude and skills' or 'moderation process'. In addition, there was a high negative relationship between 'qualification and equipment' and 'challenges' faced by teachers. Moreover, there was no significant relationship between 'qualification and equipment' and students 'attitude and motivation'. However, there was an indirect relationship between the two dimensions, whereby input created the conditions for processes to be implemented effectively. No moderation effect of school type (urban-rural) was noted but there was a moderation effect of school category (secondary-primary) found on the relationships between the evaluation dimensions.

This evaluation study was guided by the CIPP Model. In general, it can be concluded that the model developed in this study was partly helpful in explaining the interrelationship between dimensions in SBA implementation. To some extent, the CIPP Model is suitable as a means of explaining the evaluation process of SBA implementation. Elements of CIPP which are applicable in Western context are transferable to the Malaysian context. Results of SEM indicated that the model explained quite a high amount of variance. Only two variables had a direct effect on the product dimension ('attitude and skills' and 'challenge') whereas the input dimension had an indirect effect on product dimension.

6.4 Discussion of the Findings

Generally, the findings of the study revealed that SEM procedures supported the conceptual framework set out in this study. This study clearly presented the importance of the evaluation of system to follow all the four dimensions outlined in the evaluation model by Stufflebeam if we were to evaluate SBA more effectively.

6.4.1 Input, process and product evaluation dimensions of the SBA system implementation

The respondents of the study were 776 school teachers teaching several primary and secondary schools around Kelantan, a state in the north-east of Peninsular Malaysia. Apart from similar distributions between rural and urban type of schools and also between primary and secondary schools where the

respondents implement SBA, there were more female (74.7 percent) than male respondents. Most were Malays (93.6 percent) with their teaching experiences amounting to 10 to 20 years (49.1 percent). The majority (71.0 percent) are educated to degree level. Nearly half of the respondents had attended SBA courses once and have had one year experience in practising SBA in schools.

In answering the first research question, for input dimension as perceived by the teachers, to a certain extent the personnel's qualification, resources and procedures assessed in this study have met the desired result. These findings were in line with Zhang *et al.* (2011) who believed that the use of a complete assessment document, a teaching assistant and suitable training for all the personnel should be highly recommended in order to meet the needs of the given objectives of SBA. This is further supported by the initiatives taken by the MOE in Malaysia. The MOE will conduct additional special training to all teachers, administrative officers in district and state education department across Malaysia upon receiving criticisms and complaints by the teachers (Utusan Online, 2014c). Training will focus on assessment concepts and marking skills regarding the latest SBA improvement system (Sinar Harian, 2014). Up until 2014, there was still no teaching assistant in Malaysian schools (KPPK, 2014a). It is believed that Malaysian schools must have teaching assistants, not only due to the large class size but also to help the weak students or those who need more attention (Nasir, 2011). As for the complete assessment document, it was greatly needed by teachers.

In addition, with nearly three quarters of the respondents having had more than ten years of teaching experience and holding a degree in education, it might be assumed they would have equipped themselves with suitable qualifications and skills in assessment. However, research has found that there was little or no correlation between teacher qualifications or teachers' college grade and student progress (Wiliam, 2011). Furthermore, studies also found that attempts made to improve the quality of teachers already in the profession did not show a positive outcome. Since the process of deselecting existing teachers or improving the quality of entrants into the teaching profession also has its own limitations and could take years to give a noticeable impact on students' achievement, one way to improve teacher performance is to develop

the practise of serving teachers with courses and workshops along their teaching profession.

In answering the second research question, the process dimension of SBA evaluation is supposed to provide a complete description of the actual programme activities. Findings from the study indicated that teachers' belief about and attitude towards SBA were positive except for their feeling towards SBA as something burdening them. These findings were consistent with those of Othman *et al.* (2013) and Hamzah and Paramasivan (2009). Similarly, a survey conducted with 518 New Zealand primary and secondary school teachers suggested that teachers do have positive attitudes towards assessment (Brown *et al.,* 2012). Teachers believed that providing feedback to students is an integral part of teaching. However, they believed that feedback was meant to be used more in improving students' work and developing autonomy amongst them and less in enhancing well-being of students such as self-esteem or in providing task-oriented information to students.

The main challenge in implementing SBA was that the work is burdensome to teachers. Most respondents felt that SBA was burdening them although they believed that SBA was a good alternative to public examinations and their readiness level towards practising SBA was quite high. This finding supported the findings highlighted by teachers from previous research by Salmiah (2013), Chan and Gurnam (2012) and Faizah (2011). Recently, this problem has been highlighted by most teachers in Malaysia together with the support from the National Union of the Teaching Profession (Utusan Online, 2014c). Excessive workload issues have also been highlighted by Tan (2010) in her book on SBA implementation entitled 'SBA in Malaysia: Teachers' readiness, Issues and Implementation Guide'.

The second most challenging factor found in this study was no special recognition for teachers' performance. Nor Azni *et al.* (2014) in their conceptual paper on SBA implementation suggest that instructional leaders need to provide suitable incentive to teachers such as giving reinforcement to the excellent performance during meeting, through memos or school bulletin, to give recognition in a written form or to reward teachers with opportunities to develop their professionalism. The next challenging factor was the difficulty

to get related resources on SBA. A lack of proper knowledge and information will further make things worse. Davison (2007) pointed out similar issues with English language teachers conducting SBA in Hong Kong. Teachers felt that the technical resources like video cameras are necessary for reviewing few students' performance but unfortunately such resources were not accessible in their school.

A large number of respondents were aware of the courses and the role in improving their skills in assessing students yet, nearly a quarter were unsure about it. One possible explanation for this uncertainty was that some of them were not really practising SBA in their classroom. This was supported by Talib *et al.* (2014) who revealed that almost 80.0 percent of the Malaysian primary school teachers who were selected using clustered sampling from the ten districts believed that their level of AfL practises in the classroom were at the two lowest levels – unsatisfactory and basic. In other words, teachers attend courses but a failure to apply the training in their classrooms can create a state of 'ignorance' about the training and the materials and information delivered to them during the workshops or briefings.

A large percentage of respondents from this study disagreed with or were unsure about the potential of SBA in reducing absentee rates. This seemed to contradict what SBA has to offer as discussed previously in the literature review chapter. Reducing absentee rates is one of the main objectives in SBA implementation in Malaysia. It might be due to lack of knowledge amongst educators on the role of SBA in reducing absenteeism whenever SBA is implemented effectively. And yet school absenteeism in Malaysia is the biggest contributor to the problem of student discipline (Bernama, 2010). From ministry records, it has been shown that in 2010, absenteeism contributed 17.5 percent of cases (19, 545 cases of absenteeism out of 111, 484 cases of discipline problems). So, as a newly-implemented assessment system, teachers might not realise that the difference of attendance rates, if there is any, is influenced by teacher practises in classrooms.

In answering the third research question for product dimension of SBA evaluation, looking at the broad effects of the programme, the results of this study showed that the teachers thought that the students have made a

difference in terms of their attitude and knowledge on SBA and also their motivation towards learning but such changes were limited. A greater percentage of teachers in this present study agreed that SBA influences students' attitude towards assessment in a positive way. This finding corroborated previous studies in other parts of the world such as Hong Kong and New Zealand. In Hong Kong, a study by Lo (2006) found that students had positive attitudes towards assessment. Students believed that individual oral feedback from teachers allowed them to clarify specific learning points which they did not understand during teaching time or in feedback. Furthermore, feedback gives them opportunities to improve their English sentences and composition. A study in New Zealand also presented similar findings. The study used a mixed method and was conducted to 100 senior ESOL (English for Speakers of Other Languages) students in a secondary school in New Zealand (Feng, 2007). The findings revealed that students' attitudes towards formative assessments varied depending on the assessment activities conducted by teachers in their classroom. Most students favored feedback as it allowed them to know their strengths and weaknesses and also favoured self-assessment because it allowed them to understand their mistakes and helped them more on process rather than product. But students did not favour questioning because of their limitation in English language, less waiting time given by teachers and cultural sensitivity. They also did not favour peer-assessment as they did not trust their peers in marking their work.

In terms of students' motivation towards learning, SBA did increase students' interest in subjects and enabled them to determine their strengths but did not quite improve their reading interest. From the findings, according to teachers, reading habits of students seem not to have shown a great deal of improvement since SBA has been implemented. Such findings are contrary to data gained in 2010 from NILAM, a reading activity programme which has been operating since 1999 (Wan Zah *et al.,* 2005), which showed that students' involvement in reading had improved significantly. The average reading rate for primary school children was 20 and 15 books and for secondary school children, it was 10 and 8 books, for the year 2010 and 2008 respectively (Kementerian Pendidikan Malaysia, 2014). These findings from the MOE might have contributed to the difference in findings of this study. The difference

might also be due to the fact that SBA is newly introduced; and teachers are more prone to relate reading habits to the NILAM programme rather than SBA.

So, the descriptive findings suggested that: i) the teacher training in SBA should be enhanced or be given a more quality and frequent training and briefings such as an improved model compared to the present Cascade Model used by the Ministry. The Cascade Model allows dilution of information and hence its effectiveness is questionable (Norzila, 2013); ii) a teaching assistant position should be formed and the physical infrastructure and ICT should be upgraded; iii) resources and documents on SBA supplied to schools should be improved; iv) workloads should be revised and lessened; and v) special recognition for teachers would be beneficial.

6.4.2 The measurement scale for input, process and product evaluation dimension

a) Input dimension measurement model

Findings from the research failed to support the hypothesised relationship between the second-order construct and the first-order constructs. The results indicated that the final modified CFA model of input dimension consists of two 1st-order constructs instead of three – 'appropriateness of personnel's qualifications' and 'suitability of physical infrastructure and ICT' with four items altogether. One 1st-order construct, 'material and personal needs' with items on complete assessment document, teaching assistant and training of personnel was deleted due to its low factor loading value and the data also clearly showed its univariate non normality of data and the reduced variance. So, the findings clearly showed that, in order to assess input components in achieving the desired result in SBA implementation in schools, look for teachers' qualification and skills and also the physical equipment and space in the classroom.

b) Process dimension measurement model

For the process dimension, there are four measurement models – Process1, Process2, Process3 and challenges. Initially, Process1 consisted of five 1st-

order constructs. Finally, after a few model re-specifications, it ended up having three 1st-order constructs: 'attitude', 'understanding' and 'courses (skills)'. The removal of the 'readiness' construct from attitude construct did not significantly change the content of attitude construct as it was conceptualised. Moreover, these findings were in line with the theoretical arguments on attitude construct proposed by Narayanan *et al.* (2008) which consisted of 'belief and feeling' constructs as were outlined in the literature review chapter. These findings were also consistent with Heritage (2007a) who stated that those three components – knowledge, skills and positive attitude of teachers in formative assessment are the components that are highly required if teachers were to implement formative assessment effectively.

Process2 consisted of two 1st-order constructs – moderation and monitoring process. These constructs were in line with the SCAP model (Rust *et al.,* 2005) which asserted that moderation and marking are one of the main processes in assessment, and if they are implemented wisely could result in a significant improvement in the students' work. These two constructs are the main components in QA process of SBA practised in the Malaysian educational system to produce a valid, reliable, feasible and fair assessment (Lembaga Peperiksaan, 2012).

Process3 consisted of two 1st-order constructs – role and importance of SBA which added to the school improvement process in SBA. The observed variables under these two main constructs include the role of SBA in improving QA process, public confidence towards school, students' assessment practises and their various learning opportunities, teachers' professionalism and their decision making in assessment reform. All of these processes are strategies which could lead to a change in school culture and in turn could improve students' achievement. These findings were consistent with suggestions by Hopkins (in Harris, 2003b) who asserted that to improve students' achievements, strategy has to focus on two things: modifying classroom practises and adapting management arrangements within the school. Or, in other words, it is by building school-capacity through an effective assessment implementation. When Process3 measurement model was assessed on its own, it was valid. However, when pooled CFA was conducted, Process3 was highly correlated with Process1 and Product, thus discriminant validity was not

achieved. So, Process3 was deleted from process measurement model. A possible explanation for this is the respondents might perceive that all the items regarding processes to improve school are similar to their attitude and understanding towards SBA and the outcome of SBA implementation.

For 'challenge', out of 12 indicators, only six indicators remained to measure the one-factor model of challenge. These represented school climate, insufficient financial resources, extra workload, no special recognition to teachers, teachers' integrity and responses are only memorised by students. These indicators had a moderate factor loadings ranging from 0.589 to 0.749 indicating that the meaning of challenge has been preserved by these indicators. Kennedy (2013) in a seminar addressing SBA prospects and realities in Asian contexts pointed out that the main challenges in SBA implementation processes for Asian societies were increased workloads, lack of community confidence, lack of confidence by teachers and drilling of assessments by students. Excessive workload is not a new theme in SBA implementation. When Hong Kong implemented SBA in 1978, which they referred it as 'Teacher Assessment Scheme', the same main challenges were presented, namely, excessive workload, lack of supporting resources and teachers' integrity (Kennedy, 2013).

c) Product dimension measurement model

The results indicated that the final modified CFA model of product dimension consists of two 1st-order constructs out of 3-factor model as in the hypothesis – 'students' attitude towards SBA' and 'students' motivation towards learning'. A construct on 'knowledge on SBA' was deleted as it had the lowest factor loading and overlap of content with the other two constructs in product dimension. These findings were in line with the logic model (Wiliam, 2009) concerning student outcomes. This was also in line with the objectives of SBA implementation as proposed by the Malaysian government, which are to get an overall picture of students' potential, rather than simply the cognitive aspects of them. *Potential* includes various abilities that someone has which can be developed to help that person become successful; for example cognitive potential, talent, affective potential, physical or leadership. Furthermore, motivation and self-esteem of students are seen to be important elements in

influencing learning whenever assessment is practised for the purpose of improving learning (Wiliam, 2011). These two constructs, 'attitude' and 'motivation' of students seemed to limit the outcome of the system evaluation but in doing research, the number of items in a survey should be greatly considered. So, the findings clearly showed that to evaluate product components in assessing the impact of SBA on students' learning, focus on students' attitude towards SBA and also their motivation towards learning.

6.4.3 The relationship among evaluation dimensions

In this section, the study tested the hypotheses that there were interrelationships between evaluation dimensions of SBA implementation. Three main hypotheses (H_{13}, H_{14} and H_{15}) were developed as shown in Table 6.1.

Table 6.1: Hypotheses and summary of results for the relationships between evaluation dimensions

Hypotheses	Result
H_{13}: Process dimension is associated with product dimension	
i) Process1 is positively associated with product	Fail to reject
ii) Process2 is positively associated with product	Rejected
iii) Process3 is positively associated with product	Rejected
iv) Challenges is negatively associated with product	Fail to reject
H_{14}: Input dimension is associated with process dimension	
i) Input is positively associated with Process1	Fail to reject
ii) Input is positively associated with Process2	Fail to reject
iii) Input is positively associated with Process3	Rejected
iv) Input is negatively associated with challenges	Fail to reject
H_{15}: Input dimension is positively associated with product dimension	Rejected

There were nine related hypotheses developed to test the interrelationships between the three evaluation dimensions. Findings indicated that, out of nine paths, five paths were found to be significant and four paths were non-significant. Structural model 4 (final model) in Figure 5.16 showed a complete picture of it. After the structural model was finalised, the constructs of the dimensions were also finalised. From now onwards, Input is written as 'qualification and equipment', Process1 is written as 'attitude and skills', Process2 as 'moderation' and Product as 'students' attitude and motivation'. Only 'challenge' is maintained.

The data indicated that process was associated with product dimension of evaluation. This relationship indicated that in general, according to teachers, processes implemented by teachers to facilitate SBA could influence students' attitude and motivation towards learning. These findings were consistent with the CIPP Model (Stufflebeam, 1971b) which has been discussed earlier in the literature. In addition, they were also in line with the learning theories, the formative model, the logic model and the SCAP Model.

a) The relationship between teachers 'attitude and skills' and students 'attitude and motivation'.

Teachers 'attitude and skills' was associated to students 'attitude and motivation' with a strong positive relationship. The relationship between the two dimensions was supported, hence enabling an expansion of theories developed previously to the Malaysian context in evaluating SBA system. A significant and strong positive relationship between these two 2nd-order constructs indicated that teachers' who have a positive belief and feelings towards assessment and have received information especially on suitable skills in assessing students are more likely to enhance students' attitude towards assessment and their motivation towards learning.

Research by ARG (2002) suggested that providing students with explanations on the purpose of tests and constructive feedback that helps further learning or developing their self-assessment skills and the use of criteria relating to learning rather than using test performance are some of the actions to improve students' motivation. All of those skills can be new to teachers who have never had any exposure on formative assessment before. So, in order to improve students' learning, particular actions have to be taken to deliver these kinds of skills to teachers effectively, and one of the most effective ways is through teacher training. Not only materials such as modules, workbooks and guidebooks on assessment delivered to teachers during teacher training are useful in improving skills, but workshops also have their own role in establishing learning communities amongst teachers (Bennet, 2011). Similarly, Kennedy (2013) believed that in the Asian education system context, in order to gain community support to perceive SBA as a fair and reliable instrument in

assessing students as how public examinations have been treated previously, the most important thing is to develop teacher capacity.

These findings are supported by a case study conducted in Malaysia looking at the use of various formative assessment techniques in Bahasa Melayu subject such as questioning technique, using scratch cards, loud pronunciation technique, matching technique, assignments, discussions and singing technique in improving ways to gain information on students' learning without having to use too much time (Suzana and Jamil, 2012). The techniques also improve students' attitude and knowledge on assessment. And in order to use various techniques, teachers would have to equip themselves with appropriate knowledge and skills from courses and training that are proposed to them and also have a positive attitude towards the system. The findings of the present study are consistent with those of Dorman *et al.* (2006) who found that the quality of assessment tasks and classroom environment were linked positively to students' attitudinal outcomes. Studies have also found that the quality of assessment tasks and classroom environment very much depend on the specific knowledge and skills on assessment that teachers have (Heritage, 2007b). So, in short, it can be said that teachers' knowledge and skills are a very important mechanism in improving students' attitudinal outcomes. Furthermore, Heritage also added that such knowledge and skills are of no importance unless teachers had a positive attitude towards the great value that assessment can have in improving student learning. On the contrary, attitudes alone without proper skills also could not produce good outcomes.

Research conducted at the University of Sussex, United Kingdom, investigated the attitude of students and teachers towards student assessment (Iberson, 2012). It was found that both teachers and students have a positive attitude towards students' involvement in self-assessment. Furthermore, both parties also felt that students can make accurate judgments in measuring their own performance as long as sufficient training is provided to them. This again is consistent with Heritage (2007) who strictly believed that one of the most important skills for teachers to master in successfully implementing formative assessment is the ability to teach students how to self-assess their own learning and assess their peers' learning.

b) The relationship between 'Challenge' and students 'attitude and motivation'

'Challenge' is negatively associated with students 'attitude and motivation' with a moderate relationship between them. The negative significant relationship indicated that teachers facing too many challenges are less likely to produce students with a positive attitude towards assessment or an increased motivational towards learning. The main challenges involved were extra workload, insufficient financial resources, no special recognition, teachers' integrity and the act of memorizing by students. A study found that physical resources like desks, chairs, photocopy machines, projectors, books or printed media are needed to help teachers in practising formative assessment effectively in a classroom (Reyneke *et al.,* 2010). Furthermore, a lack of paper provided by schools to teachers can also make teachers feel demotivated to implement SBA while each student needs enough relevant assessment tools during the implementation process.

The present study found that teachers facing too many challenges are less likely to produce students with a positive attitude towards assessment. This finding is also aligned with a previous qualitative study by Malakolunthu and Hoon (2010). According to the teachers, with the new form of assessment regarding school-based Oral English Assessment, they were not given clearly stated criteria for grading and the assessments were very subjective, given the teachers listened to the students in general and then appointed them with suitable mark based on a given band. Rubrics or marking guidelines developed by different teachers could lead to variations in scoring amongst teachers (Reyneke *et al.,* 2010) resulting in inconsistencies in the grading of students' performance and thus in turn can influence the validity and reliability of final score.

However, using SEM analysis, a relationship of -0.289 indicated only a moderate negative relationship between challenges and the outcome of SBA implementation. One may suggest that it may not be worth countering these challenges as a great concern as they are normal processes when it comes to a newly-introduced system. Furthermore, findings by Wei (2010), who conducted a three semesters' circular action research to 227 students in the Chinese College English teaching context in China, found that SBA has proved to

benefit students in terms of an increased intrinsic motivation, to strengthen and sustain their motivation, to help them to understand more on their strengths and weaknesses and also the role of self-evaluation and many more. These few benefits are very similar to findings of this study. All of these potential benefits were gained despite facing challenges such as time-consuming, labour intensive work needed, lack of training in formative assessment, low reliability of assessment tasks and large class sizes. So, one possible explanation for this is that challenges do not have a significant influence on the product dimension as long as teachers used SBA appropriately and consistently with the availability of experienced teachers, as found by Wei.

In spite of the moderate relationship between challenges and students 'attitude and motivation' found in this study, it became apparent that instances in which efforts taken by government to reduce challenges have made teachers more willing to accept it. When the challenges of SBA implementation were improved by the Malaysian government after obtaining views from 55,000 people including academics and non-governmental organisations, the improved version of SBA appears well-accepted by the majority of educators (Bernama, 2014). Teachers could then focus more on their teaching and learning process as the improvements involved reducing data that needed to be keyed-in online. So, it clearly shows here that reducing challenges does matter but it has to be understood according to the school context. For example, in most urban schools, computer and internet access might not be a problem, whereas this is more problematic for rural schools.

c) The relationship between 'moderation' and students 'attitude and motivation'

There is no statistically significant relationship between moderation process implemented in schools and students' attitudes towards SBA and motivation towards learning, as perceived by the teachers. Looking at the descriptive data, it showed that most teachers implemented moderation and monitoring processes and were quite aware of the processes.

A valid measurement model formed in this study consisted of these two processes, moderation and monitoring but when the structural model was

formed, the monitoring construct was deleted. So, why was moderation process not significantly related to product? The possible explanation is that teachers might believe that their preparation of scores and evidences and marks submission to the central system conducted at school has nothing to do with students' attitude and motivational level. The contradictory finding suggests that there is a difference between Asian and Western teachers' perceptions in moderation processes. As explained by the SCAP model (Rust *et al.,* 2005), moderation discussions among educators on completion of the marking could improve commonly informed understandings between them. Then, it could improve their standardisation of marking which in turn could enhance their active engagement with feedback. And feedback seems to be the most potential components in promoting student learning and achievement if it is practised in an effective manner (Nicol and Macfarlene-Dick, 2004). Instead, good feedback practises encourage students to construct their positive motivational beliefs and their positive attitude towards assessment. The moderation process is a challenging task if teachers are new to it (Klenowski, 2013). Teachers therefore need plenty of time to understand the process because it involves the development of shared understanding between teachers about the interpretation of standards. In addition, teachers should also be given an opportunity to defend their judgment or to provide supporting evidence from students' work. This is not what happen with the Malaysian teachers. Not all teachers are involved directly in the moderation process as the internal moderators include only selected teachers, excellent teachers and the head teacher. On the contrary, Kennedy (2013) believed that the problem with SBA implementation in Asian context is not because of its poor assessment process but due to its unconducive conditions. He added that it is not an easy job to gain community trust and confidence in SBA to the extent that they have achieved with public examination.

Taken together, it is not too much to claim that the findings of this study revealed several issues worthy of further research. For example, to what extent do teachers practise the idea of giving students a chance to improve their task so that their scores can be improved after the moderation process has been done is yet to be discovered. And to what extent teachers try to improve their teaching practises after getting feedback from both parties, the external and internal moderators is also yet to be discovered. Furthermore, moderation

process is very important in SBA implementation. As in the case of Malaysia, moderation together with monitoring, tracking and mentoring processes were meant to contribute to the QA process proposed by the Malaysian government to produce a valid, reliable, feasible and fair assessment (Lembaga Peperiksaan, 2012).

Furthermore, as these two processes: moderation and monitoring - are the main aspects in QA process in Malaysia besides tracking and mentoring, these findings showed a weak QA process. As a reflection, these results suggest that the system implementation needs to review the procedures of moderation and monitoring processes conducted in schools. It might be that teachers do not really understand the role of these procedures in SBA implementation and they are doing it for the sake of administration only. Those groups and individuals who understood their role most might only be those officers from the ministry, the head teachers and head of panels. Training and courses should therefore be focusing more on these processes, especially on how they could be implemented and their role in relation to teachers' active engagement with feedback, so that students' outcome can be improved by those processes. In addition, training should not focus on marking scheme as a fragmented process without realising the link between the processes as have been set out in the SCAP model.

d) The relationship between Input and Process

Findings from this study provided evidence to support that input was associated significantly with process implemented in schools. This relationship enables the expansion of theories developed previously to the Malaysian context in evaluating SBA. Input dimension consists of 'qualifications' of teachers and 'physical equipment' in school. The positive relationship between both dimensions indicated that i) input facilitated teachers to have positive attitudes towards SBA and improved assessment skills from the courses that they have attended; and ii) input facilitated moderation processes amongst teachers. And iii) input was found to influence teachers' perceptions towards SBA as being less challenging. As pointed out by Wiliam (2011), formative assessment itself is a process, although in practise it is seen more frequently as an instrument rather than process. So, theoretically, from the findings it can

be concluded that any personnel, resources or procedures implemented in achieving SBA objectives would probably have an impact on formative assessment.

These findings were found to be consistent with the relationships proposed by Stufflebeam in the CIPP model. Furthermore, they were also in line with the learning theories, formative model, Logic Model and the SCAP Model. In addition, Heritage (2007b) pointed out that specific knowledge and skills in assessment are very important for teachers successfully implementing formative assessment. She further emphasised that it has to be made compulsory for pre-service teachers to attend training programmes related to knowledge and skills on assessing.

The findings of this research corroborated previous studies in the western and Asian context. A longitudinal quasi-experimental intervention study has been conducted in Singapore over two school years for the year 2006 and 2007 (Koh and Velayutham, 2009). When teachers were provided with ongoing, sustained professional development focusing on authentic assessment task design and rubric development for two years, teachers' assessment task and the quality of students work improved. Similar results are produced when a programme called 'Embedding Formative Assessment' was provided to the teachers in Cannington, a district in Greater London in the UK (Leahy and Wiliam, 2009). During this one-day training session, teachers were given briefings on formative assessment research, an overview of the five key strategies and 30 different techniques in formative assessment. In addition, all of the teachers on the session were provided with a short (30-page) booklet on formative assessment and its application. At the end of seven months, it was found that the progress made by the teachers was positive. In addition, the principals of the secondary schools were quite positive and were looking forward to the programme in the following year. So, the findings of this study seemed to bridge the western/Asian divide within the implementation of SBA.

e) The relationship between Input and Product

Input and product dimensions are assessed based on the CIPP Model (Stufflebeam, 1971a) and have been used by many researchers in various

contexts since the 1960s (Stufflebeam and Shinkfield, 2007). The CIPP Model showed an interrelationship between the two dimensions. The findings in the western context found that input could foster product. For instance, Newmann *et al.* (2000) in the Newmann's Model found that technical resources such as high-quality curriculum, books, assessment instruments, laboratory equipment, computer or adequate work space are one of the main aspects of school capacity to offer instruction which could boost student achievement. Theoretical arguments proposed by Siobhan Leahy in the Logic Model (Bennet, 2011) also supported the above statement. The presence of technical resources will guide teachers to elicit evidence from students and use them to adapt their teaching and learning to meet students' immediate learning needs. At the end of the intervention, what is expected is improved student learning outcomes. Furthermore, Bell and Cowie (2001a) stated that classroom factors such as the layout of the furniture, student-teacher interaction, opportunities provided from the teaching and learning activities or teacher questioning technique can influence the process and product of formative assessment implementation in classrooms.

From the above literature, there is a strong suggestion that input is related to product of a system. Therefore, it was expected that these two evaluation dimensions would have a significant positive relationship. In this study, the relationship between input and product was not statistically significant indicating that input was not positively associated with product. However, there was an indirect relationship between input and product as there was a strong relationship between input and process, followed by a strong relationship between process and product. Looking at the descriptive data from a 5-point Likert scale for distribution of response, little difference was found in terms of manifestation from the respondents towards input and product compared to constructs from process dimension. So, in this case it is not appropriate to conclude that all the items assessed in input dimension were not suitable for use in this study.

6.4.4 Moderating Role of School Type and School Category

'School type' consists of urban and rural schools. By definition, *urban schools* refers to their status when they were firstly registered and located inside the

Local Council administration in each district; the status of *rural schools* are those refers to when the schools were first registered and located outside the Local Council administration (Emisportal, 2013). Hypotheses H_{16}, H_{17} and H_{18} were developed to test the moderating effect of school type on the relationship between the evaluation dimensions. Findings failed to support the moderating effect of school type (urban-rural) on the relationship between all the significant paths. The possible explanation for this is it might be due to the classification of urban and rural type of schools. Schools might be registered as rural, but after some time, developments arise in that area which makes them to be similar to urban schools standard. Shaharudin (UKM, no date) states that it is timely to reassess the classification of urban and rural schools in Malaysia since the demarcation between urban and rural used by MOE in defining school type is not in line with the development process in the country.

School category consists of primary and secondary school. Hypotheses $H_{19,}$ H_{20} and H_{21} were developed to test the moderating effect of school category on the relationship between all the dimensions. School category was found to moderate all the significant relationships. However, the findings failed to support the moderating effect of school category on the relationship between input and product dimension (H_{19}) since the path is not significant. The findings suggested that the tendency of i) input influencing process dimension of evaluation; and ii) teachers' 'attitude and skills' influencing product dimension is more likely to happen in secondary schools rather than primary schools. Only Challenge influencing product is more likely to happen in primary schools. In general, the relationship of input-process and process-product relationship are stronger in secondary than primary schools. In a more general perspective, recent research has found that secondary school teachers held stronger concerns regarding the implementation of SBA in Malaysia compared to primary school teachers (Aidarwati and Abdul Ghani, 2013). Teachers' concerns include: i) awareness towards SBA implementation, involvement with SBA and the eagerness to learn more about SBA; ii) knowledge and ability on some general aspects of SBA such as its rationale, requirement, demands, moderation mechanism, types of support provided and their role; iii) management such as time management and the planning of the assessment tasks; iv) collaboration and support such as teachers' collaboration with their colleagues and other schools in enhancing the effect of SBA or

sharing information; and v) impact of SBA on students in terms of how SBA affects students' performance.

In Malaysian schools in 2014, there were 89 percent and 28 percent graduate teachers in secondary and primary schools, respectively (KPPK, 2014a). The Malaysian Government intends to increase the percent to 90 and 60 respectively by the year 2015, by allocating RM1.4 billion for the Graduate Teacher Program. Through this programme, it is hoped to increase knowledge and skills of teachers. Looking at the data descriptively, a mean value for number of days attending SBA courses for secondary and primary school is 1.16 and 1.48 times respectively; and for a mean value for experience in practising SBA is 1.14 and 1.80 years, respectively. This seems to contradict with the above statement which stated that secondary schools moderated most relationships between dimensions. The explanation for this is, at the time data was collected, secondary schools had been implementing SBA for a year whereas primary schools have been implementing it for two years already. So, basically, there are some stronger relationships between factors in the model of this study from secondary schools compared to primary schools.
In conclusion, this study clearly points the importance of the interrelationships between dimensions as outlined in the evaluation program research. It thus provides some support for the effectiveness of the SBA system implementation and also for the theoretical model such as the formative model, the logic model and the SCAP model. However, this study also suggests that in the Malaysian context, SBA is a difficult and lengthy process to implement effectively but one that is not impossible. Yet, how far SBA could present itself as an alternative educational assessment still has to be continuously evaluated.

6.5 Implications of the study

This study has attempted to play a significant role in making the educational communities aware of research on assessment and the evaluation of its implementation in schools. Furthermore, it is hoped to bring research data and findings to the attention of policy makers, researchers and practitioners. The implications of the findings of this study with regard to theoretical, methodological and practical are presented in this section.

6.5.1 Theoretical Implications

This study has given the opportunity for the issue on SBA to be expanded especially within the school context in a non-western country, namely Malaysia by expanding the existing body of knowledge on SBA evaluation. Although research on SBA is emerging in the literature but most of them come from the western world that might not be appropriate to the non-western scenario since most of the western countries have implemented SBA for quite a long time ago. In addition, little is known about the interrelationship between all four of the evaluation dimensions in SBA. So far, research has focused on the interrelationships between two dimensions only (Koh and Velayutham, 2009; Christina, 2009) This is also the main reason why the researcher has been very determined to conduct this research – investigating the interrelationships between all the four evaluation dimensions in SBA in a country which has a very short experience in the implementation. Furthermore, to the best knowledge of the researcher, this is the first study to empirically link all the four dimensions of SBA evaluation.

This doctoral research highlighted the fact that the interaction between all the four dimensions of evaluation – context, input, process and product factors as have been detailed out by Stufflebeam (1971a) - may be the case in western culture. However, in the Malaysian culture, the main implication of the strong association between input and process dimension is that resources and procedures supplied to schools should help teachers to implement SBA effectively. In other words, more input provided is more likely to support SBA implementation in schools while less input could diminish the process. Secondly, some processes also have strong association with product. Thirdly, input has an indirect relationship with product and lastly, moderation process implemented by teachers has no significant relationship with product. The last finding could be due to a lack of understanding on the role of the moderation process and also the fact that this system is newly-implemented. Furthermore, the differences in demographic characteristics of respondents should also be considered. Malaysia is a multicultural society with Malays dominated 50.33 percent, followed by Chinese (21.76 percent) and Indians (6.52 percent) (JPM,

2014). Moreover, in the Kelantan population, Malay contributes to 92.47 percent, Chinese is 3.14 percent and Indians is 0.27 percent. In addition, statistics gained on the 31st of August 2014 showed that there were 69.79 percent female teachers and 30.21 percent male teachers in Malaysia (Emisportal, 2014). The concept of input dimension proposed in this study needs further study, especially to determine the extent to which input actually improves product. But even in the absence of empirical findings on the direct relationship between input and students 'attitude and motivation', this study may have important implications for school leaders and policy makers. The literature discussed in this study on various perspectives on the concept of process dimensions of SBA and how it is strongly influenced by input and at the same time showed a strong relationship with students' 'attitude and motivation' present a good result in some school improvement initiatives. And the failure to improve input to address formative assessment implementation comprehensively may be a major reason for the disappointing results in achieving the goals of SBA which has been enacted by the Malaysian government. In conclusion, the dynamic interplay between the dimensions as outlined in the CIPP Model is supported by this study as being central importance to programme evaluation.

6.5.2 Methodological Implications

Findings of this study highlighted the fact that the choice of using a statistical analysis technique named SEM to simultaneously examine a series of interrelated relationships between variables to be promising. The measurement model was developed based on theories to determine the suitable items to measure specific constructs of interest. Byrne (2010) stated that the CFA technique is suitable to confirm a hypothesised measurement model developed from the theories when the researcher has had some knowledge on the latent constructs of the model. From there, the researcher postulates a full latent variable model which consists of both - measurement and structural models. By using SEM, data analysis is for inferential rather than descriptive purposes so the hypotheses testing of the study can be conducted. SEM provides explicit estimates of the error variance parameters which improves the accuracy of the relationships between factors of the hypothesised models especially when the errors are sizeable (Byrne, 2010). Furthermore, analysis related to the effect of

moderation variables in the relationship path between variables conducted using the multi-group CFA method also gives an alternative method to the researcher. In addition, it is more practical than using the normal modelling procedure using interaction terms as it could avoid model convergence problem or model misfit (Zainuddin, 2012).

This study developed items on school improvement process referring to the concepts suggested by Harris (2003b). Despite the strong concept proposed by Harris, findings in this study triggered some doubt on the suitability of the items. Perhaps, the items such as 'SBA reduces absentee rate' or 'SBA increases public confidence towards school' were perceived by the respondents as being similar to the items in the product dimensions. So, the suggested items for future study might be, 'Do teachers receive updated information on their SBA practices?', 'Do teachers recognize a direct involvement of students in learning?' or 'Do teachers benefitted from the collected score of students?'. However, these items still have to be empirically tested to ascertain whether they would be better in determining the processes that contribute to SBA implementation. It might also be useful to look in more detail at some 'feedback elements' practised by the teachers during the assessment process. This is because providing feedback to students is an integral part of teaching. It could be something on teacher beliefs about feedback as suggested in TCoF inventory by Brown *et al.* (2012) which modelled nine feedback factors. However, as this inventory was developed in a New Zealand context, it might need further adaptations to suit the Asian context.

6.5.3 Practical Implications

a) Implications for teacher training

Teachers develop their assessment knowledge and skills particularly from their pre-service education and then continue with formal and informal courses and training organised by schools or the ministry. Peer observation activities and interaction between teachers also play a role in this. If it is possible to identify what specific knowledge and skills are needed and also the processes that need to be focused on or type of challenges that are really burdening teachers, then it would be beneficial to devote a significant amount of energy and

financial resources to these particular areas of teacher training. Also hands on training might be more beneficial to them. This is supported by Wiliam (2007b) who pointed out that, in order to gain an effective formative assessment implementation which produces substantial increases in students' achievement, it has to be a sustained professional development focusing minute-by-minute and day-by-day formative assessment.

b) Implications for teacher practices

The implications for teachers of the implementation of SBA are that teachers can no longer rely solely on public examination at the end of a term or year. They can no longer rely on one technique of assessment type, such as tests only, as testing alone does not allow students to demonstrate their full potential (Black and Wiliam, 1998a). Furthermore, testing is an assessment which is weakly linked to the learning experiences of students. While the use of the evaluation instrument in this study might not directly improve teaching practises in the classroom, the data gained from this instrument could help training providers such as the ministry to structure their training programmes to align with teachers' belief system together with their knowledge and skill in assessment. Furthermore, it seems possible that training on SBA might be more beneficial to teachers after the extra burden and workloads are removed from them. To some extent, this instrument could also provide a useful tool for establishing a baseline for teachers' belief system on SBA and their basic knowledge and skill on assessing students as a means of teacher evaluation after they have attended training courses.

c) Implications for schools management

Research strongly suggests that students' learning, achievement and motivation for learning are more achievable when teachers focus more on formative assessment implementation and implement summative assessment only when it is necessary (ARG, 2002). So it is critical for the head teachers to ensure that assessment policy and a positive school climate are clearly communicated to all teachers. Head teachers, senior managers and officers from the ministry need to firmly adhere to the policy instructed by the government to focus more on AfL. The strict establishment on the policy and

school culture in promoting AfL involves educating staff, parents and larger communities with this new innovation. School policies should also acknowledge teachers and administrators regarding suitable time and method to implement summative assessments since they are also an important component of SBA. Although summative assessment has some weaknesses especially in motivating students for learning, it also has its advantages, as has been discussed in Chapter 3. In addition, ARG (2002) suggested that in order to avoid demotivation influenced by summative assessments, teachers are advised to assess their students when they are confident of their students' performance.

The findings of this study indicated that challenges faced by teachers are negatively associated with students' attitude and motivation. By focusing specifically on school climate issues, the findings indicated the importance of creating a positive school climate to improve the outcome. Furthermore, descriptive data also showed that nearly one third of the respondents felt that school climate seems to hinder SBA implementation in their schools. Similarly, one third of the respondents also believed that SBA does not increase public satisfaction and confidence towards school. Therefore, it might be useful for the top management to revisit the physical, social and academic dimensions, as stated in Loukas (2007) and shown in Table 3.18, especially those dimensions related to assessment implementation to ensure that a positive school climate is achieved. In order to deliver this information, not only to teachers, but also to students, parents and wider school communities, an appropriate communication and consultation are needed by the management.

d) Implications for national assessment policies

The findings of this study also suggested that policy strategies have to focus on teacher change and curriculum development in order to carry out assessment reform. As such, policies could focus on incorporating such holistic assessments to teachers in pre-service teacher education and in-service professional development realising the fact that this type of assessment not only improves students' performance in class but also produces a higher-order, collaborative, creative and critical workforce (Koh and Luke, 2009). However, it is important to say that this process is not a one-shot process or an ad-hoc

event in order to achieve a better outcome. It has to be conducted in a systematic manner in its planning, implementation and evaluation.

Teachers and administrators have to be clear on the policies relating to the purpose of formative and summative assessment to avoid confusion. They should understand the limited validity that public examination has to offer in improving student learning. They should also understand the balance between the costs of formative and summative assessment in terms of teaching time or marking. As formative assessment or AfL is now seen as having a good chance to be the most cost-effective means of assessment to the school or system if being implemented wisely (Wiliam, 2009), this newly-implemented form of assessment should be given greater concern compared to summative. In Malaysia, when PT3, a central assessment which is a summative component of SBA for Year 9 students is implemented in 2014, all marking processes are done by teachers at schools. This will cut cost, which is previously incurred by the government whenever public examinations are conducted. This in turn should lead policy makers to understand about the costs and benefits of the implementation of each component of SBA. Moreover, although implementing SBA is cost-effective to the school or system, there is no doubt that providing teachers with continuous professional development is costly (Webster *et al.,* 2012). Furthermore, nearly half of the respondents agreed that the problem with financial resources seemed to be one of the main challenges to implement SBA. This is because most teachers have to use their own internet line, papers or materials to implement SBA.

The results of this survey indicated that the impact on students motivational towards learning are not consistent with the official Malaysian government policy concerning the objectives of the National Education Assessment System (NEAS) on improving students learning. Certainly, based on this survey, we would expect researchers, practitioners or schools to be able to explore more on student performance. As such, access to data on students' performance at schools should be made more transparent so that a better picture could be gained regarding the impact of the system on students. For example, the aims and objectives of SBA which have been detailed by the ministry need to be made accessible to schools, as do the recent developments of SBA in the

country so schools can better prepare themselves to deal with this new innovation.

6.6 Limitations of the Study

While this study clearly shows that the interrelationships between evaluation dimensions did occur, a number of limitations of this study need to be pointed out. The limitations need to be addressed to support the accurate interpretations of the results derived and also to show a need for further examination or additional research. The limitations are:

1. This study is designed to evaluate the implementation of SBA in schools by using only one model of evaluation, which is the Stufflebeam's CIPP Model.
2. This study surveyed only on teachers' perceptions without considering views from the other stakeholders such as students, parents or officers from the ministry. This might limit the development and validation of the instrument in measuring factors. Furthermore, by asking teachers what they think might not be the same as what they actually practise in reality.
3. This study uses survey research design so it is not as robust as an experimental research design. However, it is difficult in establishing causality as SBA is an on-going system, so conditions are very difficult to manipulate.
4. Investigating the evaluation dimension of a newly-introduced system like SBA can be considered as a sensitive issue and teachers might have reported only the positive aspects of it, therefore there is the possibility of bias. The willingness to report positively appears to be normal as some teachers believe the success of the system implementation is dependent on their responses. Such reporting might not give the true picture of the findings. However, several preventive steps had been taken such as explanations were made to the teachers when the researcher went to the schools and this issue was also made clear in the consent letter attached to every questionnaire (See appendix C).

5. The range of items included under the input, process and product dimensions are limited with some items deleted during the CFA and SEM procedures. This is because, few items can be included at any one time due to time and resources limitations during field work. For example, the researcher believes that it is better to include classroom level data such as student grades in weekly or monthly assessment in product dimension to allow stronger inferences on the impact of the SBA to be made. However, at the time of the data collection, SBA was still newly-implemented and some schools still have not had the marks. Moreover, some schools have not keyed in the marks into the system due to problems with server. So, it is not clear from this study whether process implementation are also, or as strongly, related to basic skills achievement gains or long term cognitive goals. It is also possible, and indeed likely, that some processes which were not effective in attaining students' outcomes in this study can show a significant positive relationship if the outcomes are extended to a broader dimension.
6. This study based much if its literature on research on SBA in a Western context due to insufficient research conducted on the topic in Asian countries. In addition, some rural schools in this study context are much poorer compared to those in Western countries. However, as the evaluation scale was developed and validated by the researcher to suit the sample, this might help in influencing the findings of the study.
7. The sample of this study was chosen from school teachers in Malay-medium public schools in one of the state in the east of Peninsular Malaysia, so generalisations cannot be made to all schools in Malaysia. But since all Malaysian school teachers are following the same education system directed from the centre which is the MOE operated in Putrajaya, to some extent the results is expected to be similar to all parts of Malaysia. The researcher believes that to some extent, the centralised education system affects the way teachers think and feel about SBA in a similar way.
8. When the model fits data it shows that our hypothesis model is not rejected but still, there might be variables which are not included in our model which are more influential. There might also be another model which may fit our data as well or better. Only if we can replicate our

models over time and they keep fitting, can we be more confident that the model is accepted.

6.7 Suggestions for Future Research

Future research in this area would of course be beneficial. Thus, a future study would ideally:

1. Bring in other views (policy makers, head teachers, students, parents) in order to provide further insights and perspectives and could be a fruitful attempt to confirm a robust conclusion of the findings. Similarly, comparing perceptions of teachers from different states could also be conducted for a more rigorous result.
2. Future research might want to employ both approaches, quantitative and qualitative, so as to provide more insight and depth on evaluation. For example, the inclusion of qualitative methods, such as interviews, observations or document analysis rather than using questionnaires alone could allow for more themes and sub-constructs to be explored. Case studies can also be used as they are seen as a powerful means in getting a detailed and a wide variety of information on the processes and practices involved and also the outcomes (Muijs *et al.,* 2004).
3. The interrelationships that exist between the dimensions in this study within the Malaysian programme evaluation context suggests that this is an educational phenomenon that merits further investigation within other contexts worldwide.
4. This study is conducted in one country in the north-east of Peninsular Malaysia with most of them are Malay population. It seems appropriate for samples to include respondents from various type of schools to represent various backgrounds of respondents. Furthermore, rural schools involved in this study do not include those at the remote areas due to inaccessible factor and all the schools selected excluded schools with high-achieving students, like cluster schools or boarding schools due to administration factor limitation.

6.8 Conclusion

This study presented a detailed investigation on the evaluation of the SBA system implementation among teachers in Malaysia. The investigation seems encouraging, when it is considered that SBA is newly-implemented in Malaysia. Other Asian countries like Hong Kong, Singapore, China or Bangladesh have started to implement formative assessments much earlier. For example, Hong Kong has started to discover about formative assessment since 2001 (Kennedy, 2013).

This study identified the perception of teachers towards the input, process and product dimension of evaluation. The findings confirmed that each dimension of SBA evaluation – input, process and product did show a positive result. The findings confirmed the results from previous studies whereby the input dimension (Othman *et al.,* 2013; Adnan *et al.,* 2009), process dimension (Majid, 2011) and product dimension (Mansor *et al.,* 2013) of SBA evaluation have showed positive findings. This contributed to the literature of SBA implementation in schools especially in the non-western context. Assessment of the measurement models of input, process and product produced a valid and reliable instrument, so evaluation could be conducted in later research using these models. But even so, a review has to be made before any study is conducted to suit the different samples and study context. Furthermore, this study also investigated the interrelationships between dimensions of evaluation. Although the SBA system has been researched before, most previous research has focused on certain dimensions of the evaluation only and did not determine the influence of one dimension on another. This study investigated the relationship between input and process dimension. Input was found to be related to challenges, 'attitudes and skills' and 'moderation process' implemented by teachers. And, all the relationships were significant and strong. Next, it was found that some processes were related to product. 'Challenges' and 'attitudes and skills' were related to product whereas 'moderation process' was not significantly related to product. This has paved the way for further investigation on teachers understanding on this process and how it is linked to the goals of SBA. Furthermore, input was found not to be directly related to product, but still was indirectly related to it. This too, has paved the way for further investigation on the input constructs especially in the

Asian context. This study also revealed that secondary schools moderated most of the significant relationships between dimensions. In other words, the relationship between all the significant paths was stronger in secondary schools compared to primary schools. This was the case, although at the time data were collected by the researcher, primary schools have had two years of experience in implementing SBA whereas secondary schools have had only one year experience.

In conclusion, by looking at the dimensions being evaluated and the testing of all the hypotheses related to SBA implementation to a non-western country, this study has provided a good picture to the overall process of evaluation. This study suggests that in the Malaysian context at least, SBA is difficult to implement effectively but not impossible. Schools need to be geared up to have a clear understanding about the standards that should be applied to SBA in order to win the hearts and minds of teachers, students and the community as a whole. Furthermore, this study also provides a good literature for other countries, especially the developing countries. The development of a valid and reliable evaluation scale is hoped to contribute to the body of knowledge of SBA evaluation, hence people could use it in evaluation. Not only contributing additional knowledge to the literature of the implementation of SBA system in schools, the findings are expected to help policy-makers in planning professional development courses or providing more resources in a more effective way. In addition, knowledge on the interrelationships between dimensions of evaluation and the moderation effects that act upon the relationships could also be a valuable knowledge in assisting researchers with their future research.

Appendix A: SBA Evaluation Questionnaire (English Version)

Please answer **ALL** questions honestly. There is no right or wrong answer. Your responses will be strictly confidential and will be used only for academic purposes.

SECTION A:

Using the following 1-5 scale, please indicate by circling the most correct response, the degree to which you agree with the statements listed below:

1	2	3	4	5
Strongly disagree	**disagree**	**neutral**	**agree**	**strongly agree**

Using school-based assessment (SBA) in teaching:

1	Is efficient in improving students learning	1	2	3	4	5
2	Can help students to become more independent learners	1	2	3	4	5
3	It is good that students have more than one opportunity to be assessed	1	2	3	4	5
4	SBA can be effectively incorporated into existing lessons	1	2	3	4	5
5	Is not burdening me	1	2	3	4	5
6	Is not boring	1	2	3	4	5
7	Is compulsory	1	2	3	4	5
8	I prepare students' individual files for my subject before conducting SBA	1	2	3	4	5
9	I have plenty of opportunities to discuss SBA implementation with my colleagues who are teaching the same subject with me	1	2	3	4	5
10	I always follow the planning instructed by the MOE in SBA implementation	1	2	3	4	5
11	Educational assessment transformation through SBA leads to changes from examination-oriented assessment to an assessment which is more integrated	1	2	3	4	5
12	SBA is conducted during the teaching and learning process	1	2	3	4	5
13	For centre assessment in SBA, assignments are prepared by the Examination Board	1	2	3	4	5
14	It is easy to implement assessment if teachers are supplied with a complete assessment document	1	2	3	4	5
15	A teaching assisstant is needed to help teachers in assessment	1	2	3	4	5
16	Training of the personnel involved should be properly planned and implemented	1	2	3	4	5
17	Teachers are sufficiently qualified to implement assessment activities	1	2	3	4	5
18	Teachers' skills in assessment are adequate for implementing assessment activities	1	2	3	4	5
19	The physical equipment in classroom is suitable to conduct SBA activities	1	2	3	4	5
20	The space can be adapted to assessment activities	1	2	3	4	5
21	The ICT hardware is suitable to conduct SBA activities	1	2	3	4	5

SECTION B:

22. Have you ever attend any courses/ any training related to SBA system?
    YES ☐ NO ☐

If you have answered "NO" to question 22 please go to question 26, otherwise continue with question 23.

23. Information on SBA courses that you have attended:

No	Name of Course/ Organizer	No of days	Year

1 2 3 4 5

Strongly disagree **disagree** **neutral** **agree** **strongly agree**

24. The courses on SBA system have served me in upgrading my skills on SBA on aspects such as how:

i.	to assess students using the assessment instrument	1	2	3	4	5
ii.	to assess students documentation following stated criteria in the assessment document	1	2	3	4	5
iii	to explain assessment criteria in details to students	1	2	3	4	5
iv	to conduct assessment activities effectively	1	2	3	4	5

25. Please evaluate the following questions relate to the In-house training (IHT) on SBA at your school.

i.	IHT is conducted by the experts from the ministry or the panels of teachers	1	2	3	4	5
ii.	IHT has been conducted more than once to increase teachers' understanding	1	2	3	4	5
iii	Training includes practical forms not only theoretical	1	2	3	4	5

26. Please give your views on administration in implementing SBA at your school. Administrators..

i.	Provide reinforcement such as giving gifts or praise to those teachers that practice SBA in their classrooms	1	2	3	4	5
ii.	Give special recognition to teachers performance for conducting SBA	1	2	3	4	5

27. Please give your views on the coordination process run at your school. Each teacher

i.	prepares students' scores for his/her subjects	1	2	3	4	5
ii.	prepares evidence according to students' achievement	1	2	3	4	5
iii	records the marks of evidence in SPPBS application	1	2	3	4	5

28. Please give your views on the monitoring process run at your school.

i.	The school SBA committee appoints certain teachers as internal monitors	1	2	3	4	5
ii.	Heads of panel of each subject monitor SBA activity implementation	1	2	3	4	5
iii	Evidence of centre assessment is kept in a systematic way in a safe place before it is returned to the students	1	2	3	4	5

SECTION C:

29. Here are some possible challenges faced by the teachers in implementing SBA in schools. Please indicate

your level of agreement for each item by circling the chosen number:

1	Difficult to gain support from the headteacher	1	2	3	4	5
2	Difficult to gain cooperation from teachers	1	2	3	4	5
3	Insufficient knowledge to implement SBA process	1	2	3	4	5
4	Lack of confidence to conduct SBA	1	2	3	4	5
5	Problem to get related resources on SBA	1	2	3	4	5
6	School climate seems to hinder SBA implementation	1	2	3	4	5
7	Insufficient financial resources	1	2	3	4	5

8	Extra workload	1	2	3	4	5
9	No special recognition for teachers performance for conducting SBA	1	2	3	4	5
10	Students and parents may not trust teachers' assessment in SBA	1	2	3	4	5
11	Students may memorize responses and reproduce them during SBA sessions	1	2	3	4	5
12	The existing number of teachers is not sufficient to implement SBA activities	1	2	3	4	5

SECTION D:

Please indicate your level of agreement for each item by circling the chosen number.

1 2 3 4 5 .

Strongly disagree disagree neutral agree strongly agree

30. In my school, the implementation of SBA

i.	Reduces absentee rate	1	2	3	4	5
ii.	It contributes strongly to the quality assurance process	1	2	3	4	5
iii.	Increases public satisfaction and confidence towards school	1	2	3	4	5
iv.	Provides new learning opportunities for building on students' strengths and interests	1	2	3	4	5

31. The implementation of SBA is important in that it may

i.	Help students improve generic skills to build on their strengths and interests	1	2	3	4	5
ii.	Help teachers to improve students assessment practices	1	2	3	4	5
iii.	Help teachers to actively develop their professionalism	1	2	3	4	5
iv.	Help teachers to develop students various learning opportunities	1	2	3	4	5
v.	Make teachers and students willing to discuss learning problems in non-threatening manners	1	2	3	4	5
vi.	Make teachers willing to take decisions on assessment reform	1	2	3	4	5
vii.	Reduces pressure in public examinations	1	2	3	4	5
viii.	Improves literacy and numeracy learning	1	2	3	4	5

SECTION E:

Circle the number that is reflecting the extent to which the students have mastered the given statement. Have in your mind that:

1 2 3 4 5 .

Strongly disagree disagree neutral agree strongly agree

32. Students' attitude towards SBA:

i.	Students practice SBA in their study	1	2	3	4	5
ii.	The students are motivated to prepare for SBA.	1	2	3	4	5
Iii	Practising SBA in their study is NOT burdening them at all	1	2	3	4	5

33. Students' knowledge of SBA. Students know that:

i.	SBA tasks are part of teaching and learning of their respective subjects	1	2	3	4	5
ii.	Feedback from teachers helps them develop skills that may not be reflected in public examinations	1	2	3	4	5

34. Students' motivational source:

i.	SBA encourages students to read more books than before.	1	2	3	4	5
ii.	The students are becoming more interested in my subject than before	1	2	3	4	5
Iii	SBA helps students to understand more on their strengths and weakness in	1	2	3	4	5

	each subject.					

SECTION F: DEMOGRAPHIC INFORMATION

Instruction: Please put a *(√)* in the box for each question and answer the questions.

1. School Grade: i. Grade A ☐ ii. Grade B ☐
2. School Location: i. Urban ☐ ii. Rural ☐
3. School Category: i. Secondary School ☐ ii. Primary School ☐
4. Gender: i. Female ☐ ii. Male ☐
5. Race: i. Malay ☐ ii. Chinese ☐
   iii. Indian ☐ iv. Others ☐
6. Age: ☐☐ years
7. Teaching experience: ☐☐ years
8. Year/ Form which teacher is teaching: _____
9. Teacher's experience in practicing SBA: ☐☐ years
10. Attending any courses/training on SBA: ☐☐ times
11. Do you teach your own option? Yes ☐ No ☐
12. The main subjects taught: _____
13. Highest academic achievement:
    i. Diploma ☐ ii. Degree ☐
    iii. Master ☐ iv. PhD ☐

THANK YOU

Appendix B: SBA Evaluation Questionnaire (Malay Version)

SOAL SELIDIK GURU

Lot 260, Kg Pasir Hor,
Kota Bharu, 15150
Kelantan Darulnaim,
Malaysia
Tel: 09-7653980/+447438845908
Email:nidacmg@yahoo.com/ncmg1g11@soton.ac.uk

Saudara/ Saudari,

SOALSELIDIK BERKENAAN PELAKSANAAN SISTEM PENTAKSIRAN BERASASKAN SEKOLAH (PBS) DI KALANGAN GURU

Sukacita dimaklumkan bahawa satu soalselidik akan dilaksanakan kepada guru-guru sekolah rendah dan menengah. Tujuan soalselidik ini adalah untuk menilai pelaksanaan sistem PBS di sekolah-sekolah di Kelantan Darulnaim.

Sebagai guru yang mengajar di sekolah kerajaan di Malaysia, saudara/ saudari sudah pasti sedikit sebanyak pernah melaksanakan PBS di sekolah. Oleh itu, maklumbalas saudara/saudari amat penting dalam membantu usaha memahami kekuatan dan kelemahan pelaksanaan sistem. Soalselidik ini adalah semata-mata untuk tujuan akademik dan semua maklumat yang diberikan akan dirahsiakan. Saudara/ saudari hanya memerlukan lebih kurang 20 minit sahaja untuk menyiapkan soalselidik ini.

Saya amat menghargai pertolongan yang diberikan dalam menjayakan kajian ini. Terima kasih sekali lagi atas bantuan yang jujur dan ikhlas daripada saudara/ saudari.

Nor Hasnida Che Md Ghazali
University of Southampton

BAHAGIAN A:

Sila berikan pandangan anda tentang sejauhmana anda bersetuju dengan pernyataan di bawah dengan membulatkan nombor menggunakan skala 1-5:

1	2	3	4	5
Sangat tidak setuju	**Tidak setuju**	**neutral**	**setuju**	**sangat setuju**

Melaksanakan pentaksiran berasaskan sekolah (PBS) di dalam pengajaran:

1	Adalah berkesan dalam meningkatkan pembelajaran pelajar	1	2	3	4	5
2	Boleh membantu pelajar untuk menjadi lebih berdikari	1	2	3	4	5
3	Ianya bagus kerana pelajar mempunyai lebih dari satu peluang untuk ditaksir	1	2	3	4	5
4	PBS boleh digabungkan secara efektif ke dalam pengajaran yang sedia ada	1	2	3	4	5
5	Tidak membebankan saya	1	2	3	4	5
6	Tidak membosankan suasana pembelajaran	1	2	3	4	5
7	Ianya diwajibkan di dalam pengajaran	1	2	3	4	5
8	Saya menyediakan fail individu pelajar bagi mata pelajaran saya sebelum menjalankan PBS	1	2	3	4	5
9	Saya mempunyai banyak peluang untuk membincangkan pelaksanaan PBS dengan rakan-rakan yang mengajar subjek yang sama dengan saya	1	2	3	4	5
10	Saya sentiasa mengikuti perancangan dari pihak Kementerian Pelajaran dalam melaksanakan PBS	1	2	3	4	5
11	Transformasi pentaksiran pendidikan melalui PBS membawa kepada perubahan dari pentaksiran berorientasikan peperiksaan kepada pentaksiran yang lebih bersepadu	1	2	3	4	5
12	PBS dijalankan semasa proses pengajaran dan pembelajaran	1	2	3	4	5
13	Bagi Pentaksiran Pusat, tugasan disediakan oleh pihak Lembaga Peperiksaan	1	2	3	4	5
14	Adalah mudah untuk melaksanakan pentaksiran jika guru dibekalkan dengan dokumen pentaksiran yang lengkap	1	2	3	4	5
15	Pembantu guru diperlukan untuk membantu guru dalam pentaksiran	1	2	3	4	5
16	Latihan kakitangan yang terlibat dengan PBS haruslah dirancang dan dilaksanakan dengan teliti	1	2	3	4	5
17	Guru benar-benar layak untuk melaksanakan aktiviti pentaksiran	1	2	3	4	5
18	Kemahiran guru dalam pentaksiran adalah mencukupi untuk melaksanakan aktiviti pentaksiran	1	2	3	4	5
19	Peralatan fizikal dalam bilik darjah adalah sesuai untuk menjalankan aktiviti PBS	1	2	3	4	5
20	Ruangan dalam bilik darjah boleh disesuaikan dengan aktiviti pentaksiran	1	2	3	4	5
21	Perkakasan ICT adalah sesuai untuk menjalankan aktiviti PBS	1	2	3	4	5

BAHAGIAN B:

22. Adakah anda pernah menghadiri sebarang kursus/ perbincangan dengan ketua panitia berkaitan PBS?

PERNAH ☐ TIDAK PERNAH ☐

Jika anda telah menjawab "TIDAK PERNAH" untuk soalan 22, sila jawab soalan 26 dan seterusnya, sebaliknya teruskan menjawab soalan 23.

23. Maklumat berkaitan PBS yang pernah dihadiri (jika ada)

Bil	Nama Kursus/ Penganjur	Bilangan Hari Kursus	Tahun

1	2	3	4	5	.
Sangat tidak setuju	**Tidak setuju**	**neutral**	**setuju**	**sangat setuju**	

24. Kursus/perbincangan berkaitan PBS telah memberi peluang kepada saya meningkatkan kemahiran dalam menjalankan PBS terutamanya dalam aspek:

i.	untuk mentaksir pelajar menggunakan instrumen pentaksiran	1	2	3	4	5
ii.	untuk menilai dokumentasi pelajar mengikut kriteria yang dinyatakan dalam dokumen pentaksiran	1	2	3	4	5
iii	memperjelaskan kriteria pentaksiran secara terperinci kepada pelajar	1	2	3	4	5
iv	untuk menjalankan aktiviti pentaksiran dengan berkesan	1	2	3	4	5

25. Sila nilai latihan/ perbincangan berhubung PBS (*in-house training*) yang dijalankan di sekolah anda.

i.	Latihan dijalankan oleh pakar dari pihak kementerian/ guru panitia sekolah	1	2	3	4	5
ii.	Latihan telah dijalankan lebih daripada sekali untuk meningkatkan pemahaman guru	1	2	3	4	5
iii	Latihan bukan hanya merangkumi aspek teori sahaja tetapi juga aspek praktikal	1	2	3	4	5

26. Sila berikan pandangan terhadap pihak pentadbiran dalam melaksanakan PBS di sekolah. Pentadbir:

i.	Memberi pengukuhan seperti memberi hadiah atau pujian kepada guru yang menjalankan PBS di dalam pengajaran	1	2	3	4	5
ii.	Memberi pengiktirafan khas terhadap prestasi guru yang menjalankan PBS	1	2	3	4	5

27. Sila berikan pandangan terhadap proses penyelarasan di sekolah anda. Setiap guru seharusnya:

i.	Menyediakan skor pelajar untuk subjek mereka	1	2	3	4	5
ii.	Menyediakan evidens mengikut pencapaian pelajar	1	2	3	4	5
iii	Merekod markah evidens di dalam aplikasi SPPBS	1	2	3	4	5

28. Sila berikan pandangan terhadap proses pemantauan di sekolah anda.

i.	Jawatankuasa PBS sekolah melantik guru-guru tertentu sebagai pemantau dalaman	1	2	3	4	5
ii	Ketua panitia bagi setiap subjek memantau pelaksanaan aktiviti PBS	1	2	3	4	5
iii	Eviden pentaksiran pusat disimpan dalam cara yang sistematik di tempat yang selamat sebelum ianya dipulangkan kepada pelajar	1	2	3	4	5

BAHAGIAN C:

29. Berikut disenaraikan kekangan-kekangan yang dihadapi oleh guru dalam menjalankan PBS. Sila berikan pandangan anda dengan membulatkan jawapan.

1	Sukar untuk mendapat sokongan daripada guru besar	1	2	3	4	5
2	Sukar untuk mendapat kerjasama daripada guru-guru	1	2	3	4	5
3	Pengetahuan yang tidak mencukupi dalam melaksanakan proses PBS	1	2	3	4	5
4	Kurang keyakinan dalam menjalankan PBS	1	2	3	4	5
5	Masalah dalam mendapatkan sumber berkaitan dengan PBS	1	2	3	4	5
6	Iklim sekolah seolah-olah menghalang pelaksanaan PBS	1	2	3	4	5
7	Sumber kewangan yang tidak mencukupi	1	2	3	4	5
8	Beban kerja tambahan	1	2	3	4	5
9	Tiada penghargaan khas terhadap prestasi guru yang menjalankan PBS	1	2	3	4	5
10	Pelajar dan ibu bapa tidak percaya dengan pentaksiran guru dalam PBS	1	2	3	4	5
11	Pelajar menghafal jawapan dan memuntahkannya semula semasa sesi PBS	1	2	3	4	5
12	Bilangan guru-guru sedia ada tidak mencukupi untuk melaksanakan aktiviti	1	2	3	4	5

	PBS					

1 2 3 4 5 .

Sangat tidak setuju **Tidak setuju** **neutral** **setuju** **sangat setuju**

BAHAGIAN D:

Sila nyatakan tahap persetujuan anda untuk setiap item dengan membulatkan nombor yang dipilih.

30. Di sekolah saya, pelaksanaan PBS

i	Mengurangkan kadar ketidakhadiran	1	2	3	4	5
ii	Sangat memberi sumbangan kepada proses penjaminan kualiti	1	2	3	4	5
iii	Meningkatkan kepuasan dan keyakinan orang ramai terhadap sekolah	1	2	3	4	5
iv	Menyediakan peluang pembelajaran yang baru untuk membina minat dan kekuatan pelajar	1	2	3	4	5

31. Pelaksanaan PBS adalah penting kerana ianya boleh:

i	Membantu pelajar meningkatkan kemahiran generik dalam membina kekuatan dan minat mereka	1	2	3	4	5
ii	Membantu guru dalam meningkatkan amalan pentaksiran pelajar	1	2	3	4	5
iii	Membantu guru dalam mengembangkan profesionalisme mereka	1	2	3	4	5
iv	Membantu guru membina pelbagai peluang pembelajaran kepada pelajar	1	2	3	4	5
v	Menjadikan guru dan pelajar bersedia untuk membincangkan masalah pembelajaran dalam suasana yang tidak tertekan	1	2	3	4	5
vi	Menjadikan guru bersedia untuk membuat keputusan berkaitan reformasi dalam pentaksiran	1	2	3	4	5
vii	Mengurangkan tekanan terhadap peperiksaan awam	1	2	3	4	5
viii	Meningkatkan pembelajaran literasi dan numerasi	1	2	3	4	5

BAHAGIAN E:

Bulatkan nombor yang menggambarkan sejauhmana pelajar telah menguasai pernyataan yang diberikan.

1 2 3 4 5 .

Sangat tidak setuju **Tidak setuju** **neutral** **setuju** **sangat setuju**

32. Sikap pelajar terhadap PBS:

i.	Pelajar mempraktikkan PBS dalam pembelajaran mereka	1	2	3	4	5
ii.	Pelajar terdorong untuk mempersiapkan diri kearah sistem PBS	1	2	3	4	5
iii	Menjalankan PBS tidak membebankan pelajar	1	2	3	4	5

33. Pengetahuan pelajar tentang PBS. Pelajar tahu:

i.	Tugasan PBS adalah sebahagian daripada pengajaran dan pembelajaran subjek masing-masing	1	2	3	4	5
ii.	Maklumbalas daripada guru membantu pelajar meningkatkan kemahiran yang tidak mungkin diperolehi daripada peperiksaan awam	1	2	3	4	5

34. Sumber motivasi pelajar:

i.	PBS memotivasikan pelajar membaca lebih banyak buku berbanding sebelumnya	1	2	3	4	5
ii.	Pelajar lebih berminat terhadap subjek saya berbanding sebelum ini	1	2	3	4	5
iii	PBS membantu pelajar lebih memahami kekuatan dan kelemahan mereka dalam subjek yang dipelajari	1	2	3	4	5

BAHAGIAN F: MAKLUMAT DEMOGRAFI

Sila tandakan *(√)* dalam kotak yang berkenaan atau isikan dalam ruangan yang disediakan.

1. Gred Sekolah: i. Gred A ☐ ii. Gred B ☐
2. Lokasi Sekolah: i. Bandar ☐ ii. Luar Bandar ☐
3. Kategori Sekolah: i. Sekolah Menengah ☐ ii. Sekolah Rendah ☐
4. Jantina: i. Perempuan ☐ ii. Lelaki ☐
5. Bangsa: i. Melayu ☐ ii. Cina ☐
   iii. Indian ☐ iv. Lain-lain ☐
6. Umur: ☐☐ tahun
7. Pengalaman mengajar: ☐☐ tahun
8. Pengalaman dalam melaksanakan PBS: ☐☐ tahun
9. Menghadiri kursus/latihan berkenaan PBS: ☐☐ kali
10. Adakah anda mengajar option sendiri? Ya ☐ Tidak ☐
11. Mata pelajaran utama yang diajar di sekolah: _____
12. Pencapaian akademik tertinggi:
    ii. Diploma ☐ ii. Sarjanamuda ☐
    iii. Sarjana/Master ☐ iv. PhD ☐

TERIMA KASIH!!!

Appendix C: Consent Form (English Version)

CONSENT FORM

Title: School-Based Assessment System implementation Among teachers: An Evaluation

The following information is provided to help you decide whether you wish to participate in the present study. You should be aware that you are free to decide not to participate or to withdraw at any time without affecting your relationship with this department, the instructor, or the University.

The purpose of this study is to evaluate School-Based Assessment system implementation in primary and secondary schools. Data will be collected using survey given to primary and secondary school teachers. Do not hesitate to ask questions about the study before participating or during the study. I would be happy to share the findings with you after the research is completed. Your name will not be associated with the research findings in any way, and only the researcher will know your identity.

There are no known risks or discomforts associated with this study. The expected benefits associated with your participation are the information about the experiences in implementing SBA system.

Please sign this consent form. You are signing it with full knowledge of the nature and purpose of the procedures. A copy of this form will be given to you to keep.

_____ _____

Name and Signature of Participant Date

Appendix D: Consent Form (Malay Version)

BORANG PERSETUJUAN

Tajuk: Penilaian Perlaksanaan Sistem Pentaksiran Berasaskan Sekolah Di Malaysia

Maklumat di bawah adalah untuk menentukan samada pihak guru ingin mengambil bahagian dalam kajian di atas atau tidak. Pihak guru adalah bebas samada ingin mengambil bahagian atau tidak, atau boleh sahaja menarik diri pada bila-bila masa dan ia tidak mempengaruhi hubungan pihak guru dengan pihak pentadbiran sekolah.

Penyelidikan ini dijalankan untuk menilai sistem pentaksiran berasaskan sekolah (PBS) yang dijalankan di sekolah rendah dan menengah. Data akan dikumpul menggunakan soal selidik. Pihak guru boleh bertanya apa sahaja soalan pada pengkaji atau guru yang telah diamanahkan. Saya yang menjalankan kajian ini sangat berbesar hati untuk berkongsi segala dapatan kajian bila penyelidikan ini telah selesai. Nama guru tidak akan dipamirkan di mana-mana penerbitan kajian. Tidak terdapat sebarang risiko atau kitidakselesaan berkaitan penyelidikan ini. Manfaat yang diharapkan dapat dijana nanti ialah berkenaan maklumat berkaitan pengalaman pihak guru yang diberikan semasa mengisi soalselidik berkenaan perlaksanaan PBS di sekolah.

Sila tandatangan borang persetujuan ini menandakan tanda setuju. Tandatangan ini diturunkan dengan pemahaman yang lengkap berkenaan objektif kajian. Satu salinan borang ini akan diberikan kepada pihak guru.

……..

……..

(Nama dan tandatangan guru) Tarikh: ……

Appendix E: A Letter to Headteacher (English Version)

Nor Hasnida Che Md Ghazali
Lot 260, Kg Pasir Hor,
Kota Bharu, 15100
Kelantan,
Malaysia.

Headteacher,
Sekolah Rendah dan Menengah,
Kelantan Darul Naim,
Malaysia. 8 December 2012

Dear Sir/ Madam,

Applying For Data Collection In Primary and Secondary Schools

The above is referenced. I am Nor Hasnida Che Md Ghazali (IC no: 710726035488) staff no: K02195, a lecturer from UPSI Tanjong Malim, Perak. I am a PhD candidate from University of Southampton, United Kingdom seeking to implement data collection for my study at your place.

Title of research:	Implementation of School-Based Assessment in Malaysia
Place of data-collection:	Primary and secondary schools in Kelantan

Attached are the questionnaires which have to be filled by 50 teachers in your school. I hope that you can randomly distribute the questionnaire to the teachers especially those involved directly with 'Pentaksiran Berasaskan Sekolah'. Once finished, I hope that the school could post them to my house in Kelantan as the above address. Or, just keep it and I will go and take it in February when I go back to Malaysia.

Thank you.

Sincerely,

NOR HASNIDA CHE MD GHAZALI
UNIVERSITY OF SOUTHAMPTON

Appendix F: A Letter to Headteacher (Malay Version)

Nor Hasnida Che Md Ghazali
15, Iris Road,
Bassett, SO16 3GU,
Southampton,
United Kingdom.

Guru Besar,
Sekolah Rendah dan Menengah,
Kelantan Darul Naim,
Malaysia. 8 Disember 2012

YBhg Tuan/Puan,

Memohon Melaksanakan Kutipan Data di Sekolah Tuan/Puan Guru Besar

Perkara di atas adalah dirujuk. Saya Nor Hasnida Che Md Ghazali (no KP: 710726035488) no staf: K02195, pensyarah UPSI Tanjong Malim, Perak adalah merupakan pelajar PhD di University of Southampton, Southampton, United Kingdom memohon untuk melaksanakan kutipan data untuk pengajian saya seperti butiran berikut:

Tajuk Kajian: Penilaian Pelaksanaan Pentaksiran Berasaskan Sekolah (PBS) di Malaysia
Tempat: Sekolah menengah dan rendah sekitar Kelantan (20 sekolah menengah dan 20 sekolah rendah dengan 1000 orang guru sebagai responden)

Bersama ini saya lampirkan soal selidik untuk diisi oleh 50 orang guru yang terlibat dalam perlaksanaan PBS di sekolah tuan/puan. Dengan rendah hati, saya memohon tuan/puan mengedarkan soalselidik ini secara rawak (*random*) kepada guru-guru, sebaiknya kepada mereka yang terlibat secara langsung dengan perlaksanaan PBS. Setelah selesai diisi, diharap kerjasama baik tuan/puan untuk mengirimkan semula ke alamat rumah saya di Pasir Hor seperti alamat di bawah atau saya pergi mengambilnya di awal February bila saya pulang ke Malaysia nanti.

Kerjasama tuan/puan amatlah dihargai.

Sekian, terima kasih.

Yang Benar,

NOR HASNIDA CHE MD GHAZALI
(Lot 260, Kg Pasir Hor, 15100, Kota Bharu, Kelantan).

Appendix G: Cover letter (English version)

Nor Hasnida Che Md Ghazali
Lot 260, Kg Pasir Hor,
Kota Bharu, 15150
Kelantan,
Malaysia
Tel: 09-7653980/+447438845908
Email:nidacmg@yahoo.com/ncmg1g11@soton.ac.uk

Dear teachers,

A SURVEY ON SCHOOL-BASED ASSESSMENT SYSTEM IMPLEMENTATION IN SCHOOL

I am pleased to inform you that a survey will be conducted to the primary and secondary school teachers. The purpose of this survey is to evaluate the implementation of school-based assessment (SBA) system in schools in Kelantan.

As a teacher in government schools, I am sure to some extent, you have been involved in SBA system implementation in your school. Therefore, your feedback is very important in understanding the strength and weakness of the implementation of the system.

This survey is for academics purpose only and all the information given will be kept confidential. You will need about 20 minutes to complete the questionnaire.

I am most appreciative of your help in this research. Thank you for your sincere co-operation.

Nor Hasnida Che Md Ghazali
PhD Candidate
15, Iris Road, Bassett,
Southampton, SO163GU

Appendix H: Cover letter (Malay version)

Lot 260, Kg Pasir Hor,
Kota Bharu, 15150
Kelantan Darulnaim,
Malaysia
Tel: 09-7653980/+447438845908
Email:nidacmg@yahoo.com/ncmg1g11@soton.ac.uk

Saudara/ Saudari,

SOALSELIDIK BERKENAAN PELAKSANAAN SISTEM PENTAKSIRAN BERASASKAN SEKOLAH (PBS) DI KALANGAN GURU

Sukacita dimaklumkan bahawa satu soalselidik akan dilaksanakan kepada guru-guru sekolah rendah dan menengah. Tujuan soalselidik ini adalah untuk menilai pelaksanaan sistem PBS di sekolah-sekolah di Kelantan Darulnaim. Sebagai guru yang mengajar di sekolah kerajaan di Malaysia, saudara/ saudari sudah pasti sedikit sebanyak pernah melaksanakan PBS di sekolah. Oleh itu, maklumbalas saudara/saudari amat penting dalam membantu usaha memahami kekuatan dan kelemahan pelaksanaan sistem.

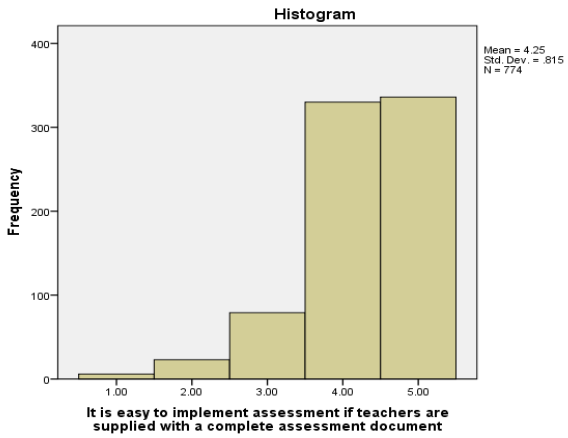
Soalselidik ini adalah semata-mata untuk tujuan akademik dan semua maklumat yang diberikan akan dirahsiakan. Saudara/ saudari hanya memerlukan lebih kurang 20 minit sahaja untuk menyiapkan soalselidik ini. Saya amat menghargai pertolongan yang diberikan dalam menjayakan kajian ini. Terima kasih sekali lagi atas bantuan yang jujur dan ikhlas daripada saudara/ saudari.

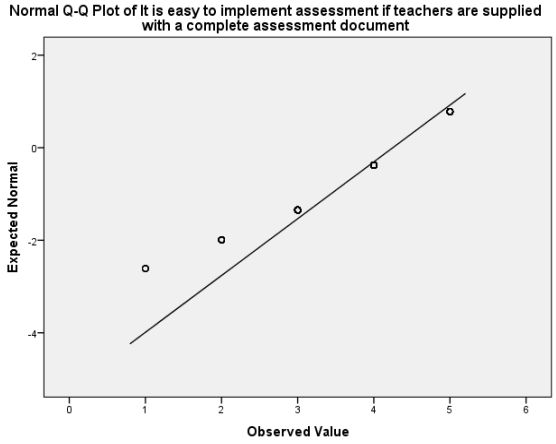
Nor Hasnida Che Md Ghazali
PhD Candidate
15, Iris Road, Bassett,
Southampton, SO163GU

Appendix I: The Frequency Histograms and the Normality Probability Plots

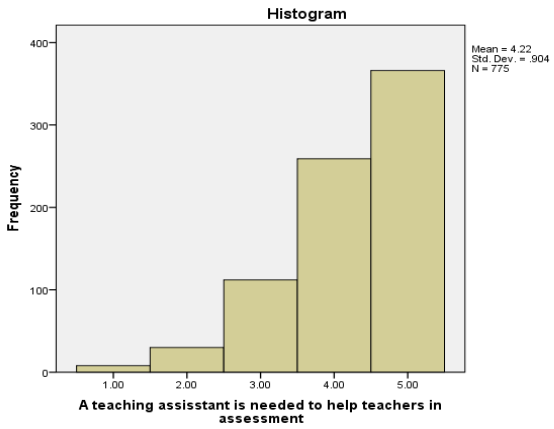
For input evaluation items:

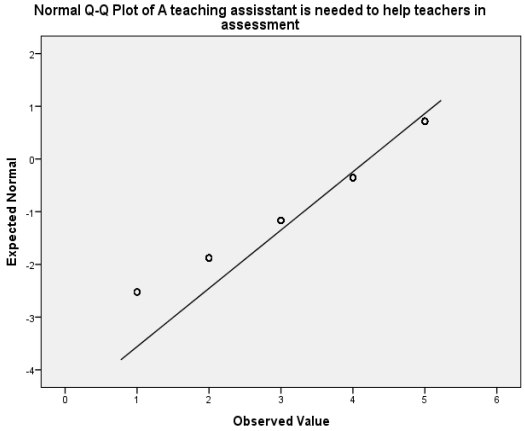
For item a14: It is easy to implement assessment if teachers are supplied with a complete assessment document



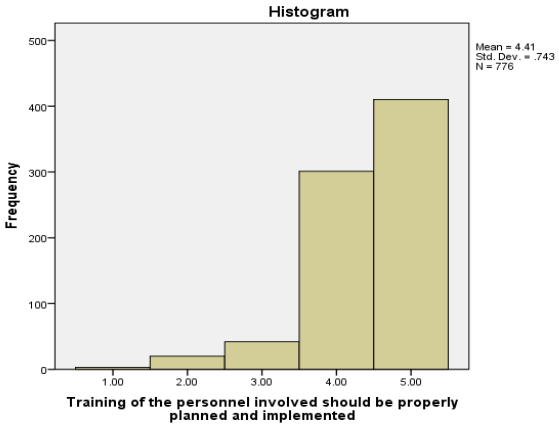


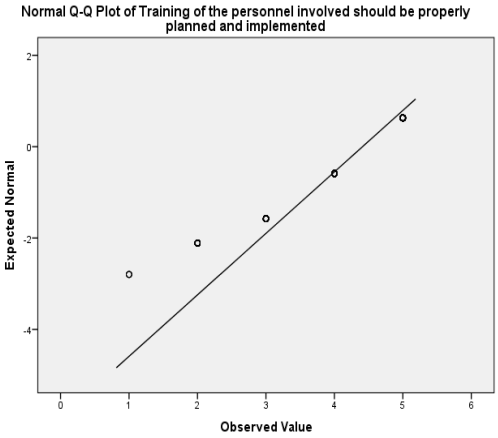
For item a15: A teaching assisstant is needed to help teachers in assessment



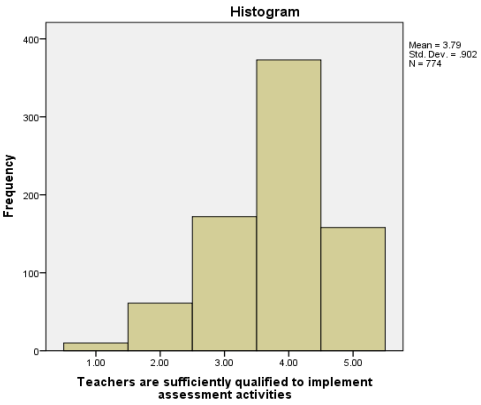


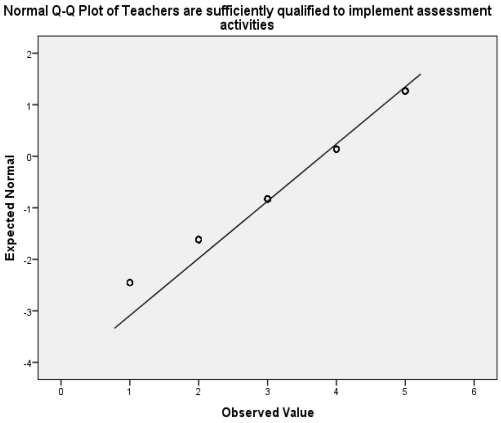
For item a16: Training of the personnel involved should be properly planned and implemented



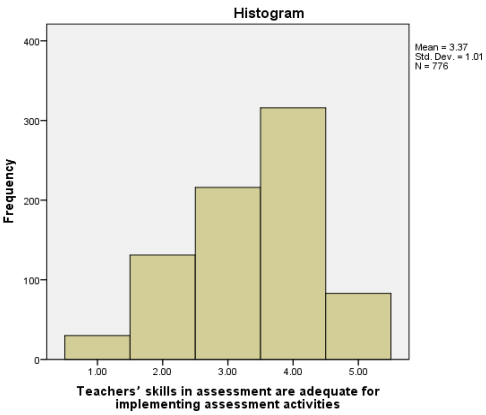


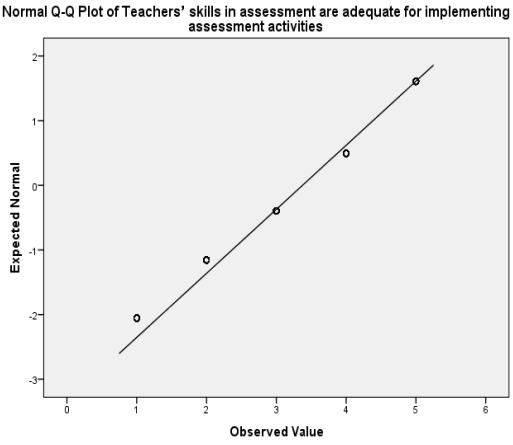
For item a17: Teachers are sufficiently qualified to implement assessment activities



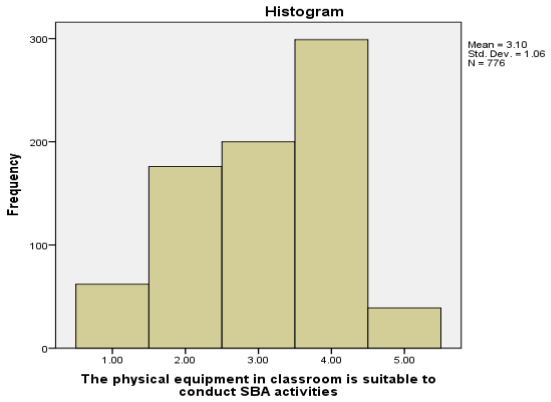


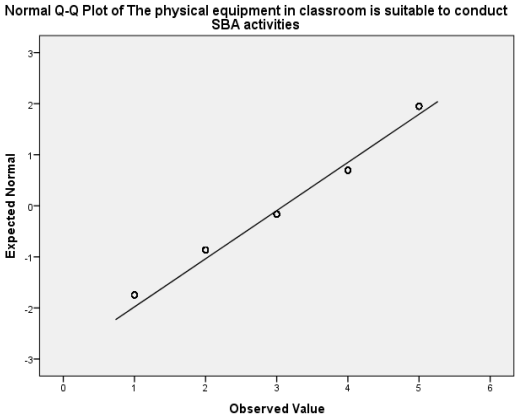
For item a18: Teachers' skills in assessment are adequate for implementing assessment activities



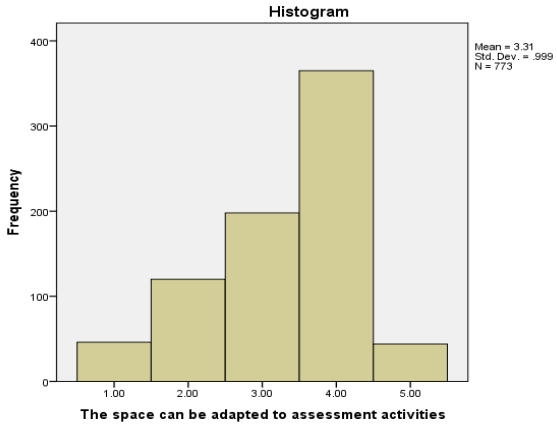


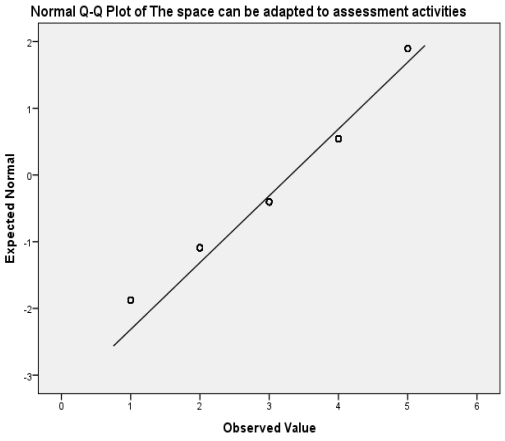
For item a19: The physical equipment in classroom is suitable to conduct SBA activities



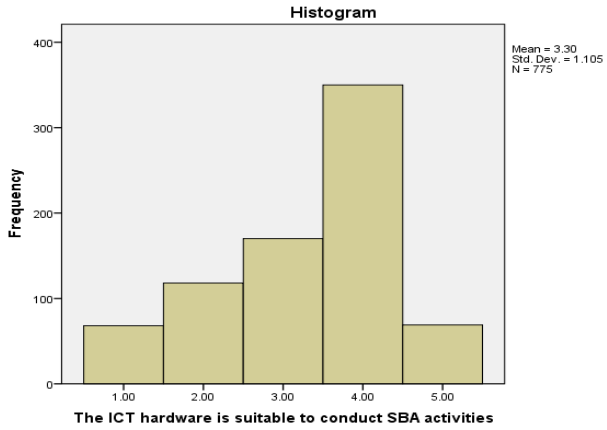


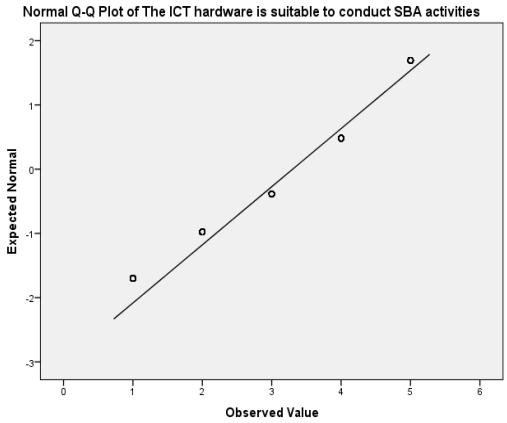
For item a20: The space can be adapted to assessment activities





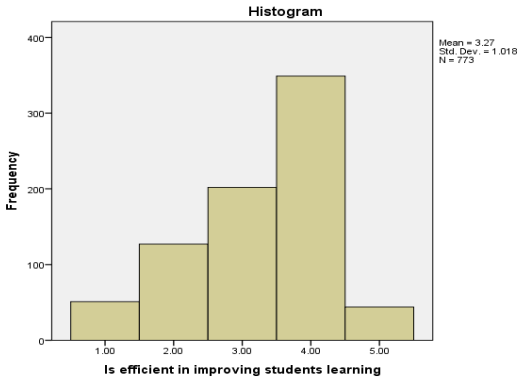
For item a21: The ICT hardware is suitable to conduct SBA activities

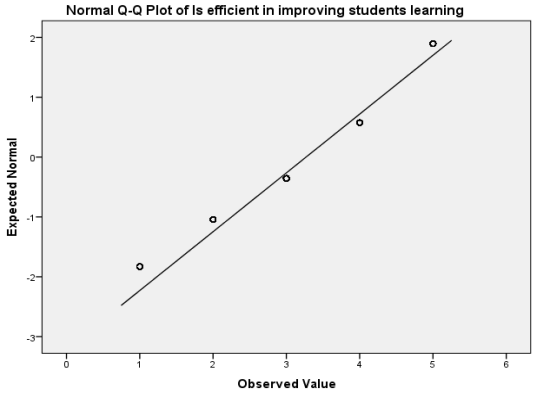




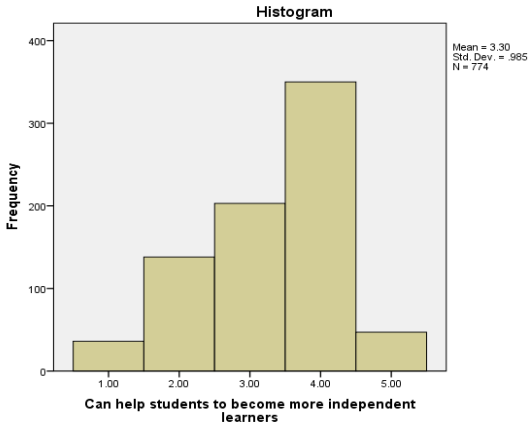
For process evaluation items:

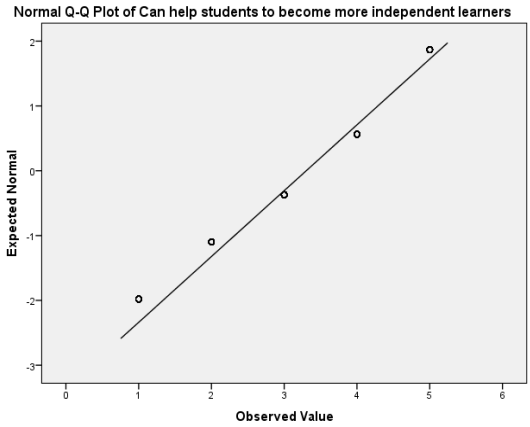
For item a1: Is efficient in improving students learning



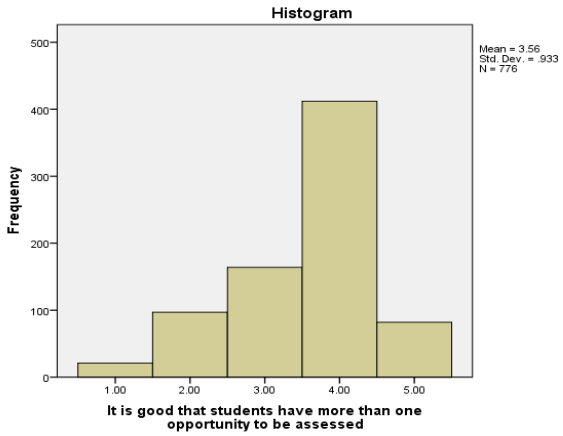


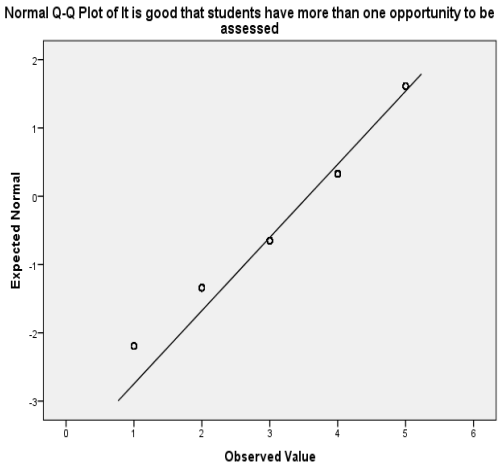
For item a2: Can help students to become more independent learners



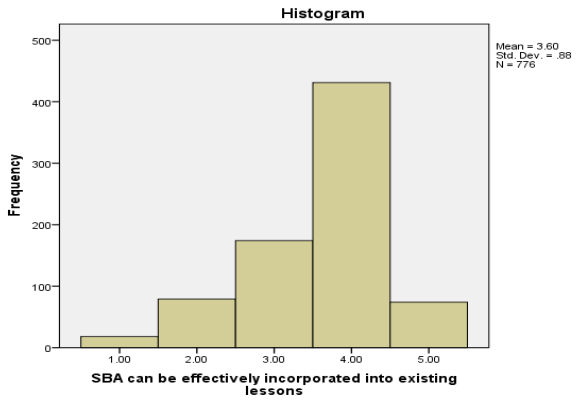


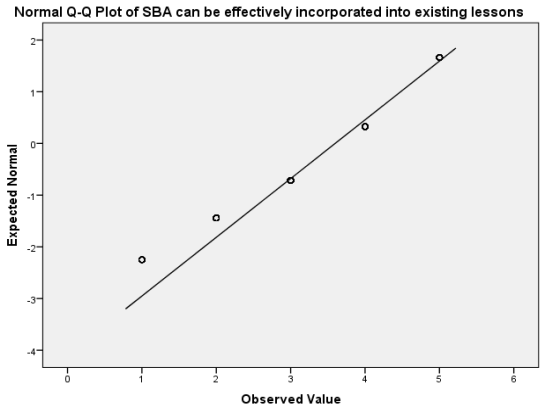
For item a3: It is good that students have more than one opportunity to be assessed



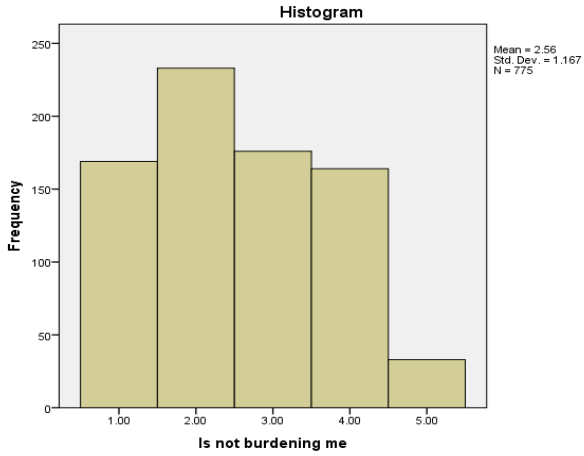


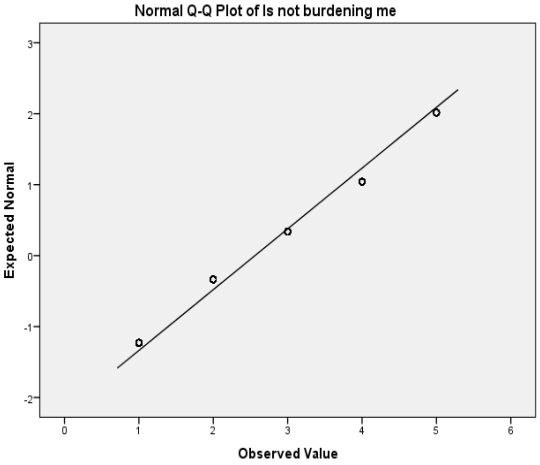
For item a4: SBA can be effectively incorporated into existing lessons



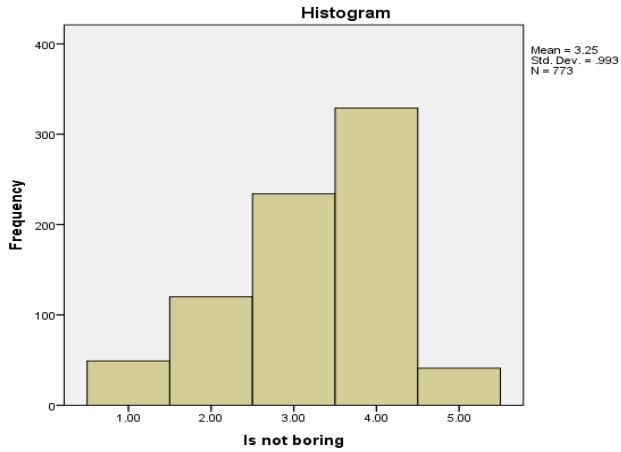


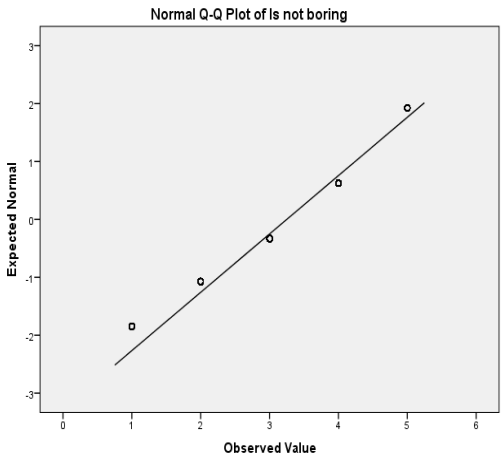
For item a5: Is not burdening me



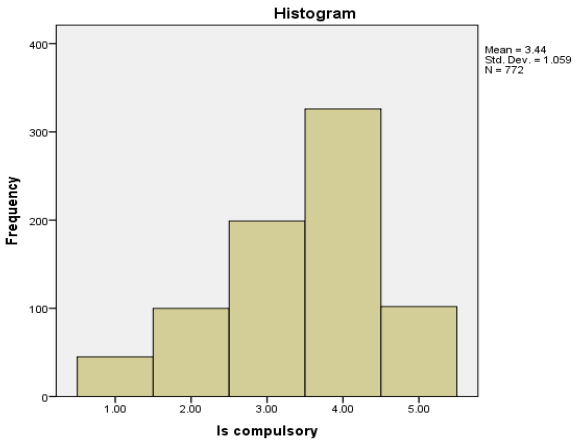


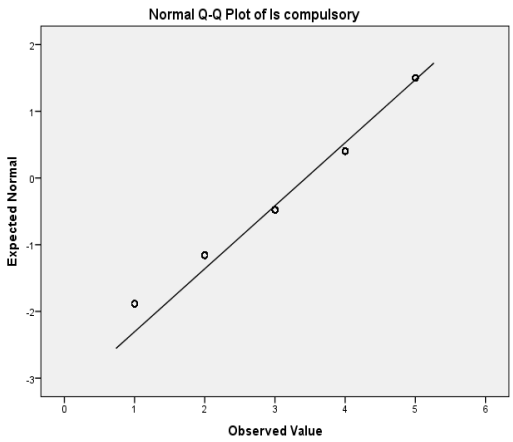
For item a6: Is not boring



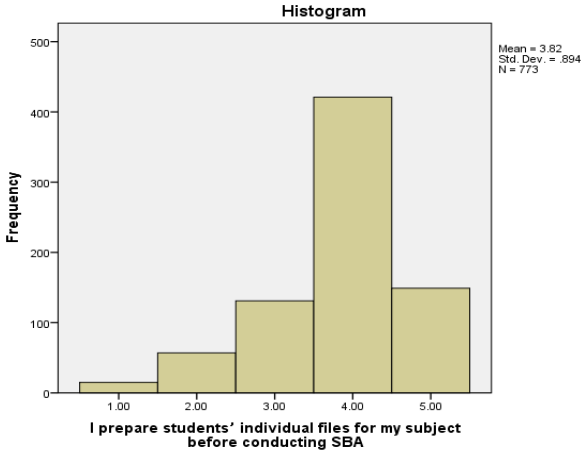


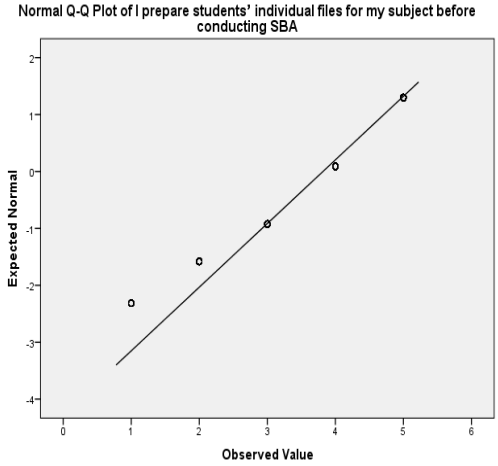
For item a7: Is compulsory



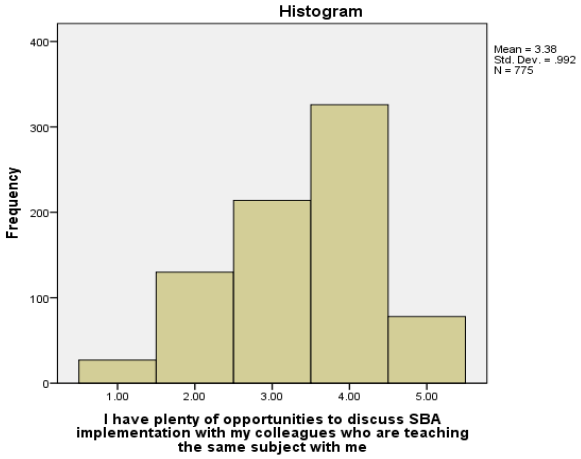


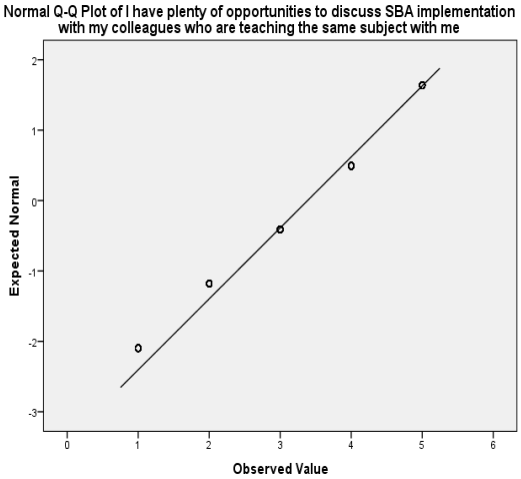
For item a8: I prepare students' individual files for my subject before conducting SBA



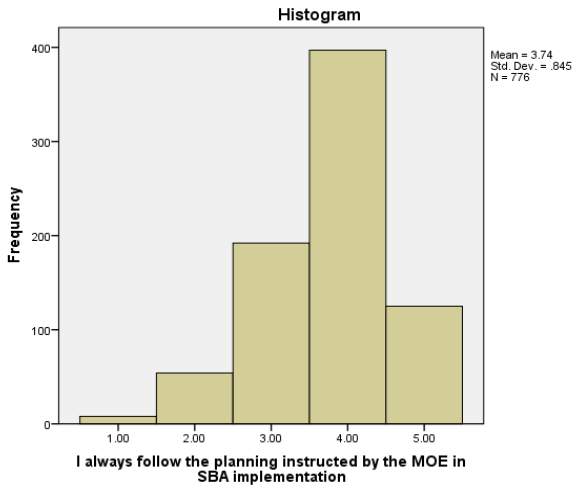


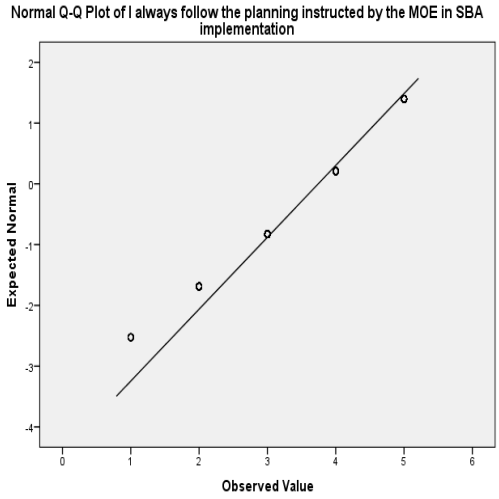
For item a9: I have plenty of opportunities to discuss SBA implementation with my colleagues who are teaching the same subject with me



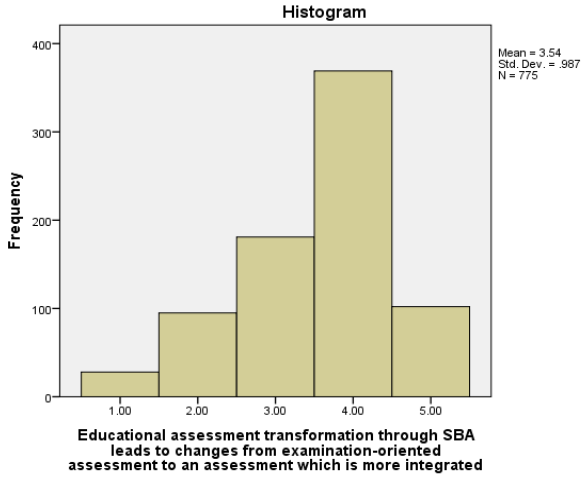


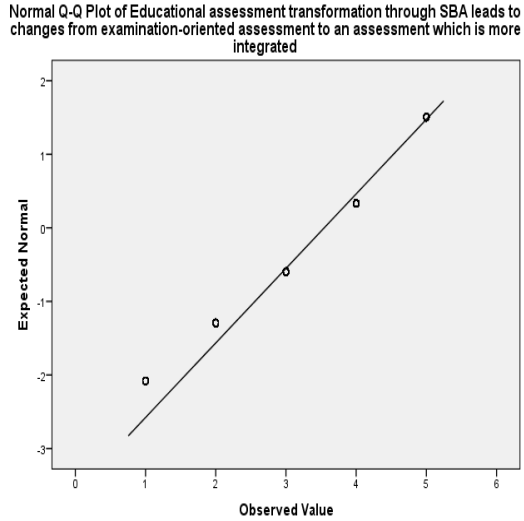
For item a10: I always follow the planning instructed by the MOE in SBA implementation



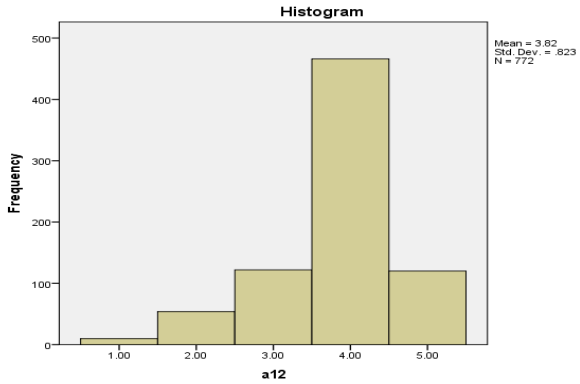


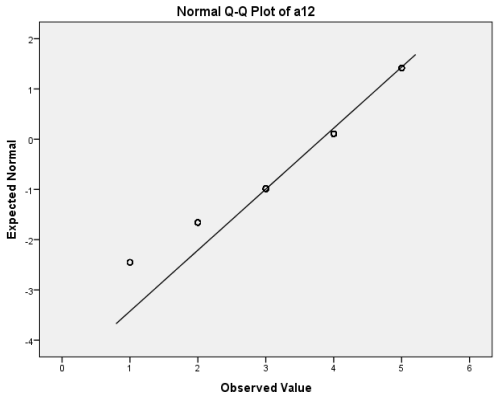
For item a11: Educational assessment transformation through SBA leads to changes from examination-oriented assessment to an assessment which is more integrated



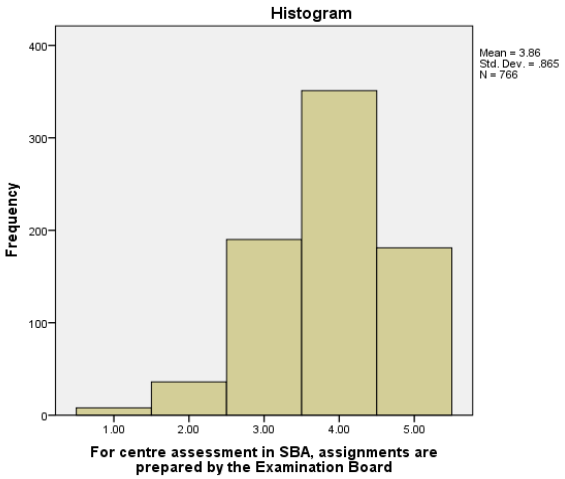


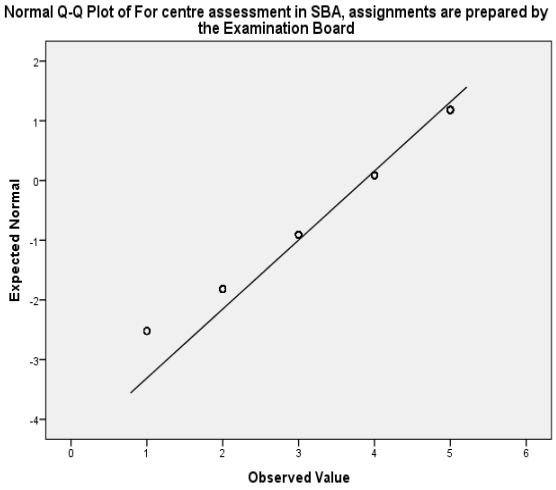
For item a12: SBA is conducted during the teaching and learning process



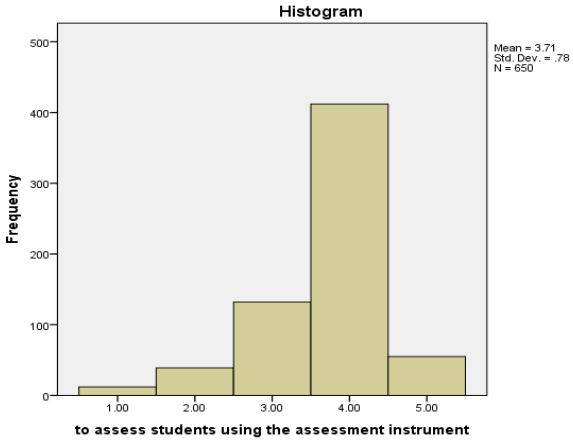


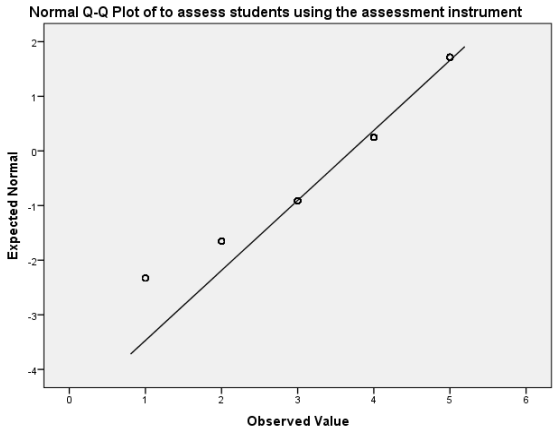
For item a13: For centre assessment in SBA, assignments are prepared by the Examination Board



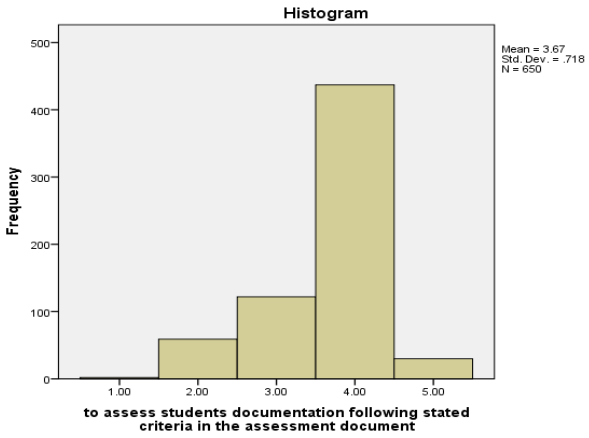


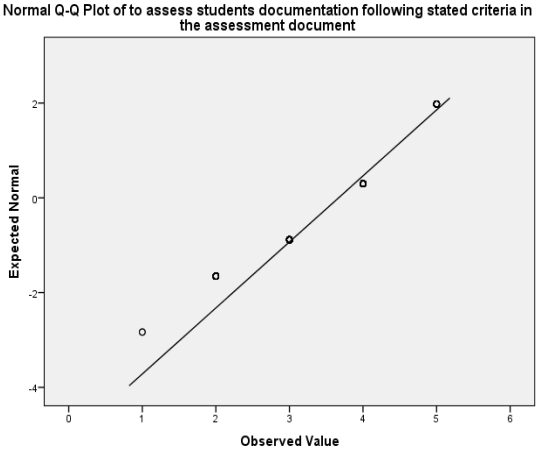
For item b24i: to assess students using the assessment instrument



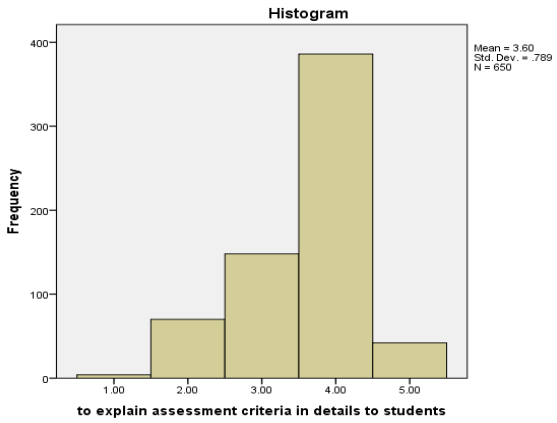


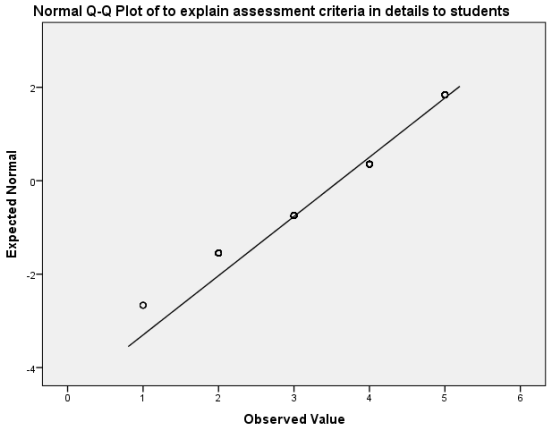
For item b24ii: to assess students' documentation following stated criteria in the assessment document



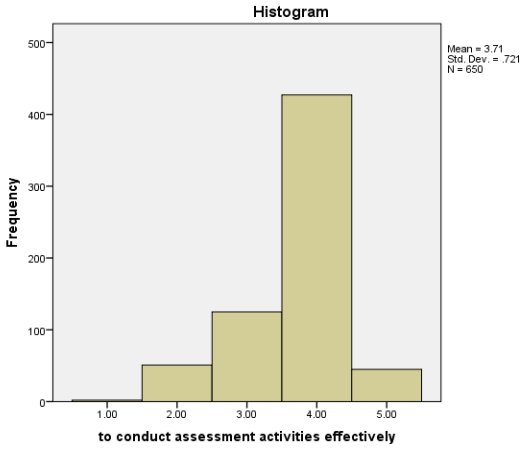


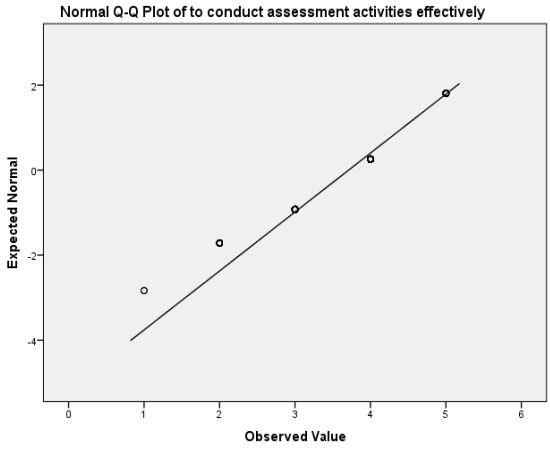
For item b24iii: to explain assessment criteria in details to students



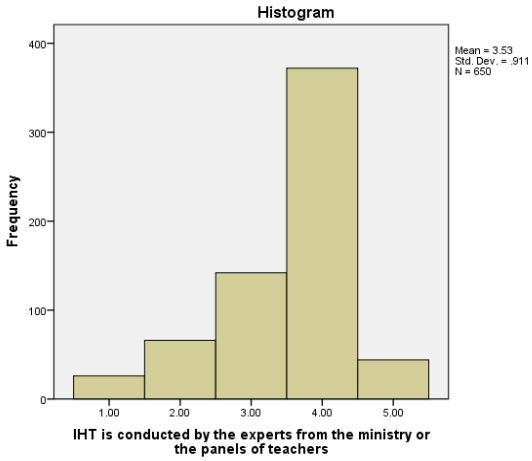


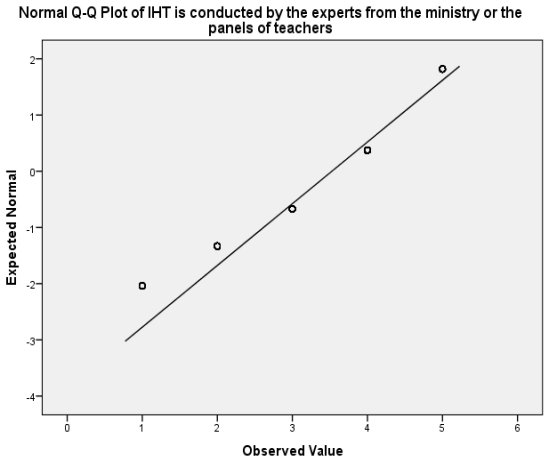
For item b24iv: to conduct assessment activities effectively



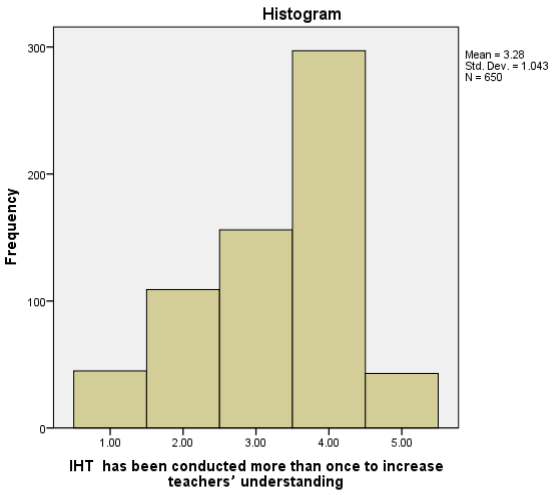


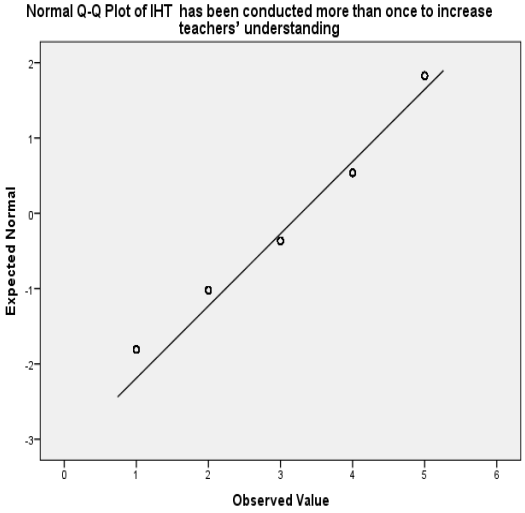
For item b25i: IHT is conducted by the experts from the ministry or the panels of teachers



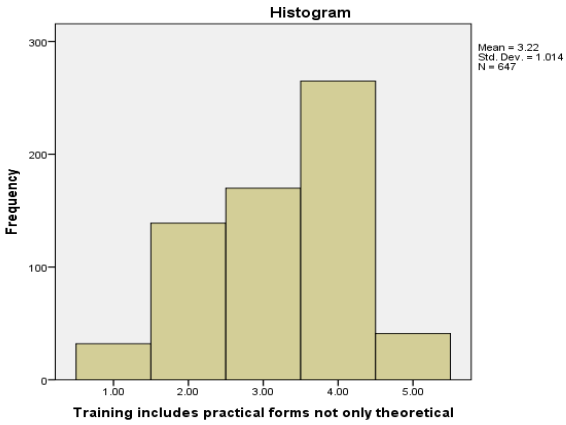


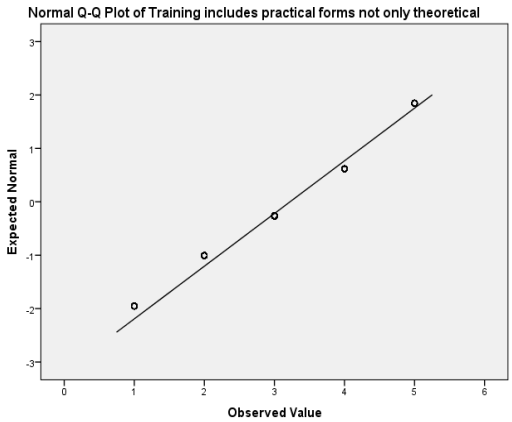
For item b25ii: IHT has been conducted more than once to increase teachers' understanding



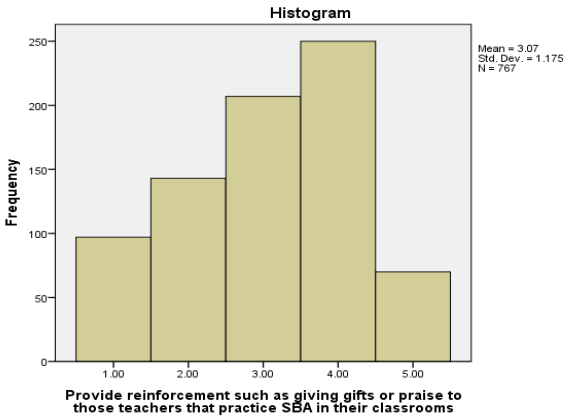


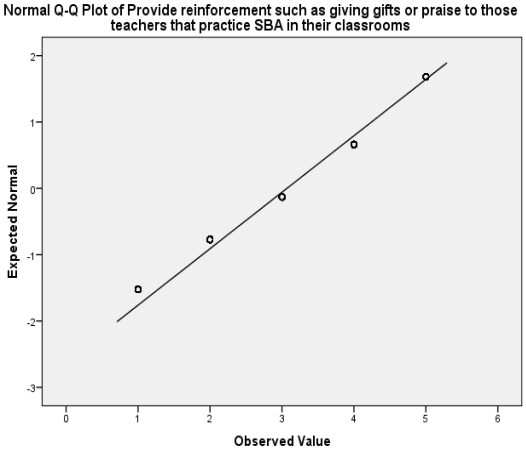
For item b25iii: Training includes practical forms not only theoretical



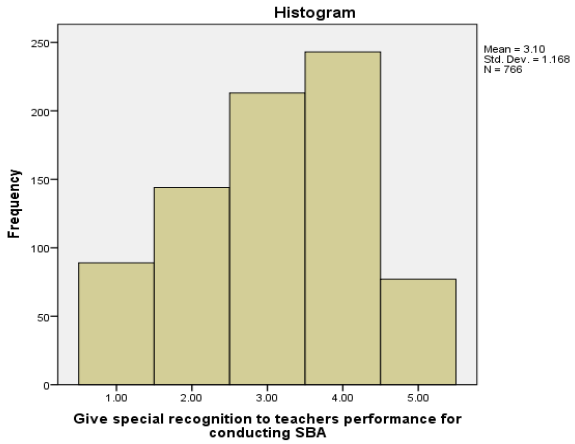


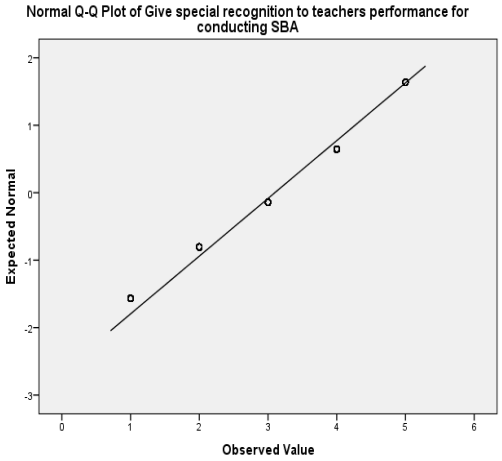
For item 26i: Provide reinforcement such as giving gifts or praise to those teachers that practice SBA in their classrooms



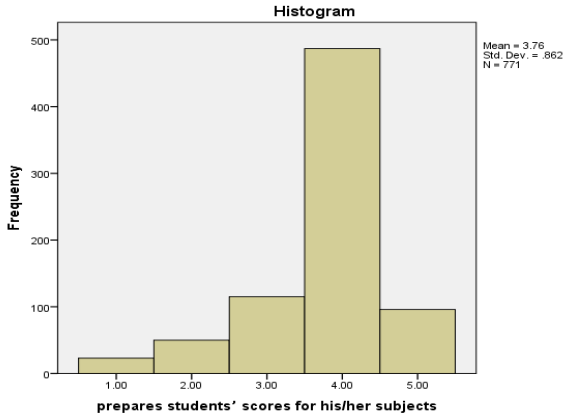


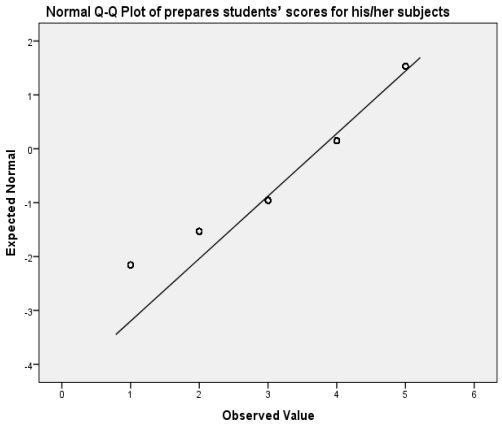
For item 26ii: Give special recognition to teachers' performance for conducting SBA



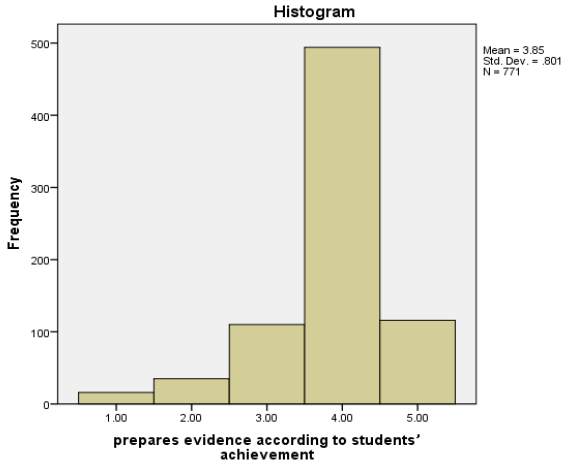


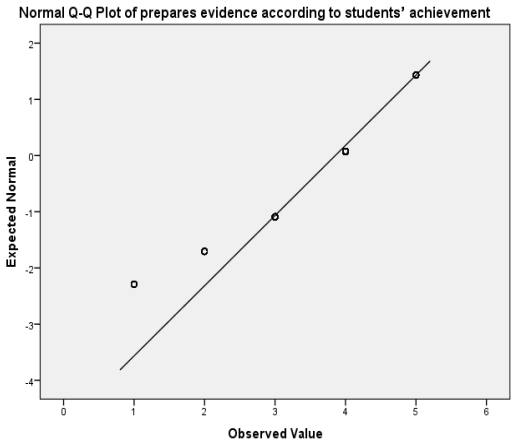
For item 27i: prepares students' scores for his or her subjects



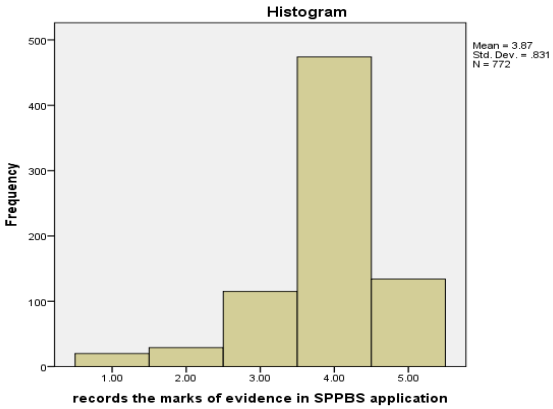


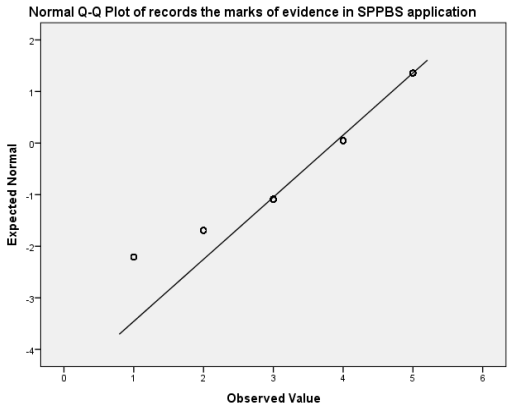
For item 27ii: prepares evidence according to students' achievement



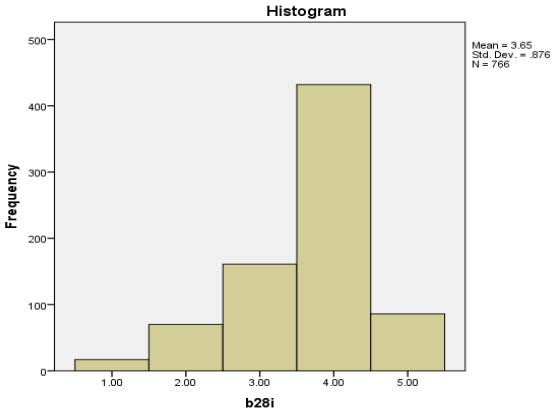


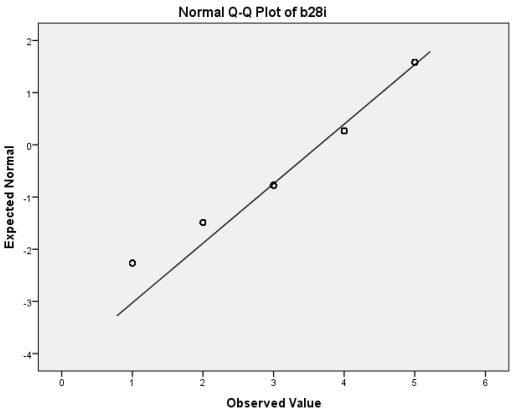
For item 27iii: records the marks of evidence in SPPBS application



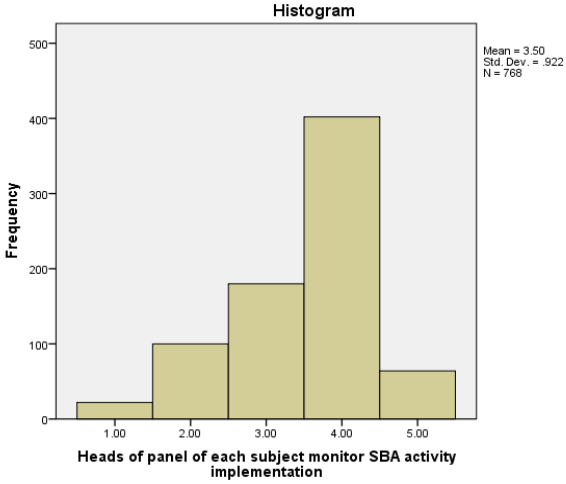


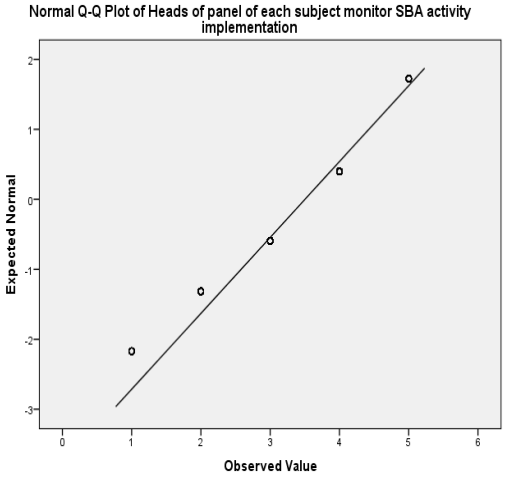
For item 28i: The school SBA committee appoints certain teachers as internal monitors



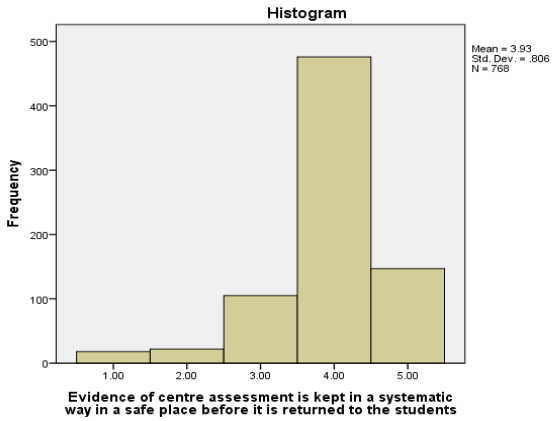


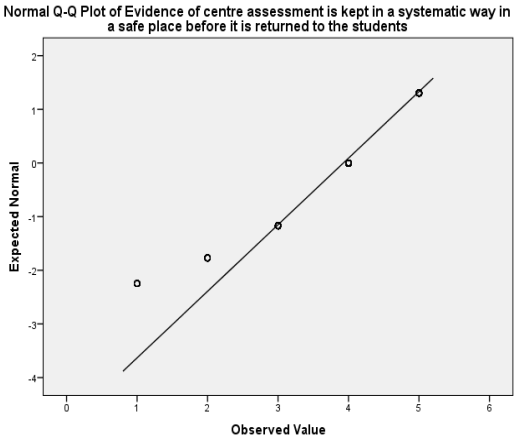
For item 28ii: Heads of panel of each subject monitor SBA activity implementation



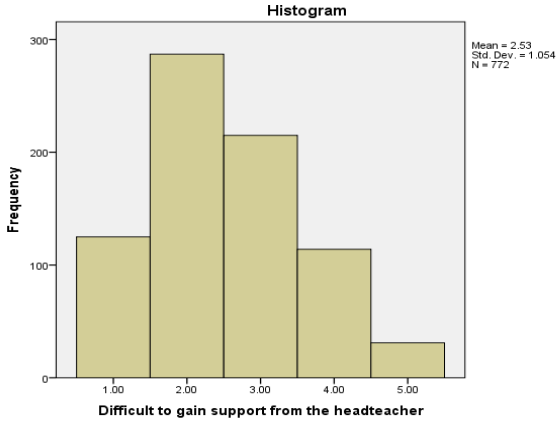


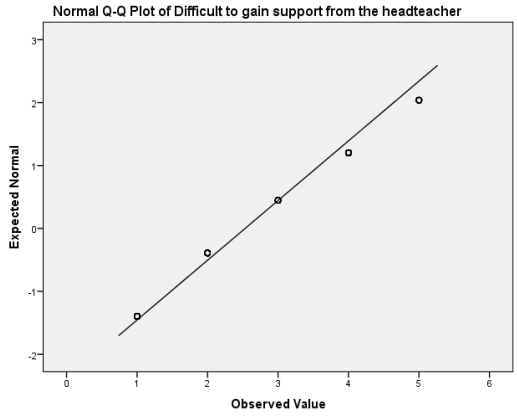
For item 28iii: Evidence of centre assessment is kept in a systematic way in a safe place before it is returned to the students



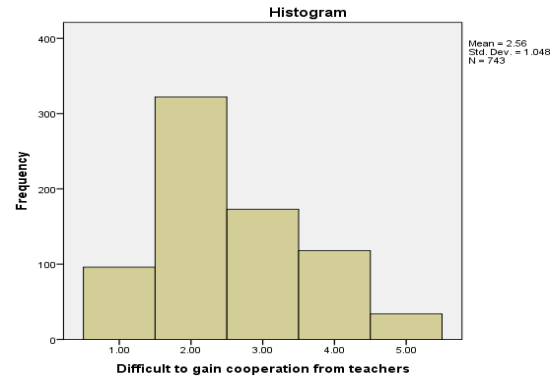


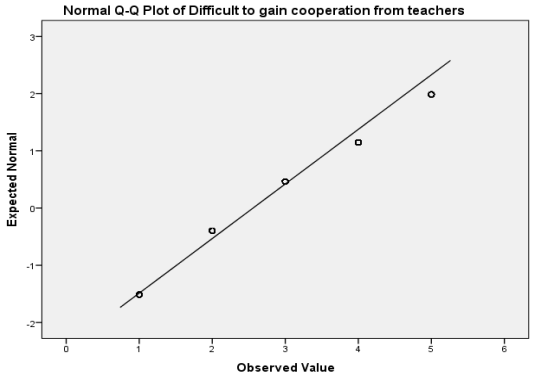
For item c1: Difficult to gain support from the headteacher



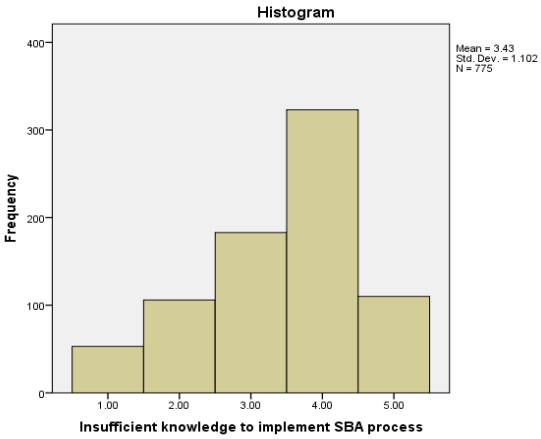


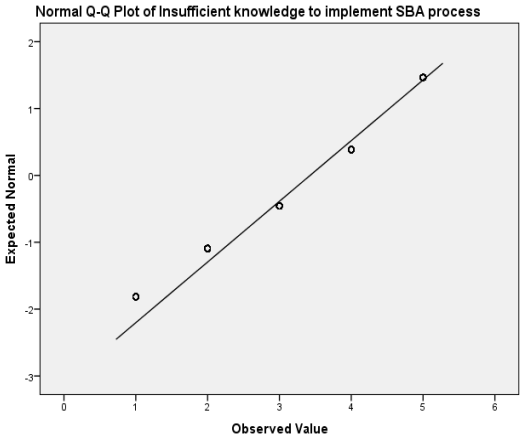
For item c2: Difficult to gain cooperation from teachers



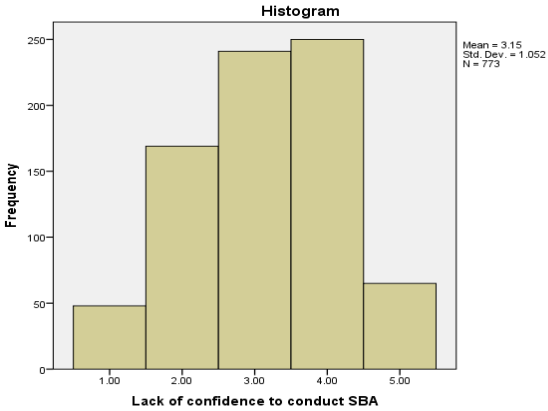


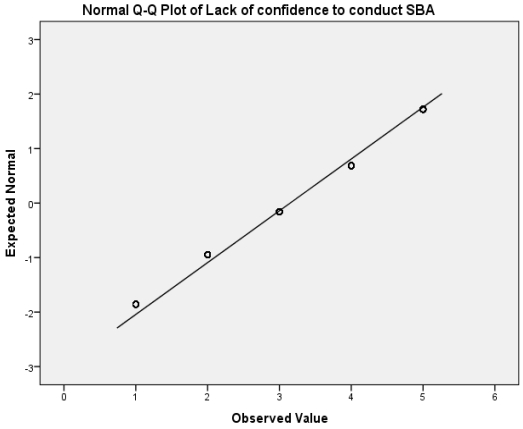
For item c3: Insufficient knowledge to implement SBA process



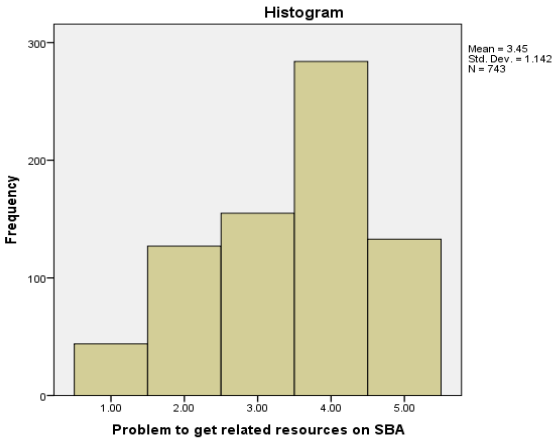


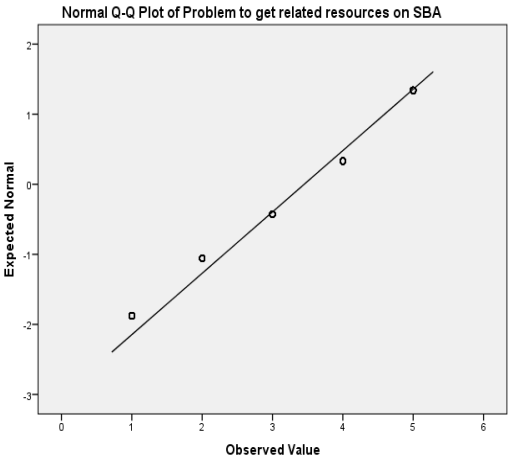
For item c4: Lack of confidence to conduct SBA



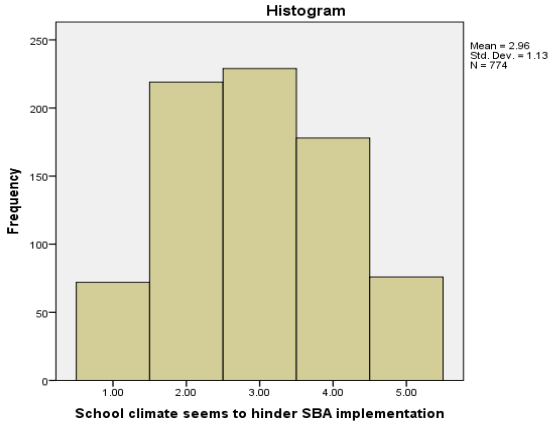


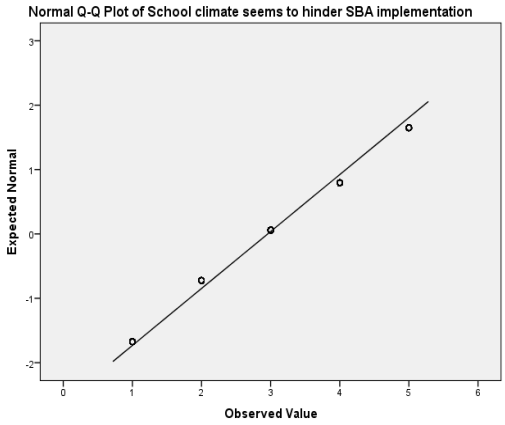
For item c5: Problem to get related resources on SBA



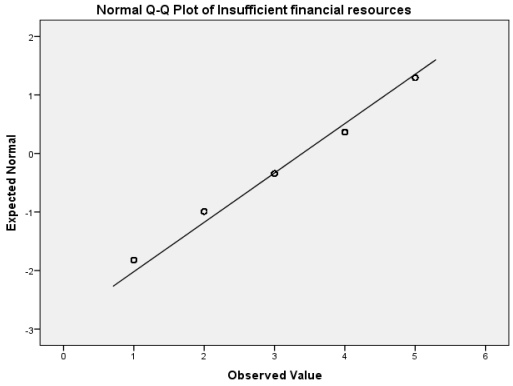


For item c6: School climate seems to hinder SBA implementation

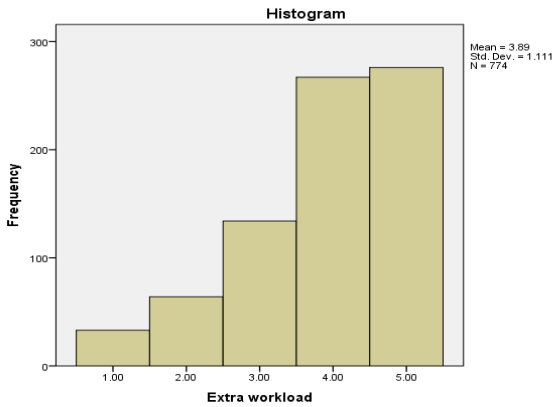


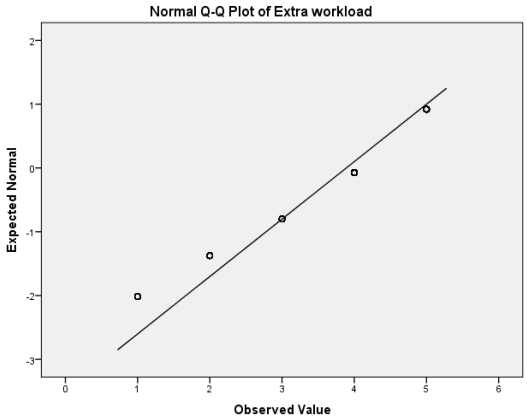


For item c7: Insufficient financial resources

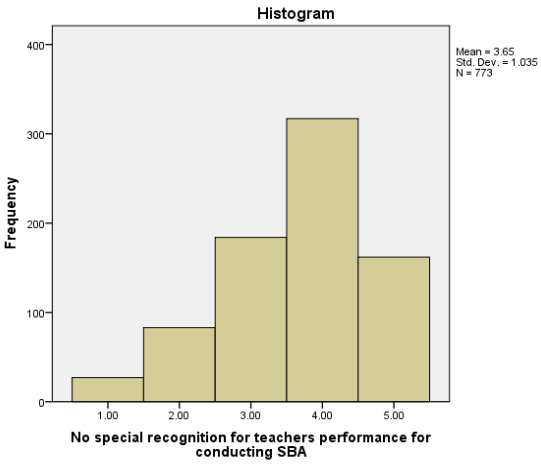


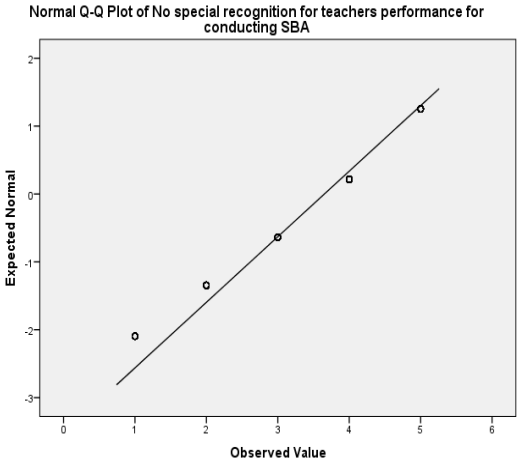
For item c8: Extra workload



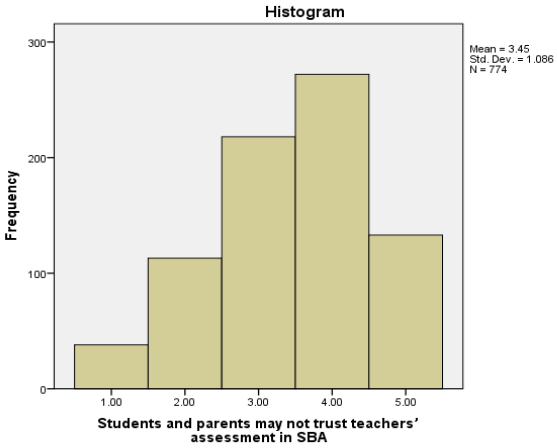


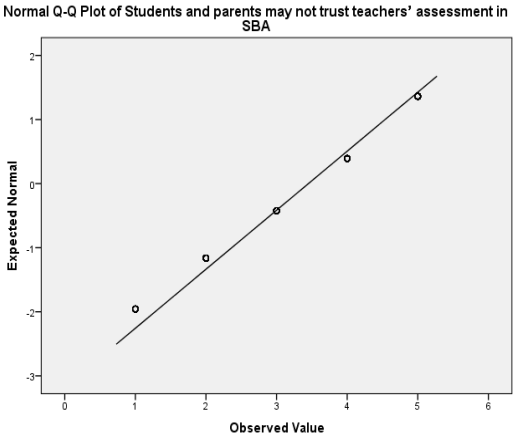
For item c9: No special recognition for teachers' performance for conducting SBA



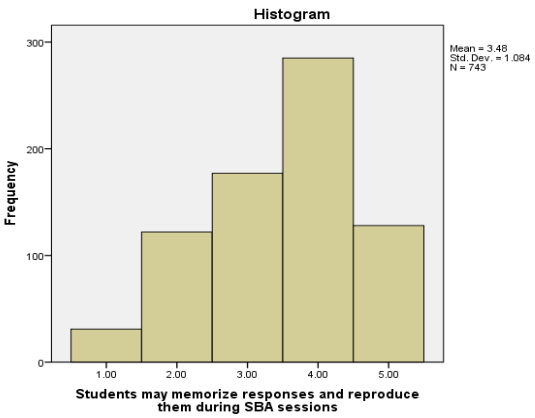


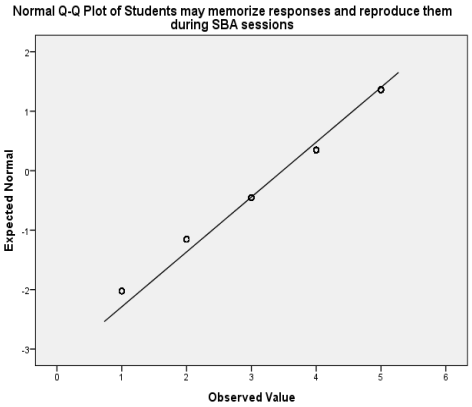
For item c10: Students and parents may not trust teachers' assessment in SBA



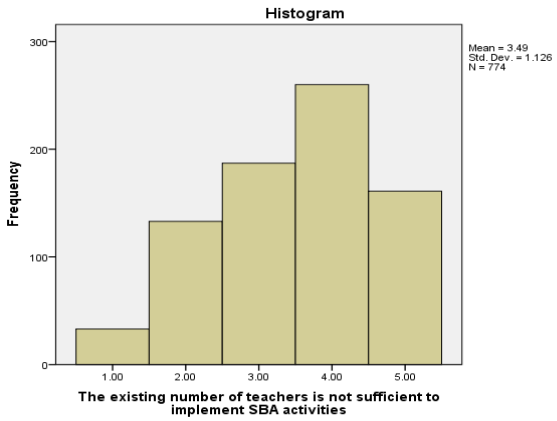


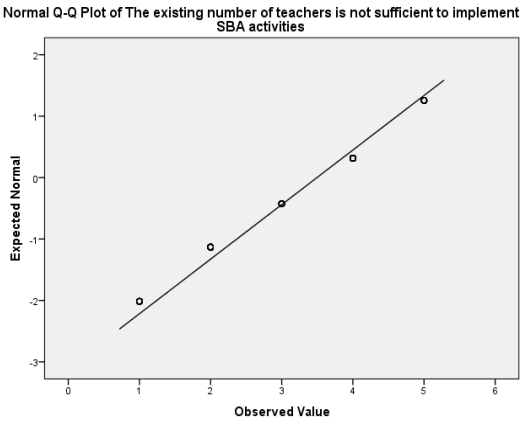
For item c11: Students may memorize responses and reproduce them during SBA sessions



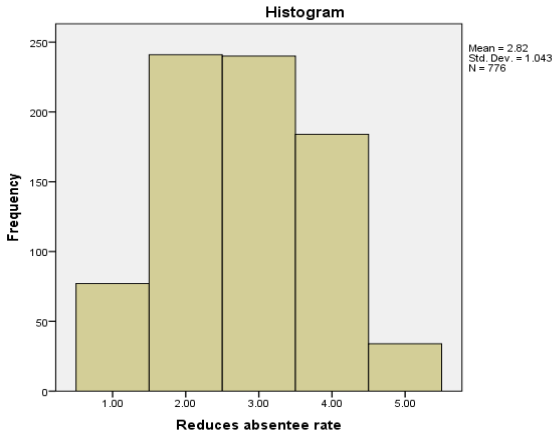


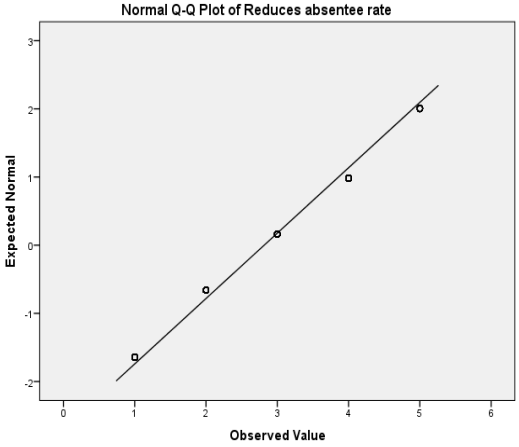
For item c12: The existing number of teachers is not sufficient to implement SBA activities



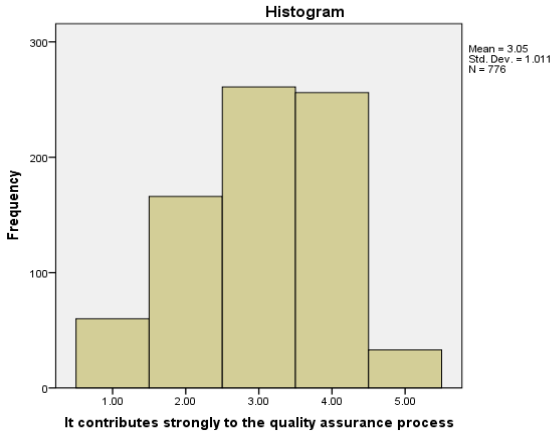


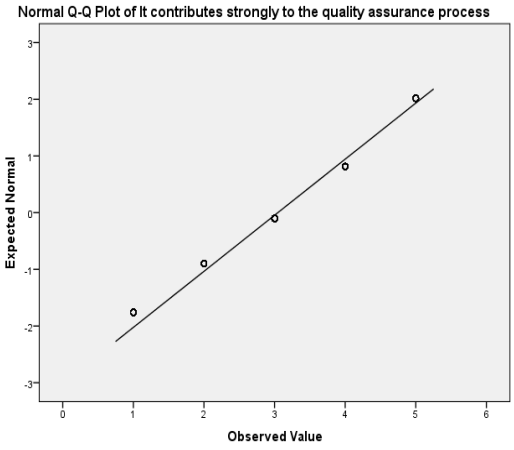
For item d30i: Reduces absentee rate



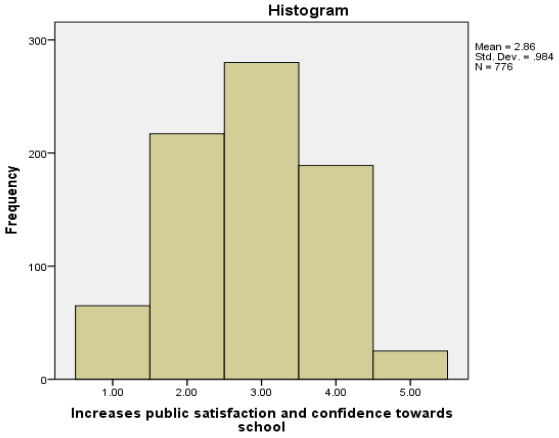


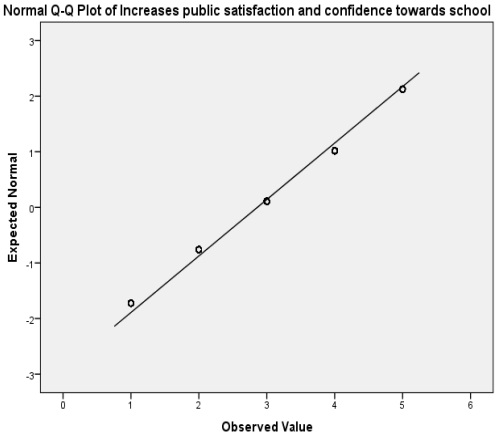
For item d30ii: It contributes strongly to the quality assurance process





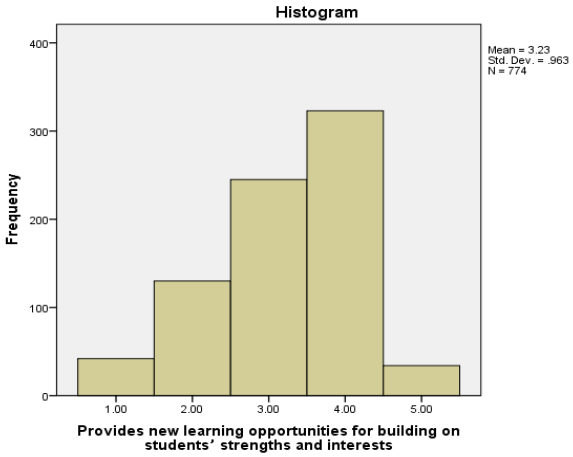
For item d30iii: Increases public satisfaction and confidence towards school

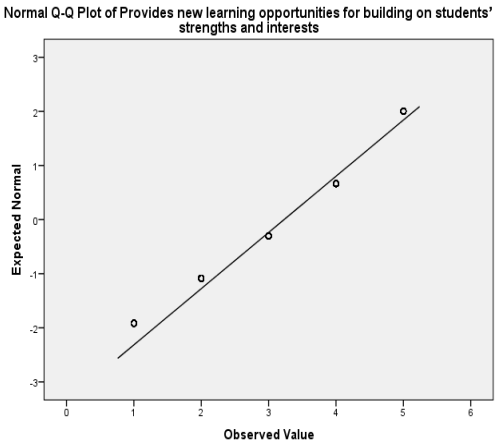




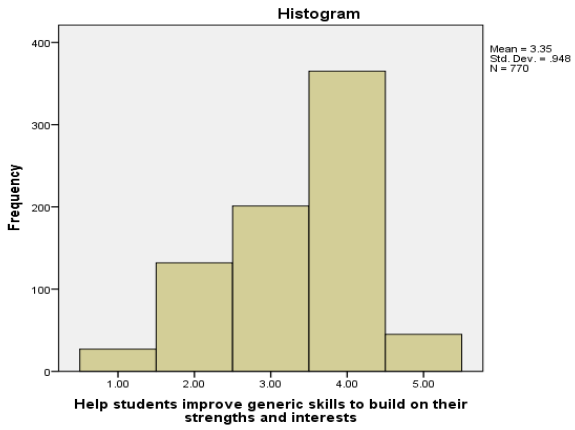
For item d30iv: Provides new learning opportunities for building on students' strengths and

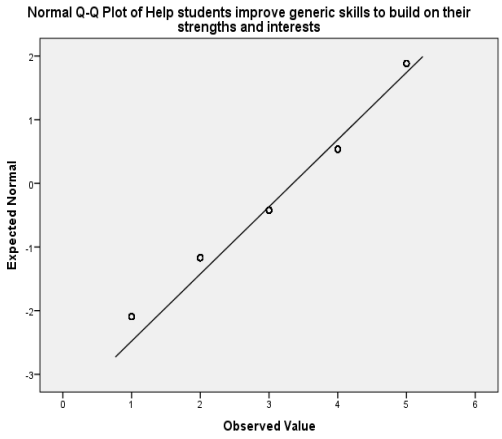
interests



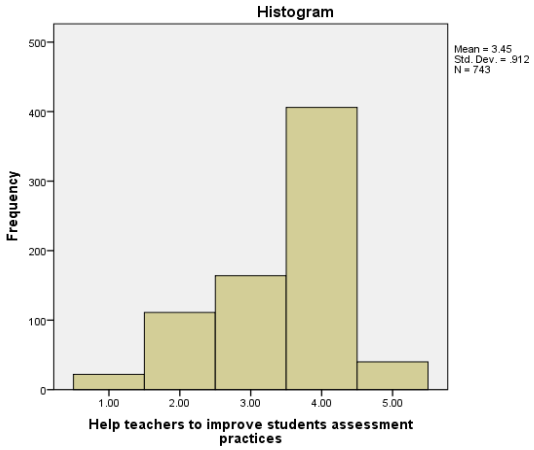


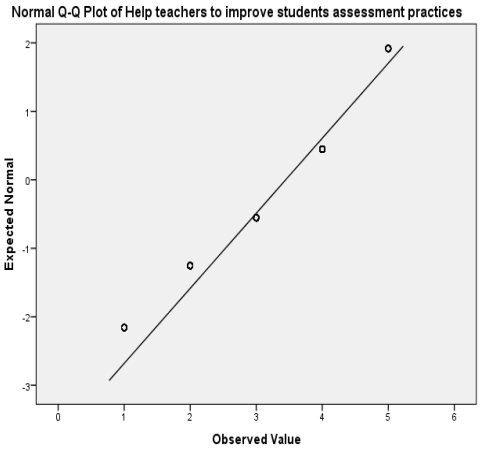
For item d31i: Help students improve generic skills to build on their strengths and interests



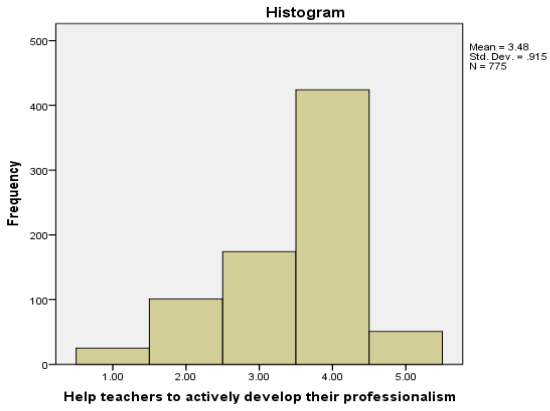


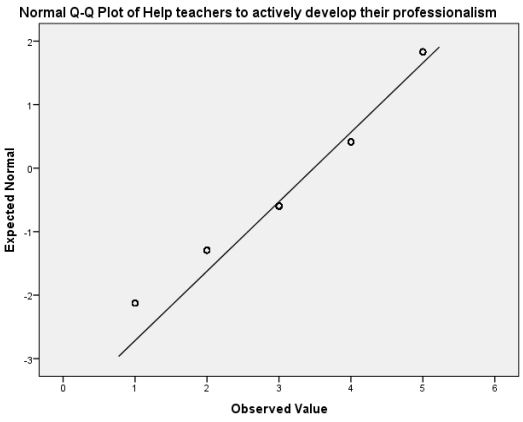
For item d31ii: Help teachers to improve students' assessment practices



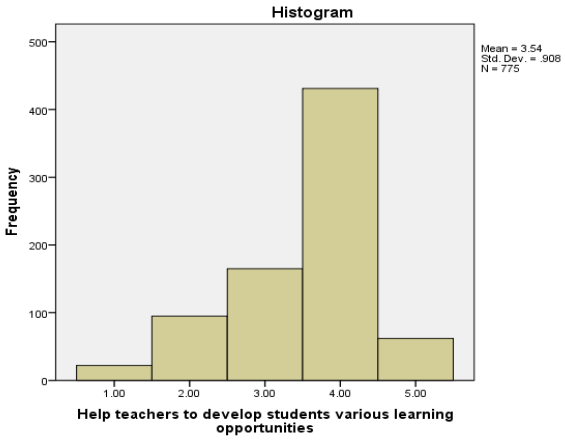


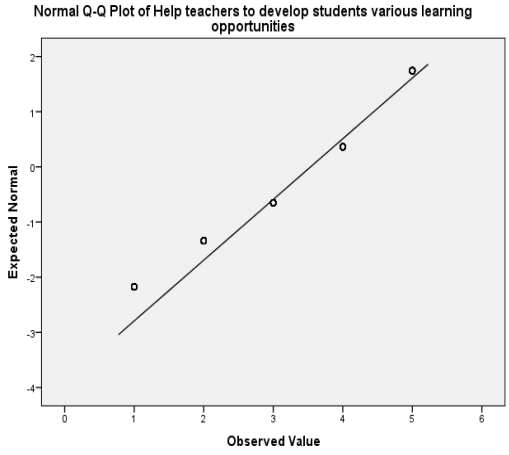
For item d31iii: Help teachers to actively develop their professionalism



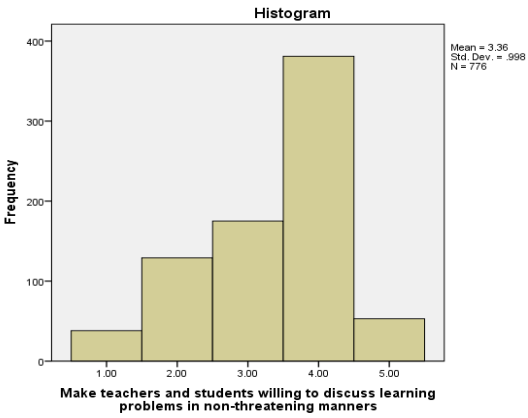


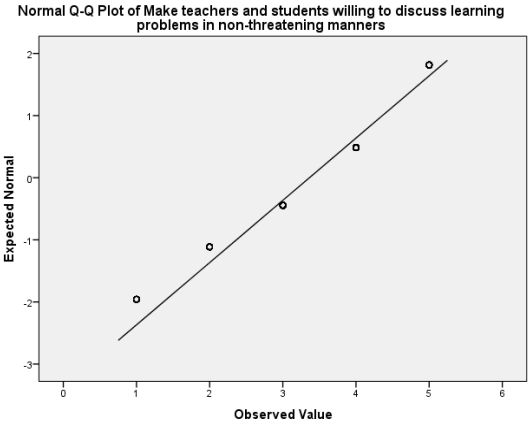
For item d31iv: Help teachers to develop students various learning opportunities



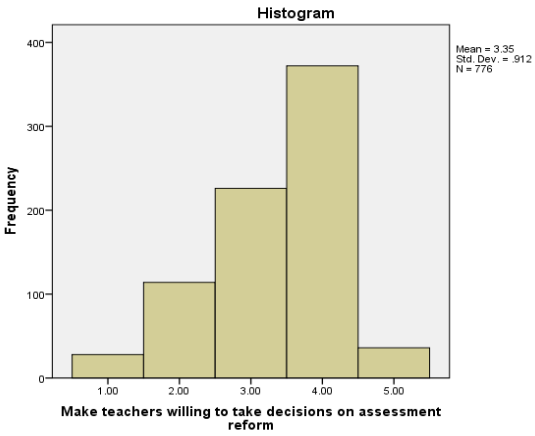


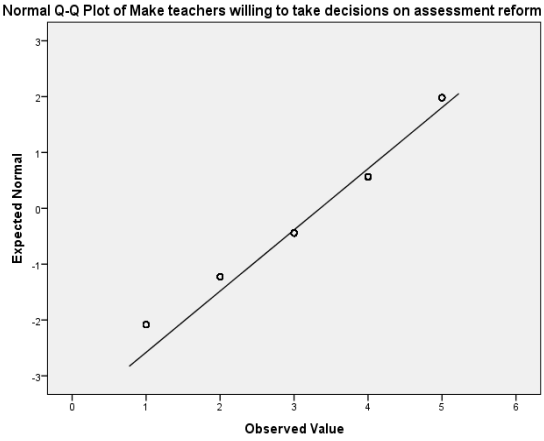
For item d31v: Make teachers and students willing to discuss learning problems in non-threatening manners



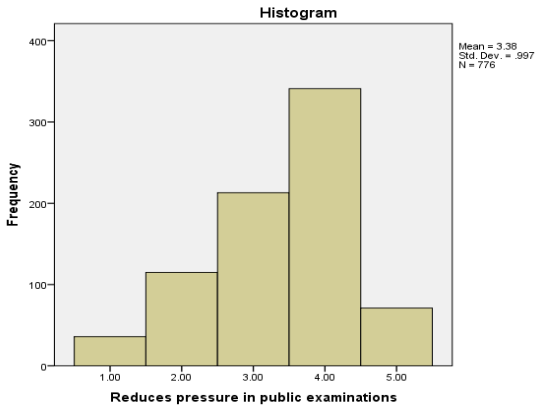


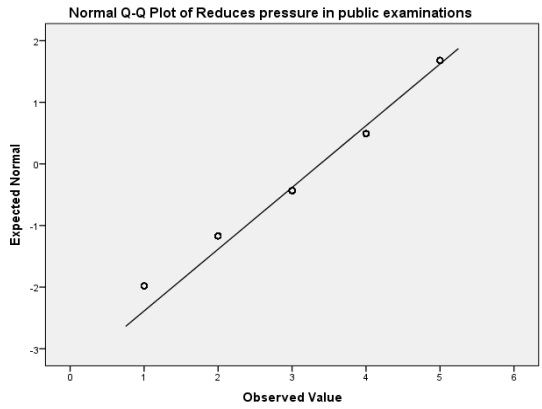
For item d31vi: Make teachers willing to take decisions on assessment reform



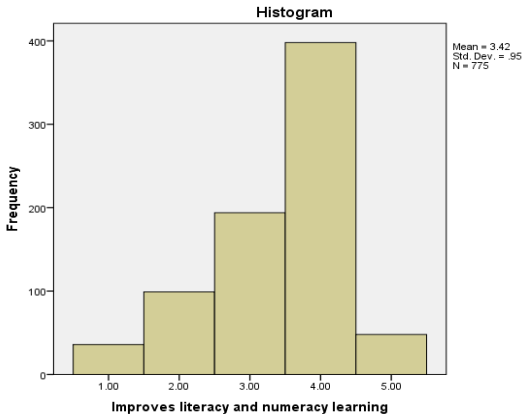


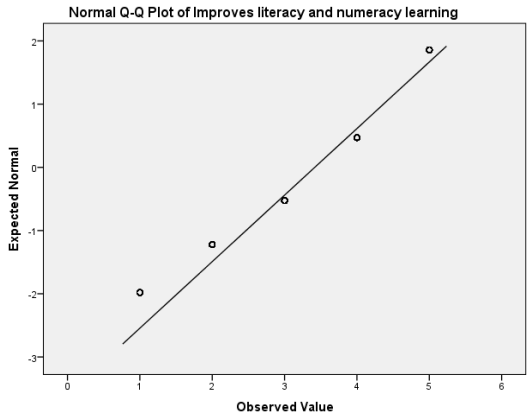
For item d31vii: Reduces pressure in public examinations





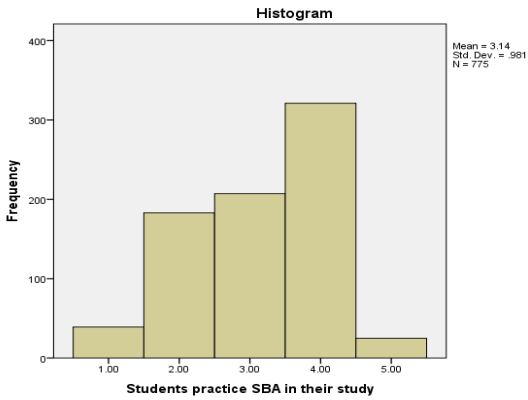
For item d31viii: Improves literacy and numeracy learning

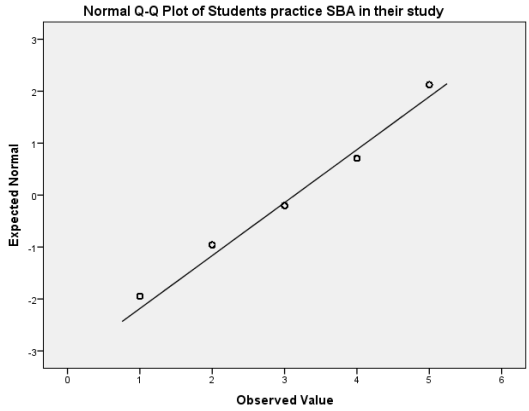




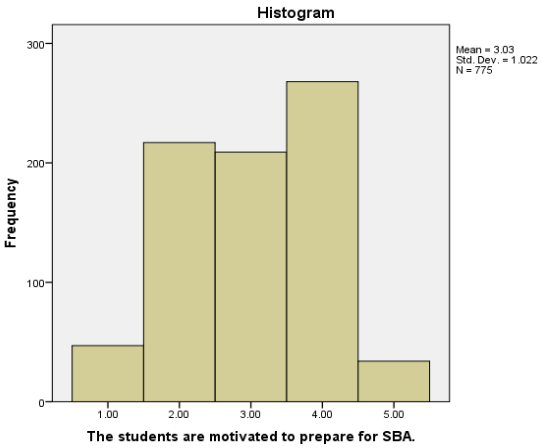
For product evaluation items:

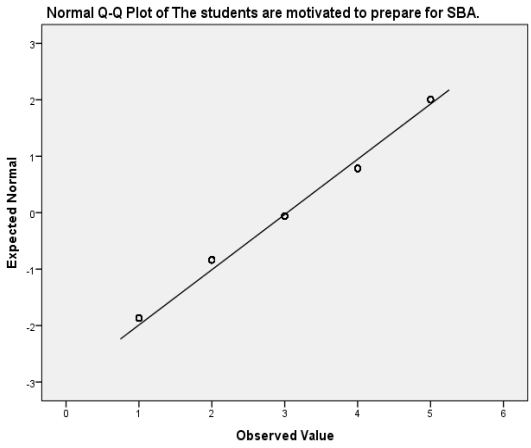
For item e32i: Students practice SBA in their study



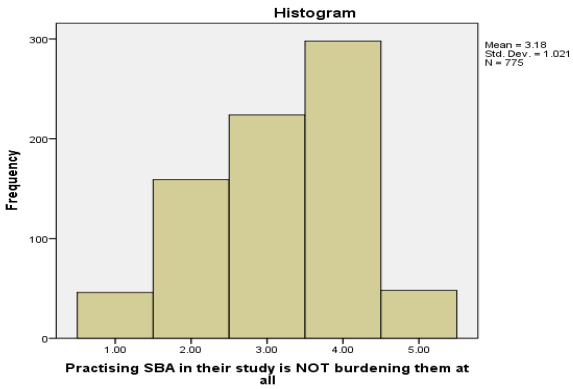


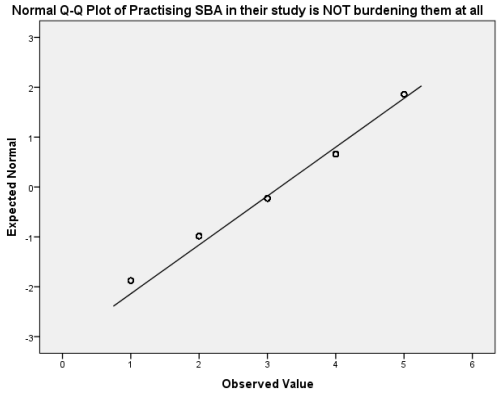
For item e32ii: The students are motivated to prepare for SBA



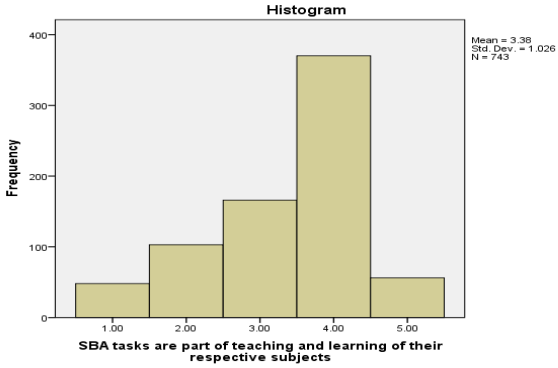


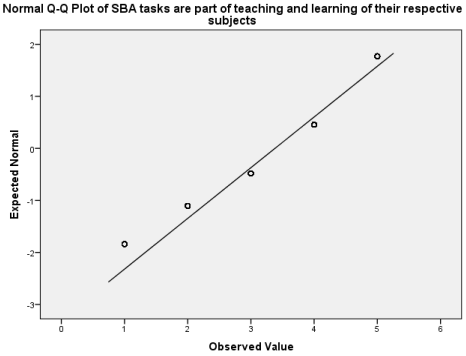
For item e32iii: Practising SBA in their study is NOT burdening them at all



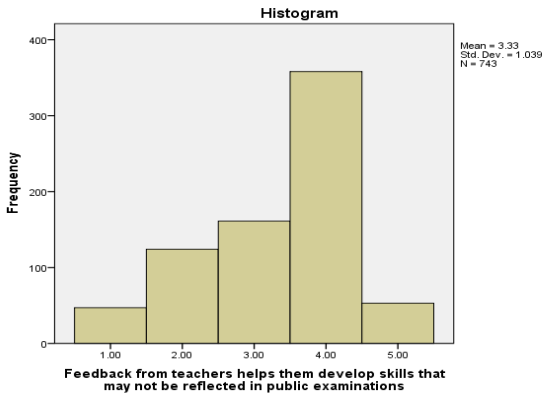


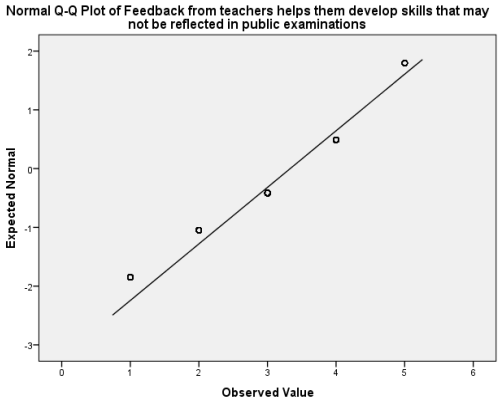
For item e33i: SBA tasks are part of teaching and learning of their respective subjects



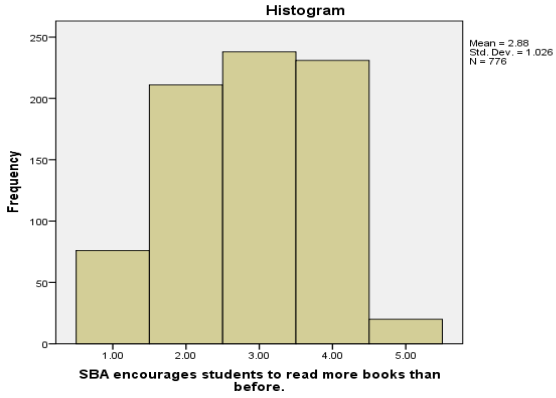


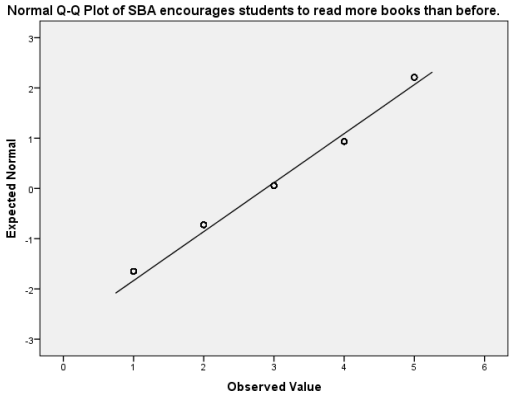
For item e33ii: Feedback from teachers helps them develop skills that may not be reflected in public examinations



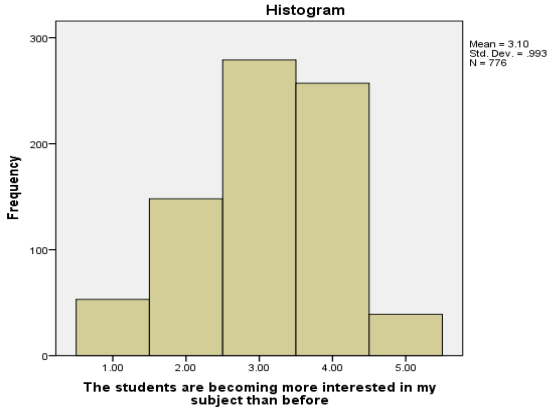


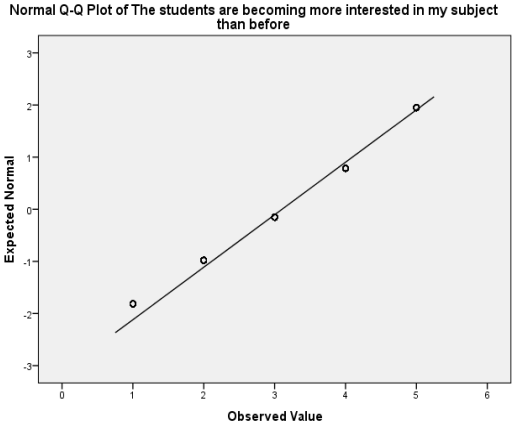
For item e34i: SBA encourages students to read more books than before



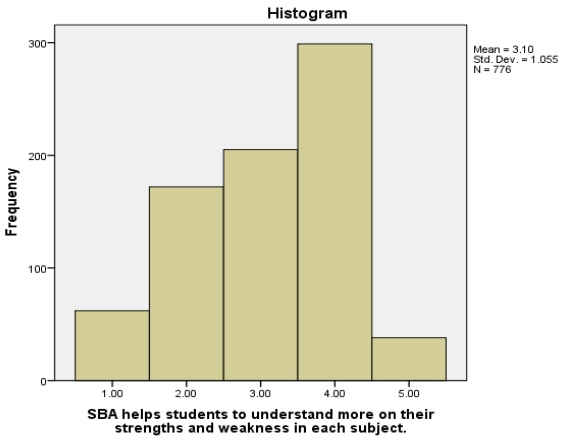


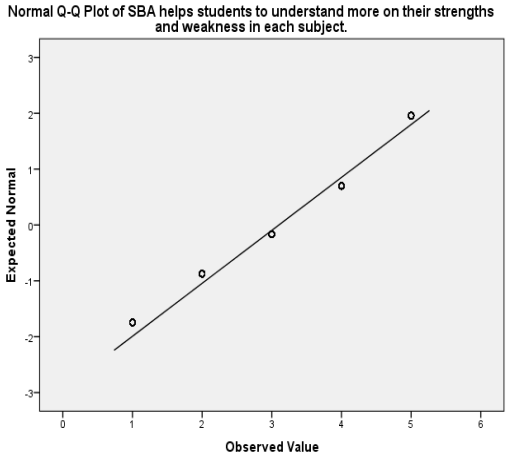
For item e34ii: The students are becoming more interested in my subject than before





For item e34iii: SBA helps students to understand more on their strengths and weakness in each subject





Appendix J: Missing data statistics by item

			Missing data		
Construct	Item	Valid N	Count	Percent	Mode
Input evaluation construct					
Material and personal needs in SBA system	a14	774	2	0.3	5
	a15	775	1	0.1	5
	a16	776	0	0.0	5
Appropriateness of personnel's qualifications	a17	774	2	0.3	4
	a18	776	0	0.0	4
Suitability of physical infrastructure and ICT	a19	776	0	0.0	4
	a20	773	3	0.4	4
	a21	775	1	0.1	4
Process evaluation construct					
Teacher's attitude: teacher's belief	a1	773	3	0.4	4
	a2	774	2	0.3	4
	a3	776	0	0.0	4
	a4	776	0	0.0	4
Teacher's attitude: teacher's feeling	a5	775	1	0.1	2
	a6	773	3	0.4	4
	a7	772	4	0.5	4
Teacher's attitude: teacher's readiness	a8	773	3	0.4	4
	a9	775	1	0.1	4
	a10	776	0	0.0	4
Teacher's understanding on SBA	a11	775	1	0.1	4
	a12	772	4	0.5	4
	a13	766	10	1.3	4
Effect of SBA courses on improving skills of SBA	b24i	650	0	0.0	4
	b24ii	650	0	0.0	4
	b24iii	650	0	0.0	4
	b24iv	650	0	0.0	4
In-house training on SBA	b25i	650	0	0.0	4
	b25ii	650	0	0.0	4
	b25iii	647	3	0.5	4

Encouragement by administration	b26i b26ii	767 766	9 10	1.2 1.3	4 4
Coordination process	b27i b27ii b27iii	771 771 772	5 5 4	0.6 0.6 0.5	4 4 4
Monitoring process	b28i b28ii b28iii	766 768 768	10 8 8	1.3 1.0 1.0	4 4 4
Challenges	c1 c2 c3 c4 c5 c6 c7 c8 c9 c10 c11 c12	772 773 775 773 770 774 772 774 773 774 773 774	4 3 1 3 6 2 4 2 3 2 3 2	0.5 0.4 0.1 0.4 0.8 0.3 0.5 0.3 0.4 0.3 0.4 0.3	2 2 4 4 4 3 4 5 4 4 4 4
School Improvement Construct					
Role of SBA	d30i d30ii d30iii d30iv	776 776 776 774	0 0 0 2	0.0 0.0 0.0 0.3	2 3 3 4
Crucial of SBA	d31i d31ii d31iii d31iv d31v d31vi d31vii d31viii	770 773 775 775 776 776 776 775	6 3 1 1 0 0 0 1	0.8 0.4 0.1 0.1 0.0 0.0 0.0 0.1	4 4 4 4 4 4 4 4
Product evaluation construct					
Student's attitude towards SBA	e32i e32ii e32iii	775 775 775	1 1 1	0.1 0.1 0.1	4 4 4
Student's knowledge in SBA	e33i e33ii	773 772	3 4	0.4 0.5	4 4
Motivational source towards learning	e34i e34ii e34iii	776 776 776	0 0 0	0.0 0.0 0.0	3 3 4
Total	68	51886	152		

Appendix K: SBA Evaluation Questionnaire for Pilot Study (English Version)

Please answer **ALL** questions honestly. There is no right or wrong answer. Your responses will be strictly confidential and will be used only for academic purposes.

SECTION A:

Using the following 1-5 scale, please indicate by circling the most correct response, the degree to which you agree with the statements listed below:

1	2	3	4	5
Strongly disagree	**disagree**	**neutral**	**agree**	**strongly agree**

Using school-based assessment (SBA) in teaching:

A1	Is efficient in improving students learning	1	2	3	4	5
A2	Can help students to become more independent learners	1	2	3	4	5
A3	It excites me	1	2	3	4	5
A4	Is not burdening me	1	2	3	4	5
A5	It is good that students have more than one opportunity to be assessed	1	2	3	4	5
A6	It is better for the students to be assessed by their teachers and not external		2	3	4	5
A7	SBA can be effectively incorporated into existing lessons	1	2	3	4	5
A8	I attend briefings on SBA conducted in schools					
A9	I prepare students' individual files for my subject before conducting SBA	1	2	3	4	5
A10	I have plenty of opportunities to discuss SBA implementation with my colleagues who are teaching the same subject with me	1	2	3	4	5
A11	I always follow the planning instructed by the MOE in SBA implementation	1	2	3	4	5
A12	Educational assessment transformation through SBA leads to changes from examination-oriented assessment to an assessment which is more integrated	1	2	3	4	5
A13	SBA is conducted during the teaching and learning process	1	2	3	4	5

A14	For centre assessment in SBA, assignments are prepared by the Examination Board	1	2	3	4	5
A15	It is easy to implement assessment if teachers are supplied with a complete assessment document	1	2	3	4	5
A16	A teaching assisstant is needed to help teachers in assessment	1	2	3	4	5
A17	Training of the personnel involved should be properly planned and implemented	1	2	3	4	5
A18	Teacher-student ratio is sufficient to implement SBA					
A19	Teachers are sufficiently qualified to implement assessment activities	1	2	3	4	5
A20	Teachers' skills in assessment are adequate for implementing assessment activities	1	2	3	4	5
A21	The physical equipment in classroom is suitable to conduct SBA activities	1	2	3	4	5
A22	The space can be adapted to assessment activities	1	2	3	4	5
A23	The ICT hardware is suitable to conduct SBA activities	1	2	3	4	5

SECTION B:

B24. Have you ever attend any courses/ any training related to SBA system?

YES ☐ NO ☐

If you have answered "NO" to question 24 please go to question 28, otherwise continue with question 25.

B25. Information on SBA courses that you have attended:

No	Name of Course/ Organizer	No of days	Year

1 2 3 4 5

Strongly disagree **disagree** **neutral** **agree** **strongly agree**

B26. The courses on SBA system have served me in upgrading my skills on SBA on aspects such as how:

i.	to assess students using the assessment instrument	1	2	3	4	5
ii.	to assess students documentation following stated criteria in the assessment document	1	2	3	4	5
iii	to explain assessment criteria in details to students	1	2	3	4	5
iv	to conduct assessment activities effectively	1	2	3	4	5

B27. Please evaluate the following questions relate to the In-house training (IHT) on SBA at your school.

i.	IHT is conducted by the experts from the ministry or the panels of teachers	1	2	3	4	5
ii.	IHT has been conducted more than once to increase teachers' understanding	1	2	3	4	5
iii	Training includes practical forms not only theoretical	1	2	3	4	5

B28. Please give your views on administration in implementing SBA at your school. Administrators:

i.	Give their support	1	2	3	4		5
ii.	Provide reinforcement such as giving gifts or praise to those teachers that practice SBA in their classrooms						
iii	Give special recognition to teachers performance for conducting SBA	1	2	3	4		5

B29. Please give your views on the coordination process run at your school. Each teacher:

i.	prepares students' scores for his/her subjects	1	2	3	4	5

ii.	prepares evidence according to students' achievement	1	2	3	4	5
iii	records the marks of evidence in SPPBS application	1	2	3	4	5

B30. Please give your views on the monitoring process run at your school.

i.	The school SBA committee appoints certain teachers as internal monitors	1	2	3	4	5
ii.	Heads of panel of each subject monitor SBA activity implementation	1	2	3	4	5
iii	Evidence of centre assessment is kept in a systematic way in a safe place before it is returned to the students	1	2	3	4	5

SECTION C:

C31. Here are some possible challenges faced by the teachers in implementing SBA in schools. Please indicate your level of agreement for each item by circling the chosen number:

1	Difficult to gain support from the headteacher	1	2	3	4	5
2	Difficult to gain cooperation from teachers	1	2	3	4	5
3	Insufficient knowledge to implement SBA process	1	2	3	4	5
4	Lack of confidence to conduct SBA	1	2	3	4	5
5	Problem to get related resources on SBA	1	2	3	4	5
6	School climate seems to hinder SBA implementation	1	2	3	4	5
7	Insufficient financial resources	1	2	3	4	5
8	Extra workload	1	2	3	4	5
9	No special recognition for teachers performance for conducting SBA	1	2	3	4	5
10	Students and parents may not trust teachers' assessment in SBA	1	2	3	4	5
11	Students may memorize responses and reproduce them during SBA sessions	1	2	3	4	5
12	The existing number of teachers is not sufficient to implement SBA activities	1	2	3	4	5

SECTION D:

Please indicate your level of agreement for each item by circling the chosen number.

1	2	3	4	5
Strongly disagree	**disagree**	**neutral**	**agree**	**strongly agree**

D32. In my school, the implementation of SBA

i.	Reduces absentee rate	1	2	3	4	5
ii.	Reduces pressure in public examinations	1	2	3	4	5
iii.	Improves literacy and numeracy learning	1	2	3	4	5
iv.	Provides new learning opportunities for building on students' strengths and interests	1	2	3	4	5
v.	Increases public satisfaction and confidence towards school					
vi.	It contributes strongly to the quality assurance process					

D33. The implementation of SBA is important in that it may

i.	Help students improve generic skills to build on their strengths and interests	1	2	3	4	5
ii.	Help teachers to improve students assessment practices	1	2	3	4	5
iii.	Help teachers to actively develop their professionalism	1	2	3	4	5
iv.	Help teachers to develop students various learning opportunities	1	2	3	4	5
v.	Make teachers and students willing to discuss learning problems in non-threatening manners	1	2	3	4	5
vi.	Make teachers willing to take decisions on assessment reform	1	2	3	4	5

SECTION E:

Circle the number that is reflecting the extent to which the students have mastered the given statement.

Have in your mind that:

1	2	3	4	5
Strongly disagree	**disagree**	**neutral**	**agree**	**strongly agree**

E34. Students' attitude towards SBA:

i.	Students practice SBA in their study	1	2	3	4	5
ii.	The students are motivated to prepare for SBA.	1	2	3	4	5
Iii	Practising SBA in their study is NOT burdening them at all	1	2	3	4	5

E35. Students' knowledge of SBA. Students know that:

i.	SBA tasks are part of teaching and learning of their respective subjects	1	2	3	4	5
ii.	Feedback from teachers helps them develop skills that may not be reflected in public examinations	1	2	3	4	5

E36. Students' motivational source:

i.	SBA encourages students to read more books than before.	1	2	3	4	5
ii.	The students are becoming more interested in my subject than before	1	2	3	4	5
Iii	SBA helps students to understand more on their strengths and weakness in each subject.	1	2	3	4	5

SECTION F: DEMOGRAPHIC INFORMATION

Instruction: Please put a *(√)* in the box for each question and answer the questions.

1. School Grade: i. Grade A ☐ ii. Grade B ☐
2. School Location: i. Urban ☐ ii. Rural ☐
3. School Category: i. Secondary School ☐ ii. Primary School ☐
4. Gender: i. Female ☐ ii. Male ☐
5. Race: i. Malay ☐ ii. Chinese ☐
   iii. Indian ☐ iv. Others ☐
6. Age: ☐☐ years
7. Teaching experience: ☐☐ years
9. Teacher's experience in practicing SBA: ☐☐ years
10. Attending any courses/training on SBA: ☐☐ times
11. Do you teach your own option? Yes ☐ No ☐
12. The main subjects taught: _____
12. Highest academic achievement:
    iii. Diploma ☐ ii. Degree ☐
    iii. Master ☐ iv. PhD ☐

THANK YOU

Glossary

List of References

Abell, S. K. and Lederman, G.L. (2007) *Handbook of Research in Science Education*. Mahwah, NJ: Lawrence Erlbaum.

Adediwura, A. A. (2012) 'Teachers' perception of SBA in Nigerian Secondary Schools', *Mediterranean Journal of Social Sciences*, 3(1), pp. 99-109.

Adnan, A., Zamri, M., Nik Mohd Rahimi, N. Y. and Mohamad, A. E. (2009) 'Penyelidikan dalam Model Pentaksiran Berasaskan Sekolah dalam Matapelajaran Bahasa Arab' *Prosiding Wacana Pendidikan Islam Siri ke-7: Peringkat Serantau,* Bangi, Universiti Kebangsaan Malaysia. Available at: http://ewarga2.ukm.my /academiccv/viewvitae.cfm?per= K009689 &dept=GGGP (Accessed: 12 November 2012).

Agrawal, M. (2004) 'Curricular reform in schools: the importance of evaluation', *J. Curriculum Studies,* 36(3), pp. 361-379.

Aishah (2014) *Dialog Rakyat*. Available at: http://dialograkyat.blogspot.co.uk /2014 /03/kemerosotan-pemikiran-di-malaysia.html (Accessed: 5 November 2013).

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

Akbari, R. and Hosseini, K. (2008) 'Multiple intelligence and language learning strategies: Investigating possible relations', *Elsevier,* 36(2), pp. 141-155.

Akpunar, B. and Dogan, Y. (2011) 'Deciphering the theory of multiple intelligences: An Islamic perspective', *International Journal of Business and Social Science*, 2(11), pp. 224-231.

Alkin, M. C. (1969) 'Evaluating theory development', *Evaluation Comment,* 2, pp. 2-7.

Alzaghoul, A. F. (2011) 'The Implication of the Learning Theories on Implementing E-Learning Courses', *The Research Bulletin of Jordan ACM*, 11(11), pp. 27-30.

Anderson, K. D. (2000) 'School Improvement and School Reform in Canada, Whose perspective is it?' *International Conference Proceedings Brdo pri Kranju, Slovenia,* College of Management in Koper, 8-10 April 2000. Available at: http://www.fm-kp.si/zalozba/ISBN/961-6268-47-3/069-085.pdf (Accessed: 20 November 2013).

Anderson, J. C. and Gerbing, D. W. (1988) 'Structural equation modelling in practice: A review and recommended two-step approach', *Psychological Bulletin,* 103(3), pp. 411-423.

Angelo, T. A. and Cross, K. P. (1993) *Classroom assessment techniques: a handbook for college teachers.* San Francisco, Jossey-Bass: A Wiley Imprint.

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

Anton, D. G. and Jordan, P. N. (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).

ARG (1999) *Assessment for learning: Beyond the black box*. Cambridge: University of Cambridge School of Education.

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

Armstrong, A. and Francis, R. (2003) 'Difficulties in evaluating crime prevention programs: What are some lessons for evaluators of community-based programs', *Paper presented at the Evaluation in Crime and Justice: Trends and Methods Conference convened by the Australian Institute of Criminology in conjunction with the Australian Bureau of Statistics,* Canberra, 24-25 March 2003. Available at: http://www.aic.gov.au/media_library / conferences/evaluation/francis.pdf (Accessed: 5 November 2014).

Assessment Reform Group (2002) *Testing, motivation and Learning.* Cambridge. Available at: http://assessmentreformgroup.files.wordpress .com/2012/01/tml.pdf (Accessed: 8 July 2013).

Arbuckle, J. L. and Wothke, W. (1999) *Amos 4.0 User's Guide*. SmallWaters Corporation, United States of America.

Armstrong, A. and Francis, R. (2003) 'Difficulties in evaluating crime prevention programs: What are some lessons for evaluators of community-based programme' *Evaluation in Crime and Justice: Trends and Methods Conference convened by the Australian Institute of Criminology in conjunction with the Australian Bureau of Statistics,* Canberra, 24-25 March

2003. Available at: http://www.aic.gov.au/media_library/conferences /evaluation/francis.pdf (Accessed: 11 July 2013).

Armstrong, A. and Ogren, E. (1986) *Evaluation Models and Strategies.* Evaluation and Training and Services Australia: Melbourne.

Arnold, J. and Fonseca, M. C. (2004) 'Multiple intelligence theory and foreign language learning: A brain-based perspective', *International Journal of English Studies*, 4 (1), pp. 119-136

Ash, D. and Levitt, K. (2003) 'Working within the Zone of Proximal Development: Formative Assessment as Professional Development', *Journal of Science Teacher Education,* 14(1), pp. 1-313.

Asli (2012) *16th MALAYSIAN EDUCATION SUMMIT 2012.* Available at: http://www.asli.com.my/?p=past-events/the-16th-malaysia-education-summit (Accessed: 12 August 2013).

Asraf, M. A. (2013) *Beban Tugas Guru Dalam Perlaksanaan PBS Di Sekolah Kebangsaan Bandar dan Luar Bandar.* Master thesis. University Teknologi Malaysia. Available at: http://www.fp.utm.my /ePusatSumber/pdffail/ ptk ghdfwP2 /ASRAFMP111466D2014TTP.pdf (Accessed: 3 July 2014).

Assessment and Reporting Unit (2005) *Current Perspectives on Assessment*. Available at: https://www.eduweb.vic.gov.au/edulibrary/public /teachlearn/ student/assessment_current_per.pdf (Accessed: 3 October 2012).

Assessment Support Material. (2001) *Assessment Support Material VCE: English/ESL*. Australia. Available at: http://www.vcaa.vic.edu.au/vce/studies/ english/EnglishASM.doc (Accessed: 12 November 2011).

Ayeni, A. J. and Adelabu, M. A. (2012) 'Improving learning infrastructure and environment for sustainable quality assurance practice in secondary schools in Ondo State, South-West, Nigeria*', International Journal of Research Studies in Education*, 1(1), pp. 61-68.

Azizi, Y., Jamaluddin, R., Shahrin, H. and Noordin, Y. (2010) *Sejauhmana Model Stufflebeam (KIPP) boleh membantu dalam Penilaian Program Pembelajaran.* Available at: http://eprints.utm.my/2256/1/AziziYahaya_ Sejaumanakah_ Model__Stufflebeam_%28KIPP%29.pdf (Accessed: 2 March 2012).

Azizi, Y., Shahrin, H., Jamaluddin, R. and Yusof, B. (2007) *Menguasai penyelidikan dalam pendidikan: teori, analisis dan interpretasi data.* Kuala Lumpur: PTS Professional.

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

Bell, B. and Cowie, B. (2001a) *Formative Assessment and Science Education*. Dordrecht: Kluwer Academic Publishers.

Bell, B. and Cowie, B. (2001b). *The Characteristics of Formative Assessment in Science Education*. University of Waikato: John Wiley and Sons, Inc.

Bennet, R. E. (2011) 'Formative Assessment: a critical review', *Assessment in Education: Principles, Policy & Practice,* 18(1), pp. 5-25.

BERA Ethical Guidelines (2011) *Ethical Guidelines For Educational Research*. Available at: www.bera.ac.uk/system/files/3/BERA-Ethical-Guidelines-2011.pdf (Accessed: 21 June 2013).

Berita Harian (2010) *Kepadatan pelajar dalam kelas perlu ditangani dengan bijak*. Available at: http://perkhidmatanpelajaran.blogspot.co.uk /2010/05/ kepadatan-pelajar-dalam-kelas-perlu.html (Accessed: 8 August 2014).

Bernama (2010) *NKRA: Improving Student Outcomes*. Available at: http://www. pmo.gov.my/?menu=newslist&news_id=2265&news_cat=63&page=1731&so rt_year=2010&sort_month (Accessed 13 June 2013).

Bernama (2014) *Sinar Harian*. Available at: http://www.sinarharian.com.my /nasional/penambahbaikan-pbs-diterima-golongan-pendidik-1.273118 (Accessed 13 June 2013).

Bernat, E. and Gvozdenko, I. (2005) 'Beliefs about Language Learning: Current Knowledge Pedagogical Implications and New Research Directions', *TESL-EJ*, 9(1), pp. 1-21.

Berryannouncement (2011) *School Improvement Programme (SIP) Initiative*. Available at: http://berryberryeasy.com/2011/09/school-improvement-programme-sip-initiative/ (Accessed: 22 January 2013).

Bertram, D. (2006) *Likert scales*. Available at: http://poincare.matf. bg.ac.rs/~ kristina/topic-dane-likert.pdf (Accessed: 7 January 2014).

Black, P. (1998) *Testing: Friend or Foe? Theory and Practice of Assessment and Testing*. London: Falmer Press.

Black, P. (2004) *The Nature and Value of Formative Assessment For Learning*. King's College London. Available at: http://www.stivoschool.org/L_T /files2/The%20 original%20Black%20and%20Williams%20paper %20on% 20Formative%20Assessment.pdf (Accessed: 19 July 2013).

Black, P. (2007) *Formative Assessment: Promises or problems?* Available at: http://www.mantleoftheexpert.com/studying/articles/Paul%20Black2007.pdf/ (Accessed: 6 July 2013).

Black, P. (2011) 'Assessment: its purposes in pedagogy and the roles of teachers' *ESRC seminar Glasgow,* King's College London, 12 Sept 2011. Available at: http://www.yumpu.com/en/document/view/7599583 /assessment-its-purposes-in-pedagogy-and-the-roles-school-of- (Accessed: 7 November 2013).

Black, P and Jones, J. (2006) 'Formative assessment and the learning and teaching of MFL: sharing the language learning road map with the learners', *The Language Learning Journal*, 34(1), pp. 4-9.

Black, P. and Wiliam, D. (1998a) *Inside the Black Box: Raising standards through classroom assessment.* London: School of Education, King's College.

Black, P. and Wiliam, D. (1998b) 'Assessment and classroom learning', *Assessment in Education: Principles, Policy and Practice*, 5(1), pp. 7-74.

Black, P. and Wiliam, D. (2004) 'Classroom Assessment Is Not (Necessarily) Formative Assessment (and Vice-versa)', *Yearbook of the National Society for the Study of Education*, 103, pp. 183-188.

Black, P. and Wiliam, D. (2009) 'Developing the theory of formative assessment', *Educational Assessment, Evaluation and Accountability*, 21, pp.5-31.

Black, P., Harrison, C., Hodgen, J., Marshall, B. and Serret, N. (2011) 'Can teachers' summative assessments produce dependable results and also enhance classroom learning?', *Assessment in Education: Principles, Policy and Practice,* 18(4), pp. 451-469.

Black, P., Harrison C. L., Marshall, B. and Wiliam, D. (2002) *Working inside the black box: Assessment for learning in the classroom.* London: School of Education, King's College.

Boon, Y. and Shaharuddin, M. (2011) *Kepemimpinan Guru Besar Dalam Pelaksanaan Pentaksiran Berasaskan Sekolah (PBS) Di Sekolah Kebangsaan Daerah Kota Tinggi, Johor.* Fakulti Pendidikan. Universiti Teknologi Malaysia. Available at: http://eprints.utm.my/11929/1/Kepemimpinan _Guru_Besar_ Dalam_Pelaksanaan_Pentaksiran_Berasaskan_Sekolah.pdf (Accessed: 8 March 2012).

Borneo Post (2012) *Infrastruktur sekolah luar bandar tidak berubah.* Available at: http://www.theborneopost .com/2013/11/21/infrastruktur-fizikal-sekolah-luar-bandar-tidak-berubah/ (Accessed: 7 November 2014).

Boud, D. (2000) 'Sustainable assessment: rethinking assessment for the learning society', *Studies in Continuing Education*, 22(2), pp. 151-167.

Boyd, R. D. and Apps, J. W. (1980) *Redefining the Discipline of Adult Education*. San Francisco: Jossey-Bass.

BPK (Bahagian Pembangunan Kurikulum, KPM) (2005) *Format Pentaksiran Tahap Dua*. Available at: http://www.scribd.com/doc/15651833 /Panduan-Pelaksanaan-PKSR-Tahun-5 (Accessed: 3 March 2012).

Branden, K. V. D. (2012) 'Sustainable education: basic principles and strategic recommendations', *School Effectiveness and School Improvement*, 23(3), pp. 285-304.

Brannen, J. (2005) *Mixed method research: A discussion paper*. ESRC National Center for Research Methods: Institute of Education, University of London. Available at: http://eprints.ncrm.ac.uk/89/1/MethodsReviewPaperNCRM-005.pdf (Accessed: 5 March 2012).

Brinkerhoff, R. D., Brethower, D. M., Hluchyj, T. and Nowakowski, J. R. (1983) *Program Evaluation: A practitioner's guide for trainers and educator.* Boston: Kluwer Academic Publishers.

Brooks, M. L. (2002) *Drawing to learn*. Unpublished PhD thesis. University of Alberta, Canada.

Brown, G. T. L., Harris, L. R. and Harnett, J. (2012) 'Teacher beliefs about feedback within an AfL environment: Endorsement of an improved learning over student well-being', *Teaching and Teacher Education*, [Online] DOI:10.1016/j.tate.2012.05.003. (Accessed: 9 January 2010).

Brown, S. (2004) Assessment for Learning, *Learning and Teaching in Higher Education*, Issue 1, pp. 81-89.

Brown, T. (2006) *Confirmatory factor analysis for applied research.* USA: The Guidford Press.

Burns, J. (2012) *BBC News*. Available at: http://www.bbc.co.uk/news/ education -17481888 (Accessed: 5 April 2013).

Butler, D. L. and Winnie, P. H. (1995) 'Feedback and self-regulated learning: a theoretical synthesis', *Review of Educational Research,* 65(3), pp. 245-281.

Brualdi, A. C. (1996) *Pareonline*. Available at: http://pareonline.net/getvn. asp?v=5&n=10 (Accessed: 9 January 2010).

Bryman, A. (2001) *Social Research Methods*. Oxford: Oxford University Press.

Buhagiar, M. A. and Murphy, R. (2008) 'Teachers' assessments of students' learning of mathematics', *Assessment in Education: Principles, Policy & Practice*, 15(2), pp.169-182.

Byrne, B. M. (2010) *Structural Equation Modelling with AMOS. Basic concepts, applications and programming.* Mahwah: New Jersey: Earlbum.

Caffrey, E. D. (2009) *Assessment in Elementary and Secondary Education: A Primer, Congressional Research Service.* Available at: http://www.fas .org /sgp/crs/misc/R40514.pdf (Accessed: 5 April 2013).

Cauley, M. C. and McMillan, J. H. (2010) 'Formative Assessment techniques to support student motivation and achievement. Clearing House'*, AJ. Edu. Strat, Issues and Ideas*, 83(1), pp. 1-6.

CERI (Centre for Educational Research and Innovation) (2008) *Assessment for Learning Formative Assessment.* Available at: http://www.oecd.org/dataoecd /19/31/40600533.pdf (Accessed: 9 March 2012).

Chan, Y. F. (2008) 'Effective Training for School Improvement: A Case Study in Malaysia' *3rd International Conference on Principal ship and School Management*, University of Malaya, 10-13 March 2008. [Online]. Available at: http://ikp.um.edu.my/images/ipk/doc/Chan%20Yuen%20Fook.pdf (Accessed: 23 August 2013).

Chan Y. F. and Gurnam K. S. (2012) 'School-based Assessment Among ESL Teachers in Malaysian Secondary Schools', *Journal of the Malaysian Education Deans' Council*, Vol. 9, pp. 1-18.

Chan, Y. F., Sidhu, G. K. and Yunus, R. (2006) University Teknologi Mara Intitutional Repository. Available at: http://eprints.uitm.edu.my/3181/ (Accessed: 23 August 2013).

Chapman, C. (2011) 'School improvement research and practice', in Chapman, C. (ed.) *School Effectiveness and Improvement Research, Policy and Practice, Challenging the orthodoxy?.* New York, Taylor and Francis: Routledge.

Che Noraini, H., Adlina, A. and Nurhidayah, M. H. (2013) '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.Available at: http://www.iium.edu.my/medc2013 /documents/ProsidingMEDC.pdf (Accessed: 8 October 2014).

Cheah, U. (2010) 'Assessment in primary Mathematics Classrooms in Malaysia' *Tsukuba International Conference: Innovation of Mathematics Teaching and Learning through Lesson Study - Connection between Assessment and Subject Matter.* Tsukuba, 17-21 February. Available at: http://www.criced.tsukuba.ac.jp/ math/apec/apec2009/doc/pdf_20-21/CheahUiHock-paper.pdf (Accessed: 11 November 2011).

Chen, C. (2009) *A case study in the evaluation of English training courses using a version of the CIPP model as an evaluation tool, Durham theses.* PhD thesis. Durham University. [Online]. Available at: http://etheses.dur.ac.uk /2912/1 /2912_742.pdf (Accessed: 14 November 2013).

Chen, H. T. (1990) *Theory-driven Evaluations.* London: SAGE Publications.

Cheng, L., Andrews, S. and Yu, Y. (2011) 'Impact and consequences of SBA: students' and parents' views of SBA in Hong Kong', *Language Testing*, 28(2), pp. 221-250.

Cheung, D. and Yip, D. (2003) 'School-based assessment of chemistry practical work: Exploring some directions for improvement', *Education Journal*, 31(1), pp. 133-152.

Children England (2011) *School Reforms. Department for Education, United Kingdom*. Available at: http://www.childrenengland.org.uk/upload /OAP%20Schools%20Briefing.pdf (Accessed: 20 December 2012).

Chin, W. W. (1998) 'Issues and opinion on structural equation modelling', *MIS Quarterly*, 22(1), pp. 7-16.

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, pp. 18-29.

Chua, Y. P. (2006) *Kaedah dan Statistik Penyelidikan: Kaedah Pendidikan*. Kuala Lumpur: McGraw Hill.

Chun, L. Y. (2006) 'Practice and Challenges of School-based Formative Assessment' *IAEA, Singapore 32nd Annual Conference*. Department of Curriculum & Instruction, The Hong Kong Institute of Education (n.d). Available at: http://www.iaea.info/documents/paper_1162a1dcb9.pdf (Accessed: 22 August 2013).

Clark, I. (2011) 'Formative Assessment and motivation: Theories and themes', *Prime Research on Education (PRE),* 1(2), pp. 27-36.

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

CMEC (2005) *OECD Study on Enhancing Learning through Formative Assessment and the Expansion of Teacher Repertoires. Canadian Report.* Available at: http://www.cmec.ca/publications/lists/publications /attachments/78/oecd_formative.en.pdf (Accessed: 9 August 2014).

Cohen, J. and Cohen, P. (1983) *Applied multiple regression/correlation analysis for the behavioral science.* 2nd edn. Hillsdale: NJL Lawrence Erlbaum Associates, Inc.

Cohen, L. and Manion, L. (1994) Research Methods in Education, 4th ed., London: Routledge Publishers.

Cohen, L., Manion, L. and Morrison, K. (2007) *Research Methods in Education.* 6th edn. London: Routledge.

Communication of Learning and UNESCO. (2004) *Case Study Malaysia*. Available at: http://portal.unesco.org/education/en/files/30973/10848 83345 1Malaysia_Case_Study.pdf/Malaysia%2BCase%2BStudy.pdf (Accessed: 7 November 2013).

Corner, J. (1991) 'In search of more complete answers to research questions. Quantitative versus qualitative methods: is there a way forward', *Journal of Advanced Nursing*, 16, pp. 718-727.

Cowie, B. and Bell, B. (1999) 'A Model of Formative Assessment in Science Education', *Assessment in Education: Principles, Policy & Practice*, 6(1), pp. 101-116.

Creswell, J. W. (2002) *Educational Research: planning, conducting and evaluating quantitative and qualitative research*. New Jersey: Pearson Education, Inc.

Creswell, J. W. (2003) *Research Design, Qualitative, Quantitative and Mixed Methods Approaches.* London: SAGE Publications.

Cresswell, J. W. and Garrett, A. L. (2008) 'The movement of mixed methods research and the role of educators', *South African Journal of Education,* Vol 28, pp. 321-333.

Cronbach, L. J. (1990) *Essentials of PsychologicalTesting*. 5th edn. New York: Harper & Row.

Cronbach, L. J. (1963) 'Course improvement through evaluation', *Teacher College Record*, 64, pp. 672-683.

Crossouard, B. and Pryor, J. (2012) 'How Theory Matters: Formative Assessment Theory and Practices and Their Different Relations to Education', *Stud Philos Educ*, 31, pp. 251-263.

CSEC (2010) Caribbean Examinations Council. Available at: https://www.cxc. org/SiteAssets/MANUALS/SBA_Manual_2012.pdf (Accessed: 3 May 2013).

Curran, P. J., West, S. G. and Finch, J. F. (1996) 'The robustness of test statistics to nonnormality and specification error in CFA', *Psychological Methods*, 1(1), pp.16-29.

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

Darling-Hammond, L. and McCloskey, L. (2008) 'Assessment for Learning Around the World: What Would it Mean to be Internationally Competetive?', *Phi Delta Kappan*, 90(4), pp. 263-272.

Darling-Hammond, L. and Pecheone, R. (2010) *National Conference on Next Generation Assessment Systems, Stanford University*. Available at: http://www. k12center.org/rsc/pdf/Darling-HammondPechone System Model.pdf (Accessed: 3 May 2013).

Darmody, M., Smyth, E. and McCoy, S. (2007) *Working Paper, The Economic and Social Research Institute (ESRI), Dublin, No. 212.* Available at: http://www.econstor.eu/bitstream/10419/67998/1/548400237.pdf (Accessed: 9 May 2013).

David, N. and Debra, M. (2004). *Rethinking Formative Assessment in HE: a theoretical model and seven principles of good feedback practice.* Available at: http://www.heacademy.ac.uk/assessment/ASS051D_SENLEF_model.doc (Accessed: 19 November 2013).

Davison, C. (2004) 'The contradictory culture of teacher-based assessment: ESL teacher assessment practices in Australian and Hong Kong secondary schools', *Language Testing,* 21, pp. 305-354. Available at: http://ltj.sagepub .com/content/21/3/305. (Accessed Nov 2011).

Davison, C. (2007) 'Views from the chalk face: SBA in Hong Kong', *Language Assessment Quarterly*, 4(1), pp. 37-68.

Deakin-Crick R., Lawson, H., Sebba J., Harlen W., and Yu, G. (2005) *EPPI-Centre, Protocol.* Available at: http://eppi.ioe. ac.uk/eppiwebcontent/reel /review_ groups /assessment/assessment_protocol5.pdf (Accessed: 1 November 2013).

Delanshere, G. G and Jones, J. H. J. (1999) 'Elementary teachers' belief about assessment in mathematics: A case of assessment paralyses', *Journal of Curriculum and Supervision*, 14(3), pp. 216-240.

Department of Education and Skills (2011) *Literacy and Numeracy for Learning and Life.* Available at: http://www.education.ie /en/Publications /PolicyReports /lit_num_strategy_full.pdf (Accessed: 5 November 2013).

Department of Statistics (2010) *Pembangunan Modal Insan*. Available at: http://www.statistics.gov.my/portal/download_Stats_Malaysia/files/MMS/2010/BM/05_Pembangunan_Modal_Insan.pdf (Accessed: 19 November 2013).

Dodge, J. (2009) *25 Quick Formative Assessment for a Differentiated Classroom*. New York: Scholastic Teaching Resources.

Donaldson, S. I. and Scriven, M. (2008) *Evaluating Social Programs and Problems: Visions for the New Millenium.* Routledge.

Dorman, J. P., Fisher, D. L. and Waldrip B. G. (2006) 'Learning environments, students' perceptions of assessment, academic efficacy and attitudes to science: A LISREL analysis', in D. L. Fisher & M. S. Khine (ed.) *Contemporary approaches to research on learning environments: world views.* Singapore: World Scientific, pp. 1-28.

Durgahee, T. (1998) *The National Mission 2006-2020*. Available at: http:// www3.pmo.gov.my/RancanganWeb/Rancangan1.nsf/MisiNasionalEng?open Form (Accessed: 4 November 2013).

Drummond, M. J. (2003) *Assessing children's learning*. 2nd edn. London: David Fulton.

Earl, L. (2004) *Assessment As Learning: Using classroom achievement to maximize student learning. Experts in Assessment.* Carlifornia: Corwin Press Inc. Thousand Oaks.

Economic Planning Unit. (2010) *Chapter 5, Developing and retaining a first-world talent base.* Available at: http://www.moe.gov.my/userfiles/file/ RMK10 _Education1.pdf (Accessed: 9 August 2013].

Eiselen, R. and Uys, T. (2005) *Questionnaire Design.* Available at: http://www. uj.ac.za /EN /postgrad/ pgh/Statkon /Documents/Statkon% 20Questionaire %20Design.pdf (Accessed: 3 August 2013).

Elliot, W. E. and Michael, D. D. (2008) *Handbook of Research and Policy in Art Education*. New Jersey: Lawrence Erlbawn Assoc.

Emisportal (2007) *Sekolah, kelas dan bilik darjah, Bab 2.* Available at: http://emisportal.moe.gov.my/emis/emis2/emisportal2/doc/fckeditor/File/BukuPerangkaan07/05_Perangkaan%2007_Chap%202_95-104_pdf.pdf (Accessed: 8 December 2012).

Emisportal (2013) *Maklumat Asas Pendidikan, Portal Sistem Maklumat Pengurusan Pendidikan.* Available at: http://emisportal.moe.gov.my /emis /emis2/ emisportal2/mainpage. php?module= Maklumat&kategori=47 (Accessed: 7 November 2013).

Emisportal (2014) *Maklumat Asas Pendidikan, Portal Sistem Maklumat Pengurusan Pendidikan.* Available at: http://emisportal.moe.gov.my /emis /emis2/ emisportal2/mainpage. php?module= Maklumat&kategori=47 (Accessed: 8 November 2014).

Entwistle, N. J. and Nisbet, D. J. (1972) *Educational Research in Action*. London: University of London Press Ltd.

EPRD (2013) *Quick Facts 2013, Malaysian Educational Statistics*. Available at: http://emisportal.moe.gov.my/emis/emis2/emisportal2/doc/fckeditor/File/Quickfacts_2013/quickfacts2013.pdf?PHPSESSID=dfc6abfe0c9b71c8405f8ed8f9812988 (Accessed: 11 November 2013).

Epu (2004) *Malaysian Quality of Life*. Available at: http://www.epu.gov.my/c/document_library/get_file?uuid=1680adba-7ddf-4dcc-8a97-5e420f3c988b&groupId=283545 (Accessed: 1 November 2013).

Eseryel, D. (2002) 'Approaches to Evaluation of Training: Theory and Practice', *Educational Technology & Society*, 5(2).

ESRC (2012) *Research Ethics Framework*. Available at: http://www.gold.ac.uk/media/ESRC_Re_Ethics_Frame_tcm6-11291.pdf (Accessed: 3 August 2013).

Esther, S. H. (2012) *Student Learning Assessment, Asia Pacific Education System Review Series, UNESCO Bangkok.* Available at: http://unesdoc.unesco.org/images/0021/002178/217816E.pdf (Accessed: 20 April 2013).

Fairfax County Public Schools (2012) *Literature Review on Formative Assessment.* Available at: http://www.fcps.edu/pla/ope/docs/final_ docs_ for_ posting/ecart _appendices.pdf (Accessed: 3 February 2015).

Faizah, A. M. (2011) 'School-based assessment in Malaysian Schools: The Concerns of the English teachers', *Journal of US-Chine Education Review*, 8(10), pp. 1-15.

Fan, L. (2011) *Performance Assessment in Mathematics: Concepts, Methods, and Examples from Research and Practice in Singapore Classrooms*. Singapore: SG, Pearson.

Feng, H. (2007) *Senior ESOL Students' Experiences of and attitudes towards Formative Assessment in Mainstream Secondary Classrooms.*Available at:http://ir.canterbury .ac.nz/bitstream /10092/1058/1 /thesis_ fulltext.pdf (Accessed: 3 November 2014).

Fishbein, M., and Ajzen, I. (1975) *Belief, attitude, intention, and behavior.* Reading, MA: Addison-Wesley.

Fives, H. and Gill, M. G. (2014) *International Handbook of Research in Teacher Belief.* New York: Routledge.

Fleishman, H. L. and Williams, L. (1996) *An Introduction to Program Evaluation for Classroom Teachers. Development Associates, Inc.* Available at: :http://teacherpathfinder . org/School/Assess/assess.html. (Accessed: 22 April 2013).

Frye, A. W. and Hemmer, P. A. (2012) *Medical Teacher*. Available at: http://www .med.unc.edu/aoe/aoe-events/videos-and-materials-from-curricular-sessions/materials-from-curricular-sessions-2014-2015/program-evaluation-guide (Accessed: 2 January 2015).

Fullan, M. (2000) *The Role of the Head in School Improvement. National College of School Leadership, England.* Available at: http://www.michaelfullan.ca /media/13396043200.pdf (Accessed: 2 January 2014).

Fullan, M. (2001) *Whole School Reform: Problems and Promises. Ontario Institute for Studies in Education, Chicago Community Trust.* Available at: http://www.michaelfullan.ca/media/13396044810.pdf (Accessed: 3 January 2014).

Fullan, M. (2002) *Principals as Leaders in a Culture of Change.* Available at: http://www.michaelfullan.ca/media/13396053050.pdf (Accessed: 3 January 2014).

Fullan, M. (2005) *The Tri-Level Solution, School/District/State Synergy. Education Analyst - Society for the Advancement of Excellence in Education.* Available at: http://www.michaelfullan.com/media/13396062320.pdf (Accessed: 17 April 2013).

Fullan, M. (2006) *Change theory, A Force for School Improvement, CSE, Seminar Series Paper No 157, Victoria.* Available at: http://www. michaelfullan. ca/media/13396072630.pdf (Accessed: 4 January 2014).

Fullan, M. (2007) *The New Meaning of Educational Change*. New York: Teachers College Press.

Fullan, M. (2009a) 'Large-scale reform comes of age', *The Journal of Educational Change*, 10, pp. 101-113.

Fullan, M. (2009b) *The Challenge of Change: Start School Improvement Now!,* 2nd ed. London: CA Corwin Press,

Fullan, M. (2010a) 'The Big Ideas Behind Whole System Reform', *Education Canada*, 50(3), pp. 24-27.

Fullan, M. (2010b) *All Systems Go: The Change Imperative for Whole System Reform.* Carlifornia: Corwin Press.

Fullan, M. (2010c) 'The Role of the District in Tri Level Reform', in Penelope, P., Eva, B., and McGaw, B. (ed.) *International Encyclopedia of Education.* Oxford: Elsevier, pp. 295-302.

Fullan, M. (2011) *Choosing the wrong drivers for whole system reform*. Available at: http://www.michaelfullan.ca/media/13501655630.pdf (Accessed: 3 December 2013).

Fullan, M. and Levin, B. (2009) *The Fundamentals of Whole-System Reform, A Case Study From Canada.* Available at: http://www.michaelfullan .com/media/13396080290.pdf (Accessed: 9 December 2012).

Fullan, M., Rolheiser, C., Mascall, B. and Edge, K. (2001) *Accomplishing Large Scale Reform: A Tri-Level Proposition, University of Toronto.* Available at: http://www.michaelfullan.ca/media/13396045990.pdf (Accessed: 4 December 2012).

Fynewever (2008) *Calvin.* Available at: http://www. calvin.edu/academic/ chemistry/research/fynewever/fynewever-research.html (Accessed: 4 December 2014).

Galloway, A. (2008) *The Assessment Systems of Finland and Queensland*, Available at: http://www.sqa.org.uk/files_ccc/ PNP_ResearchReport1 _The Assessment SystemsFinlandQueensland.pdf (Accessed: 6 December 2014).

Galvin, J. C. (1983) 'What can trainers learn from educators about evaluating management training?', *Training and Development Journal*, 37(8), pp. 52-57.

Gardner, H. (1983) *Frames of Mind*. New York: Basic Books.

Gardner, J. (2006) *Assessment and Learning*. London: Sage Publications.

Gardner, J. (2007) 'Is teaching a 'partial' profession?', *Making the grade*, Summer: 18-21.

Gedye, S. (2010) *Formative assessment and feedback: a review. University of Plymouth. Planet Issue 23.* Available at: http://www.gees .ac.uk/planet/p23 /Gedye.pdf (Accessed: 12 December 2012).

Ginsburg, H. P. (2009) 'The Challenge of Formative Assessment in Mathematics Education: Children's Minds, Teacher's Minds', *Human Development*, 52, pp. 109-128.

Glatthorn A. A., Boschee, F., Whitehead, B. M. and Boschee, B. F. (2012) *Curriculum Leadership: Strategies for Development and Implementation. Thousand Oaks, California.* Available at: http://www.sagepub.com/upm-data/44333_12.pdf (Accessed: 14 November 2013).

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

Grima, G. (2003) 'School-based Assessment: Implementation Issues and Practices', *21st Annual AEAA Conference: Assessment and Certification in a Changing Educational, Economic and Social Context,* Cape Town, South Africa, 25-30 August. Available at: http://www.slideshare .net/HajarHamdan/school-based-assesment (Accessed: 14 November 2013).

GTI Media (2013) *Pre-University, A Level, SAM, IB.* Available at: http://www. doctorjob.com.my/sectors/pre-university/stpm-sijil-tinggi-persekolahan-malaysia-form-6 (Accessed: 5 November 2013).

Guskey, T. R. (2003) 'How Classroom Assessment can improve learning', *Educational Leadership*, 60(5), pp. 7-11.

Hager, P. and Butler, J. (1996) 'Two Models of Educational Assessment', *Assessment and Evaluation in Higher Education,* 21(4), pp. 367-378.

Hailmi, A. M. (2013) 'Luahan hati guru kerana PBS', *CiKGUHAiLMI*, 3 March. Available at: http://www.cikguhailmi.com/2013/03/guru-luahan-hati-guru-kerana-pbs. html (Accessed: 3 August 2013).

Hair, J. F., Anderson, R. E., Tatham, R. L. and Black, W. C. (1998) *Multivariate data analysis.* 5th edn. New Jersey: Prentice-Hall International, Inc.

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E. and Tatham, R. L. (2006) *Multivariate Data Analysis.* 6th edn. New Jersey: Prentice-Hall International, Inc.

Hair, J. F., Black, W. C., Babin, B. J. and Anderson, R. E. (2010) *Multivariate Data Analysis.* 7th edn. Englewood Cliffs, NJ: Prentice Hall.

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

Hamid, Y. and Mahmood, S. (2010) 'Understanding constructive feedback: A commitment between teachers and students for academic and professional development', *J Pak Med Assoc.,* 60(3), pp. 224-227.

Hammond, L., Austin, K., Orcutt, S. and Rosso, J. (2001) *Introduction - How people learn*. Available at: http://web.stanford.edu/class/ed269/hplintro chapter.pdf (Accessed: 9 February 2015).

Hamp-Lyons, L. (2006) *Fairness as an issue in school-based assessment. Faculty of Education, University of Hong Kong*. Available at:sba.edu.hku.hk/ new _sba/ doc/Seminar Series LHL--Fairness as an issue in school-based assessment.ppt (Accessed: 9 February 2012).

Hamsiah, M. D. and Raja Maznah, R. H. (2011) 'Web-based Assessment in Teaching and Learning Management System (e-ATLMS)', *Malaysian Journal of Education Technology, 11 (2),* pp.5-11.

Hamzah, M. O. and Paramasivan, S. (2009) 'Between the ideal and reality: Teachers' perception of the implementation of school-based oral English assessment', *The English Teacher,* XXXVIII: pp. 13-30.

Hargreaves, E. (2005) 'Assessment for learning? Thinking outside the (black) box', *Cambridge Journal of Education*, 35(2), pp. 1-23.

Harlen, W. (2004) A *systematic review of the evidence of the impact on students, teachers and the curriculumof the process of using assessment by teachers for summative purposes,* Research Evidence in Education Library. London: EPPI-Center, Social Science Research Unit, Institute of Education.

Harlen, W. (2005) 'Teachers' summative practices and assessment for learning – tensions and synergies', *The Curriculum Journal*, 16(2), pp. 207-223.

Harrington, D. (2009) *Confirmatory factor analysis.* New York: Oxford University Press.

Harris, A. (2003a) 'Behind The Classroom Door: The Challenge of Organisational and Pedagogical Change', *Journal of Educational Change*, 4, pp. 369-382.

Harris, A. (2003b) *School Improvement, What's in it for schools?.* Routledge Falmer: London.

Harris, A. (2010) 'Leading System Transformation', *International Congress for School Effectiveness and School Improvement (ICSEI),* pp. 1-18.

Harris, A. and Muijs, D. (2003) *National College For School Leadership*. Available at: http://dera.ioe.ac.uk/5132/1/download %3Fid%3D17417% 26 filename%3Dteacher-leadership-principles-practice-full-report.pdf (Accessed: 4 December 2013).

Harris, A. and Jones, M. (2010) 'Professional learning communities and system improvement', *Improving Schools*, 13(2), pp. 172-181.

Harris, A., Day, C., Goodall, J., Lindsay, G. and Muijs, D. (2005) *What difference does it make? Evaluating the impact of continuing professional development in schools.* Available at: http://www.almaharris.co.uk/files/cpd-article.pdf (Accessed: 4 December 2013).

Hempenstall, K. (2001) 'School-based reading assessment: Looking for vital signs', *Australian Journal of Learning Disabilities,* 6, pp. 26-35.

Henseler, J. (2010) *Covariance-based SEM: Foundations and Applications*. Available at: http://www.sensometric.org/Resources/Documents/2010/ Meeting/ Presentations/013-000Henseler2010.pdf (Accessed: 7 January 2014).

Heritage, M. (2007a) 'Formative Assessment: What Do Teachers Need to Know and Do?', *Phi Delta Kappan,* 89(2), pp. 140-145.

Heritage, M. (2007b) *Formative Assessment Model. Assessment and Accountability Center (AACC/ CRESST), University of Carlifornia, Los Angeles.* Available at: http://www.gwaea.org/iowacorecurriculum /docs/ Assessment ForLearning_LitReviewFinal.pdf (Accessed: 9 July 2013).

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).

Heritage, M. (2009a) *The process of formative assessment*. Available at: http://www.aea267.k12.ia.us/system/assets /uploads/files/108 /assessment forlearning_brief1.pdf (Accessed: 4 July 2013).

Heritage, M. (2009b) *Understanding formative assessment and utilizing it to improve Classroom instruction.* Available at: http://www.aea267.k12.ia.us /system / assets/uploads/files/108/assessmentforlearning_brief1.pdf (Accessed: 5 July 2013).

Heritage, M. (2010a) *Formative Assessment and Next-Generation Assessment Systems: Are We Losing an Opportunity?.* Council of Chief State School Officers (CCSSO), CRESST: Los Angeles.

Heritage, M. (2010b) *Formative Assessment. CRESST/UCLA, Heartland AEA.* Available at: www.aea11.K12.ia.us/educators/ICC/institute/Heritage.ppt (Accessed: 9 February 2012).

Heritage, M. (2010c) *Formative assessment: Making it happen in the classroom*. Thousand Oaks, CA: Corwin Press.

Herman, J. L. (1997) *The Regents of the University of Carlifornia. CSE Technical Report 446.* Available at: http://www.cse.ucla.edu/products /reports /TECH446.pdf (Accessed: 5 March 2012).

Herman, J. L., Osmundson, E., Ayala, C., Schneider, S. and Timms, M. (2006) *Center for Assessment and Evaluation of Student Learning (CAESL).* Available at: http://www.cse.ucla.edu/products/reports/r703.pdf (Accessed: 2 December 2013).

Hinton, C, Miyamoto, K. and Della-Chiesa, B. (2008) 'Brain Research, Learning and Emotions: implications for education research, policy and practice', *European Journal of Education*, 43(1), pp 87-103. Available at: http://www. udesa.edu.ar/files/programas/NEUROCIENCIAS/Brainresearch_learning_and _emotions_Hinton.pdf (Accessed: 2 March 2012).

HKEAA (2009) *School-based Assessment: Changing the assessment culture.* Available at: http://www. hkeaa.edu.hk/en/sba/ (Accessed: 6 March 2012).

HKEAA (2012a) *Information on School-based Assessment. Hong Kong.* Available at: www.hkeaa.edu.hk/en/sba/sba_hkdse (Accessed: 9 March 2012).

HKEAA (2012b) *HHDSE English Language Examination. Introduction to the SBA Component. Faculty of Education, The University of Hong Kong.* Available at: http://www.hkeaa .edu.hk/ DocLibrary/SBA/HKDSE/Eng_DVD /sba _practice.ht (Accessed: 7 March 2012).

HKEAA (2013) *Hong Kong Diploma of Secondary Education Examination*. Available at: http://www.hkeaa .edu.hk/DocLibrary/SBA/HKDSE/ SBAhandbook-2013-ENG-311210.pdf (Accessed: 9 March 2012).

HM Government (2010) *The Coalition: Our program for government. Cabinet Office, London*. Available at: http://www.direct.gov.uk/prod_consum_ dg/groups/dg_digitalassets/@dg/@en/documents/digitalasset/dg_187876. pdf (Accessed: 3 March 2012).

Hobby, R. (2004) *A Culture for Learning. The Hay Group Management Limited*. Available at: http://www.transforminglearning.co.uk/homepage/Culture _for_Learning_Report.pdf (Accessed: 19 March 2012).

Hoe, S. L. (2008) 'Issues and Procedures in Adopting Structural Equation Modelling Technique', *Journal of Applied Quantitative Methods*, 3(1), pp.76-83.

Hogan, R. L. (2007) 'The Historical Development of Program Evaluation: Exploring the past and present', *Online Journal of Workforce Education and Development*, 11(4).

Hooper, D., Coughlan, J. and Mullen, M. R. (2008) 'Structural Equation Modelling: Guidelines for Determining Model Fit', *The Electronic Journal of Business Research Methods* 6(1), pp. 53-60.

Hopkins, D. (2005) *Every school a great school. IARTV Seminar Series Paper No 146.* Victoria, Australia: IARTV.

Hopkins, D. (2006) *Realizing the Potential of System Reform. Companion in Education Series, Routledge or Falmer*. Available at: http://www.davidhopkins .co.uk/articles/systemreform.pdf (Accessed: 8 December 2012).

Hopkins, D. (2009) *The changing landscape of school improvement. School leadership Today.* Available at: library.teachingtimes.com/.../changing-landscape-school-improveme... (Accessed: 4 March 2012).

Hopkins, D. and Levin, B. (2000) 'Educational Reform and School Improvement' , *NIRA Review*. pp. 21-26.

Hopkins, D., Ainscow, M. and West, M. (1994) *School Improvement in an Era of Change*. London: Cassell.

Hopkins, D., Harris, A., Stoll, L. and Mackay, T. (2011) *24th International Congress of School Effectiveness and School Improvement.* Available at: http://www.icsei.net/icsei2011/State_of_the_art/State_of_the_art_Session_C .pdf> (Accessed: 9 December 2013).

Horng, J, Teng, C. and Baum, T. (2009) 'Evaluating the quality of undergraduate hospitality, tourism and leisure programmes', *JHLSTE*, 8(1), pp. 37-54.

HPRC (2012) *Research Ethics.* Available at: http://www.southampton.ac.uk /hpgr/current_students/ethics.html (Accessed: 9 April 2013).

Human Rights Commission of Malaysia (2006) *Suhakam's Report on the Human Rights Approach to the Millennium Development Goals*. Available at: http://www.suhakam.org.my/c/document_library /get_file?p_l_id= 24206& folderId=30507&name=DLFE-2116.pdf (Accessed: 9 November 2011).

Humphries, M. (2013) *Missing Data & How to Deal: An overview of missing data*. Available at: http://www.texaslonghornsl.com/cola/centers/prc /_files /cs /Missing-Data.pdf (Accessed: 8 October 2014).

Hunt, E. and Pellegrino, J. W. (2002) *Issues, Examples and Challenges in Formative Assessment*, New Directions For Teaching and Learning, no 89.

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).

Iberson, H. (2012) 'An investigation into learners' and teachers' attitudes towards learners' self-assessment according to CEFR1 scales', *Language at the University of Essex (LangUE) 2011 Proceedings,* University of Essex, United Kingdom. Available at: https://www.essex.ac.uk/langling/documents /langue/langue_2011_proceedings.pdf (Accessed: 4 October 2014).

Ibnuabbas (2013) 'Luahan hati seorang guru di sekolah Kementerian Pelajaran Malaysia', *TKO Tranungkite*, 27 Februari. Available at: http://www.tranungkite . net/v12/ modules.php? name= News&file=article&sid =4047 (Accessed: 5 August 2013).

Ibrahim, D. M., Nakajjo, A. and Isoke, D. (2008) *Socioeconomic Determinants of Primary Schools Dropout: The Logistic Model Analysis*. Available at: http:// www. eprc.or.ug/pdf_files/researchseries/series54.pdf (Accessed: 9 November 2011).

Ifeoma, O. E. (2010) 'Roles and Actions of School Principals in Managing Curricular Reforms in Nigeria', *International Journal of Educational Research and Technology*, 1(1), pp. 85-90.

IKLIK (2014) *Isu Perlaksanaan Sistem Aplikasi PBS.* Available at: http://pmr. penerangan.gov.my/index.php/component/content/article/16-isu-nasional/16994-isu-pelaksanaan-sistem-aplikasi-pentaksiran-berasaskan-sekolah-pbs.html (Accessed: 7 October 2014).

Indiana University (2006) *Research Analytics. University Information Technology Services.* Available at: http://www.indiana.edu/-statmath /stat/all/cfa/cfa1.html (Accessed: 25 November 2011).

IowaCORE (2011) *Literature Review Assessment For Learning*. Available at: http://www.gwaea.org/iowacorecurriculum/docs/AssessmentForLearning_LitReviewFinal.pdf (Accessed: 25 November 2014).

Isaac, S. and Michael, W. B. (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.

Jackson, S. L. (2003) *Research Methods and Statistics, A Critical Thinking Approach*. USA: Thomson Wadsworth.

Johnson, B. and Onwuegbuzie, A. J. (2004) 'Mixed Methods Research: A Research Paradigm Whose Time Has Come', *Educational Researcher,* 33(7), pp. 14-26.

Joughin, G. and Macdonald, K. (2003) *A model of assessment in higher education institutions. The Higher Education Academy*. Available at: http://www.llas.ac.uk/resourcedownloads/2968/Joughin_and_Macdonald_model_assessment.pdf (Accessed: 24 November 2012).

JPK (2013) *Official Portal Kelantan Education Department*. Available at: http:// jpn. moe .gov.my/jpnkelantan/v2// (Accessed: 22 August 2013).

JPM (2010) *Rancangan Malaysia KeSepuluh*. Available at: http://www.pmo.gov.my/dokumenattached/RMK/RMK10_Mds.pdf (Accessed: 5 November 2014).

JPM (2014) *Sumber statistic rasmi Malaysia*. Available at: http://www.statistics.gov. my/portal/index.php?lang=bm (Accessed: 5 November 2014).

JPN Kelantan (2015) *Portal Rasmi Jabatan Pendidikan Negeri Kelantan*. Available at: http://jpnkelantan.moe.gov.my/v1/ (Accessed: 30 January 2015).

Kahan, B. (2008) *Excerpts from Review of Evaluation Frameworks*. Available at: http://idmbestpractices.ca/pdf/evaluation-frameworks-review.pdf (Accessed: 5 November 2014).

Kementerian Pelajaran Malaysia (2011a) *KSSR. Bahagian Pembangunan Kurikulum*. Available at: http://www.moe.gov.my/bpk/kssr_docs/01 _Konsep_KSSR /Konsep%20KSSR.pdf> (Accessed: 4 November 2011).

Kementerian Pelajaran Malaysia (2011b) *KSPK & KSSR. Bahagian Pembangunan Kurikulum.* Available at: www.ippm.edu.my/KSSR/KSSR.ppt (Accessed: 3 November 2011).

Kementerian Pelajaran Malaysia (2014) *Pautan Tambahan, Arkib Media-Surat kepada pengarang*. Available at: http://www.moe.gov.my/v/arkib-media-surat-kepada-pengarang-view?id=1054&> (Accessed: 7 August 2014).

Kennedy, K. (2013) *High Stakes SBA and Cultural Values: Beyond Issues of Validity*. Available at: http://www.cambridgeassessment.org.uk /images/ 139719 -sba-seminar-papers.pdf (Accessed: 22 November 2013).

Kerlinger, F. N. (1970) *Foundations of Behavioral Research*. New York: Holt, Rinehart and Winston.

King, J. A. (2008) *An Evaluation of a Character Education Program at an Elementary School, Nova Southern University*. Available at: http://www.cpcc. edu /pd /resources-1/doctoral-research-group/dissertations /jking _dissertation.pdf (Accessed: 24 November 2013).

Kirkpatrick, D. L. and Kirkpatrick, J. D. (2006) *Evaluating Training Programs.* 3rd edn. San Francisco: Berrett-Koehler Publishers.

Klenowski, V. (2013) '*School-based Assessment, Standards, teacher judgment and moderation: Messages from research'.* Available at: http://www.cambridge assessment.org.uk/images/139719-sba-seminar-papers.pdf (Accessed: 4 November 2013).

Kline, R. B. (2005) *Principles and practice of SEM*. New York: Guilford Press.

Knapp, M. S., Copland, M. A. and Talbert, J. E. (2003) *Leading For Learning: Reflective Tools for Schools and District Leaders*. Available at: http://depts. washington.edu/ctpmail/PDFs/LforLSummary-02-03.pdf (Accessed: 22 November 2013).

Knowles, M. S., Holton, E. F. and Swanson, R. A. (1998) *The Adult Learner: The definitive classic in adult education and human resource development.* Houston, TX: Gulf Publishing Company.

Koh, K. and Luke, A. (2009) 'Authentic and conventional assessment in Singapore schools: an empirical study of teacher assignments and student work', *Assessment in Education: Principles, Policy and Practice,* 16(3). pp. 1-33.

Koh, K. and Velayutham, R. L. (2009) *Improving teachers' assessment literacy in Singapore schools.* Available at: http://www.nie.edu.sg/ nie_cma/ attachments/topic/14e8a745eaKV/NIE_research_brief_09_002.pdf (Accessed: 12 November 2013).

Kosmo (2013) *Kekurangan guru akan diatasi*. Available at: http://www.kosmo .com.my/kosmo /content.asp?y= 2012&dt= 0619&pub =Kosmo& sec=Negara &pg=ne_03.htm/ (Accessed: 8 December 2013).

Koszalka, T. A. (2000) 'The validation of a measurement instrument: Teachers' attitudes towards the use of web resources in the classroom', *Quarterly Review of Distance Education*, 1(2), pp. 139-144.

KPM (2014) *Dasar-dasar KPM. Soalan Lazim*. Available at: http://www.kpkk. gov. my/pdf /Dasar-dasar%20KPM%20BM%20%28NCM%29.pdf (Accessed: 3 November 2013).

KPPK (2014a) *60 000 guru ijazah 2015*. Available at: http://www.nutp .org/ new/ms/node/673 (Accessed: 8 November 2014).

KPPK (2014b) *Sedia pembantu guru jika kerja sembilan jam*. Available at: http://www.nutp.org/new/ms/node/830 (Accessed: 22 November 2014).

Kumar, R. (2008) *Research Methodology*. New Delhi: APH Publishing Company.

Kumutha, R. and Hamidah, Y. (2014) 'Barriers Teachers Face in Integrating ICT During English Lessons: A Case Study', *The Malaysian Online Journal of Educational Technology*, 2(3), pp. 11-19

Leahy, S., and Wiliam, D. (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).

Leahy, S., Lyon, C., Thompson, M. and Wiliam, D. (2005) 'Classroom Assessment: Minute by minute, Day by day', *Educational Leadership*, 63(3), pp.1-7.

Learning Point Associates (2009) Connecting Formative Assessment Research to Practice. Available at: http://www.learningpt. org / pdfs/ Formative Assessment .pdf (Accessed: 3 November 2011).

Lee, R. E. (2008) 'Evaluate and assess research methods in work education: Determine if methods used to evaluate work education research are valid and how assessment of these methods is conducted', *Online Journal of Workforce Education and Development*, Volume III, Issue 2.

Lehr, C. A., Johnson, D., Bremer, C. D., Cosio, A. and Thompson, M. (2004). *NCSET*. Available at: http://www.ncset.org/ publications/essentialtools / dropout /part1.3.asp (Accessed: 3 November 2011).

Lei, P. and Wu, Q. (2007) *Introduction to Structural Equation Modelling: Issues and Practical Considerations.* Available at: http://ncme.org/linkservid

/47EFEB5A-1320-5CAE-6EC90BFDF09AA39E /showMeta/0/ (Accessed: 7 November 2013).

Lembaga Peperiksaan (2010) *Pentaksiran Berasaskan Sekolah*. Available at: http://buletinkpm.blogspot.co.uk/2012/11/pentaksiran-berasaskan-sekolah-pbs.html (Accessed: 7 November 2011).

Lembaga Peperiksaan (2012) *Panduan Pengurusan PBS. KPM.* Available at: http://www.moe.gov.my/lp/files/pbs/bahan /Buku%20Panduan %20 Pengurusan%20Pentaksiran%20Berasaskan%20Sekolah%20%28PBS%29%202012.pdf (Accessed: 26 November 2013).

Lembaga Peperiksaan (2014) *Portal Rasmi Lembaga Peperiksaan.* Available at: http://apps2.moe.gov.my/lponline/v1/index.php?option=com_content&view=article&id=242:buku-panduan-pengurusan-pentaksiran-berasaskan-sekolah-pbs&catid=14:pbs-rendah&Itemid=113&lang=enb (Accessed: 7 November 2014).

Lembaga Peperiksaan Malaysia (LPM) (2007) *Penataran Pentaksiran Berasaskan Sekolah (PBS), Mulai 2007.* Available at: http://sites.google.com/site/pbskedah/ (Accessed: 7 February 2012).

Lembaga Peperiksaan Malaysia (LPM) (2008) *Pentaksiran Berasaskan sekolah mulai tahun 2008. Unit UPSR, SPPMRR.* Available at: www.ppdsetiu.net/unit /bahasa/pbs/MODUL PBS upsr.ppt (Accessed: 9 February 2012).

Lembaga Peperiksaan Malaysia (2009) *SPPK Dalam Transformasi Pendidikan. Mesyuarat Pengurusan Pendidikan KPPM. Bil 2/2009.* Available at: http://www. slideshare.net/hmd1610/s-p-p-k-28-feb-k-p-p-m-u-p-s-i (Accessed: 28 November 2011).

Lembaga Peperiksaan Malaysia (2011a) *Panduan dan Peraturan PBS. Putrajaya.* Available at: http://www.scribd.com/faridooi/d/81532201-Panduan-Dan-Peraturan-Pbs-2011. (Accessed: 22 November 2011).

Lembaga Peperiksaan Malaysia (2011b) *Pentaksiran matematik di peringkat peperiksaan pusat*. Available at: ipgmkdri.edu.my/jm/documents /Pentaksiran Maths.ppt (Accessed: 29 November 2011).

Lembaga Peperiksaan Malaysia (2013) *Penjaminan Kualiti PBS Penyelarasan, Sektor Pengurusan Akademik JPN Pahang*. Available at: http://www.slideshare .net/spajpnp/penjaminan-kualiti-pbs-penyelarasan (Accessed: 28 November 2013).

Leshem, S and Trafford, V. (2007) 'Overlooking the conceptual framework', *Innovations in Education and Teaching International,* 44(1), pp. 93-105.

Levin, B. (2000) 'Putting students at the centre in education reform', *Journal of Educational Change, Kluwer Academic Publishers*, 1, pp. 155-172.

Levin, B. (2008) 'IN CANADA: 20 Minutes to Change a Life?', *Phi Delta Kappan*, 90(5), pp. 384-385.

Levin, B. (2010a) 'The challenge of large-scale literacy improvement, School Effectiveness and School Improvement', *An International Journal of Research, Policy and Practice*, 21(4), pp. 359-376.

Levin, B. (2010b) 'GLOBAL VOICES IN CANADA: What did you do in school today?', *Phi Delta Kappan,* 91(5), pp. 89-90.

Levin, B. and Fullan, M.G. (2008) 'Learning about systems renewal', *Educational Management Administration and Leadership,* 36(2), pp. 289-303.

Levin, B., Glaze, A. and Fullan, M. (2008) 'Results Without Rancor or Ranking: Ontario's Success Story', *Phi Delta Kappan*, 90(4), pp. 273-280.

Levin, H. M. (2001) 'Learning from School Reform', *International Conference on Rejuvenating Schools through Partnership*, Hong Kong, Chinese University of Hong Kong. Available at: http://www.tc.columbia.edu/centers /coce/pdf_files /b4.pdf (Accessed: 29 April 2013).

Lim, C. S. And Zhao, X. P. (2005) 'Assessment and Examination System: A Comparative Study between Malaysia and Shanghai (China)', *EARCOME* 3, TSG 6(3), pp. 1-11.

Linda, B., Patricia L. S., Ann A. and John A. (1998) 'Influence of Experience, Grade Level, and Subject Area on Teachers' Assessment Practices', *The Journal of Educational Research,* 91(6), pp. 323-330.

Linn, R. L. and Miller, M. D. (2005) *Measurement and Assessment in Teaching* 9th edn. Upper Saddle River, NJ: Prentice Hall.

Lo, Y. C. (2006) 'Practice and Challenges of School-based Formative Assessment'. *32nd Annual Conference of International Association for Educational Assessment*, Singapore. Available at: http://repository.ied.edu.hk /dspace/handle/2260.2/3148 (Accessed: 22 April 2013).

Losby, J. and Wetmore, A. (2012) *CDC Coffee Break: Using Likert Scales in Evaluation Survey Work.* Available at: http://www.cdc.gov/dhdsp /pubs/ docs/CB_February_14_2012.pdf (Accessed: 22 December 2013).

Loukas, A. (2007) 'NAESP: What is School Climate?', *Leadership Compass*, 5(1), [Online]. Available at: http://www.naesp.org/resources/2 /Leadership _Compass/2007/LC2007v5n1a4.pdf (Accessed: 5 August 2013).

Lukman, S. A. and Uwadiegwu, A. A. (2012) 'SBA as an innovation in Nigerian educational system: The Implementation Challenges', *Knowledge Review*, 25(1).

Mackenzie, N. and Knipe, S. (2006) 'Research dilemmas: Paradigms, methods and methodology', *Issues In Educational Research*, 16(2), pp. 193-205.

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

Maheran, Z. (2010) *An analysis of ethical judgments of Malaysian auditors*. Phd Thesis. University Sains Malaysia. Available at: http://eprints.uitm.edu.my/ 5633/1/MAHERAN% 20BINTI%20ZAKARIA%2010_24.pdf (Accessed: 3 November 2014).

Malakolunthu, S. and Hoon, S. K. (2010) 'Teacher perspectives of SBA in a secondary school in Kuala Lumpur', *Procedia Social and Behavioural Sciences*, 9, pp. 1170-1176.

Malaysian Education Blueprint (2012) *Preliminary Report, Malaysian Education Blueprint 2013-2025. September 2012.* Available at: http://www.moe.edu.my /btp/ wp-content/uploads/2012/Blueprint/Preliminary-Blueprint-BM.pdf (Accessed: 29 August 2013).

Mansel, W., James, M. and Assessment Reform Group (2009) *Assessment in schools. Fit for purpose? A Contemporary by the Teaching and Learning Research Programme.* London: Economic and Social Research Council, Teaching and Learning Research Programme.

Mansor, A. N., Ong, H. L., Rasul, M. S. and Raof, R. A. (2013) 'The Benefits of School-Based Assessment', *Asian Social Science*, 9(8), pp. 101-106.

Maria, T., Bailles, E., Caseras, X., and Martinez, A. (2009) 'Formative Assessment and academic achievement in pre-graduate students of health sciences', *Advanced in Health Science Education*, 14, pp. 61-67.

Martha L.A. (2001) *OAPA Handbook. COURSE-Based Review and Assessment: Methods for Understanding Student Learning.* Available at: www.umass. edu/oapa/oapa/publications/online.../course_based.pdf (Accessed: 6 November 2011).

Martinez, P. (2001) 'Effectiveness and improvement: school and college research compared' *Annual Conference of the British Educational Research Association*. Leeds University. Available at: http://www.leeds.ac.uk/educol /documents/00001835.htm (Accessed: 29 August 2013).

Marzita, A. (2010) *Pentaksiran diinsankan dan dimesrakan*. Available at: http: //www.tutor.com.my/tutor/dunia.asp?y=2006&dt=1009&sec=Berita&pg=be _01.htm (Accessed: 6 August 2013).

Marzita, A. (2012) *Kekurangan guru sukar diatasi, Utusan Online.* Available at: http://www.utusan.com.my/ utusan/info.asp?y=2012 &dt=0619&pub=Utusan_Malaysia&sec =Dalam _Negeri&pg=dn_02.htm (Accessed: 8 August 2013).

Mason, J. (2006) *Six strategies for mixing methods and linking data in social science research.* ESRC National Center for Research Methods: University of Manchester.

McNamara, C. (2002) *A Basic Guide to Program Evaluation. Authenticity Consulting, LLC.* Available at: http://www.tgci.com/magazine /a%20basic% 20guide%20to%20program%20evaluation.pdf (Accessed: 6 March 2012).

McNamara, G., Joyce, P. and O'Hara, J. (2010) *Evaluation of Adult Education and Training Program*. Available at: http://sites.jmu.edu/theflog/ files /2013/03 /evaluationandtraining.pdf (Accessed: 4 December 2014).

MEB (Malaysian Education Blueprint) (2012) *Preliminary Report MEB 2013-2025.* Available at: http://www.moe.gov.my/userfiles/file/PPP/Preliminary-Blueprint-Eng.pdf (Accessed: 4 December 2013).

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).

Merriam-Webster Dictionary (2013) *Definition of evaluate*. Available at: http://www. merriam- webster.com/dictionary/evaluate (Accessed: 7 November 2013).

Miezan Adiliah, A. I. (2001) *Pengurusan Perkhidmatan Personalia Murid.* Shah Alam, Selangor: Cerdik Publications.

Miles, M.B. and Huberman, A. M. (1984) *Qualitative data analysis: a sourcebook of New Methods.* London: SAGE Publications, Inc.

Ministry of Education (2006) *Penambahbaikan UPSR dan PMR.* Available at: ppdkerian.edu.my/download/files/PBS.pptx (Accessed: 2 January 2012).

Ministry of Education (2010) *KSPK and KSSR*. Available at: *www.ippm. edu.my/KSSR/KSSR.ppt* (Accessed: 8 February 2012)

Ministry of Education (2011) *Penambahbaikan UPSR dan PMR*. Available at: http://www.slideshare.net/Fadzliaton/penilaian-berasaskan-sekolah-pbs (Accessed: 9 November 2011).

Ministry of Education (MOE) (2013) *Portal Rasmi KPM*. Available at: http://www.moe.gov.my/ (Accessed: 9 November 2013).

Mintz, S. (no date). Assessing Student Learning. Columbia University. [Online]. Available at: http://www.columbia.edu/cu/tat/pdfs/assessment.pdf (Accessed February 2012).

Mitchell, M and Jolley, J. (2001) *Research Design Explained*. Texas: Harcourt College Publishers.

Mitchem, K., Wells, D. and Wells, J. (2003) 'Using evaluation to ensure quality professional development in rural scholars', *Journal of Research in Rural Education,* 18(2), pp. 96-103.

MOE (2012) *PP dalam PBS*. Available at: http://www.slideshare .net/ laichunseong /3-pentaksiran-pusat (Accessed: 5 November 2013).

MOE (2013) *Pentaksiran Berasaskan Sekolah, Soalan-soalan Lazim*. Available at: http://www.moe.gov.my/lp/files/faq/FAQ%20PBS%202013.pdf (Accessed: 9 November 2013).

Mohamad, A. Y. (2014) 'Bolehkah guru terus menjadi lilin?', *Sinar Harian*, 10 Februari. [Online]. Available at: http://www.sinarharian.com.my/mobile/karya /bolehkah-guru-terus-menjadi-lilin-1.249293 (Accessed: 3 October 2014).

Mohamad Najib, A.G. (2011). *Pembinaan dan analisis ujian bilik darjah.* Johor Bharu: Penerbit UTM Press.

Mohamad Najib, A. G. (2009). *Prinsip Asas Penilaian Program Pendidikan*. Universiti Teknologi Malaysia.

Mohamad, R. and Alias, A. A. (2012) *Evaluating the Effectiveness of a Training Program Using the Four Level Kirkpatrick Model in the Banking Sector in Malaysia. Open University Malaysia*. Available at: http://eprints.oum.edu .my/717/1/evaluating_rosmah.pdf (Accessed: 1 August 2013).

Mohd Burhan, I. (2005) *A hierarchical linear modelling of school effects and student factors affecting Penilaian Menengah Rendah (PMR) achievement.* PhD thesis. International Islamic University Malaysia, Kuala Lumpur. [Online]. Available at: http://www.lib.iium.edu.my/mom2/cm /content/view/view

.jsp?key=yL2sewYIoHD0rZmLRGIJthR7WahdCHq720061017094830375 (Accessed: 12 August 2013).

Mohd Isa, N. and and Abu Naim, H. (2013) '*Kajian Keperluan Pembinaan Modul Pentaksiran Berasaskan Sekolah Bagi Tajuk Dinamik Dalam Mata Pelajaran Sains Tingkatan 2'*. Available at: http://educ.utm.my/wp-content/uploads/2013/11/4.pdf (Accessed: 12 November 2014).

Mohd Majid, K. (1990) *Kaedah Penyelidikan Pendidikan*. Kuala Lumpur: DBP.

Mohd Noor, N. and Sahip, S. (2010) *Pelaksanaan Pentaksiran Kerja Kursus Berasaskan Sekolah Bagi Matapelajaran Kemahiran Hidup Di Sekolah Menengah Kebangsaan Daerah Johor Bahru, Kawasan Skudai. pp. 1-7.* Available at: http://eprints.utm.my/10688/ (Accessed: 8 November 2011).

Mokhtar, I. (2009) *Pentaksiran Pendidikan*. Kuala Lumpur: Dewan Bahasa dan Pustaka.

MPM (2012) *Garis panduan umum pengurusan dan pengendalian PBS*. Available at: http://www.mpm.edu.my/web/guest/panduan-umum-pbs (Accessed: 29 August 2013).

mStar (2013) *Sektor pendidikan terus terima peruntukan terbesar*. Available at: http://www.mstar.com.my/cerita.asp ?sec=mstar_berita&file= /2013/10 /25/mstar_berita/20131025174352 (Accessed: 7 December 2013).

Muhammad, A. W. (2008) 'Meta evaluation of a teachers' evaluation programme using CIPP Model'*, Archives Des Sciences*, 65(7), pp. 230-252.

Muhyiddin, Y. (2012) *Teks Ucapan YAB Tan Sri Dato' Hj Muhyiddin bin Hj Mohd Yassin. Kementerian Pelajaran Malaysia 2012.* Available at: http://www. jpnperlis.gov.my/v11/download/TeksUcapanMenteri.pdf (Accessed: 9 February 2012).

Muijs, D. (2007) 'Understanding how pupils learn: Theories of learning and intelligence', in Brooks, V. Abbott, I. and Bills, L. (ed.) *Preparing to Teach in Secondary Schools: A Student Teacher's Guide to Professional Issues in Secondary Education*. Maidenhead: Open University Press, pp. 45-59.

Muijs, D. (2011a) 'Leadership and organizational performance: from research to prescription?', *International Journal of Educational Management*. 25(1), pp. 45-60.

Muijs, D. (2011b) *Doing Quantitative Research in Education with SPSS*. London: SAGE Publications Ltd.

Muijs, D. and Harris, A. (2006) Teacher led school improvement: Teacher leadership in the UK. *Teaching and Teacher Education*. pp. 1-12. Available

on line at: http://zhaomr01.jxcg2011.nwnu .edu.cn/source /lw/L16%20 Teacher%20led%20school%20improvement.pdf (Accessed: 9 January 2014).

Muijs, D. and Lindsay, G. (2008) 'Where are we at? An empirical study of levels and methods of evaluating continuing professional development', *British Educational Research Journal,* 34(2), pp. 195-211.

Muijs, D. and Reynolds, D. (2003) 'The effectiveness of the use of learning support assisstants in improving the mathematics achievement of low achieving pupils in primary school', *Educational Research*, 45(3), pp. 219-230.

Muijs, D. and Reynolds, D. (2011) *Effective teaching: evidence and practice.* 3rd edn, London, GB: Sage.

Muijs, D., Day, C., Harris, A. and Lindsay, G. (2004) Evaluating CPD: an overview, in C. Day and J. Sachs (eds) *International Handbook on the Continuing Professional Development of Teachers.* Maidenhead: Open University Press. ….goodall et al….

Mukhari, A. W. and Md. Amin, R. (2010) *UTM Institutional Repository.* Available at: http://eprints.utm .my/11549/ (Accessed: 8 November 2011).

Narayanan, R., Rajasekaran, N. and Iyyappan, S. (2008) 'Factors affecting English learning among engineering and technology students', *Iranian Journal of Language Studies,* 2(4), pp. 485-512.

Nasir, A. G. (2011) *UKM NewsPortal*. Available at: http://www.ukm.my/news /index.php/ms/ berita-kampus/862-teaching-assistants-have-become-a-must-says-education-specialist.html (Accessed: 5 November 2014).

Nazim, A. and Ahmad, S. (2013) 'A Comparison Between OLS and SEM Methods in Estimating The Influential Factors of 8th Grades', *International Journal of Scientific & Engineering Research*, Volume 4, Issue 7, pp. 717-722.

NCTE (2010) *Teacher Learning Communities*. Available at: http://www.ncte .org/library/NCTEFiles/Resources/Journals/CC/0202-nov2010 /CC0202 Policy.pdf (Accessed: 7 January 2014).

Nea (1995) *Summative and formative feedback. Update, 1(4).* Available at: www.nea.org/assets/docs/HE/upv1no5.pdf (Accessed: 7 July 2013).

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

Newsome (2012) *Practical Approaches in Dealing with Nonnormal and Categorical Variables.* Available on line at: http://www.upa.pdx.edu /IOA/ newsom/semclass/ho_estimate2.pdf (Accessed: 9 November 2013).

NIC (Northern Ireland Curriculum) (2007) *Assessment for Learning for Key Stages 1 & 2. Council for the Curriculum, Examinations and Assessment (CCEA).* Available at: http://www.nicurriculum.org.uk/docs /assessment_ for_learning/training/AfL-Guidance-KS12.pdf (Accessed: 1 March 2012).

Nicol, D. and Macfarlene-Dick, D. (2004) *Rethinking Formative Assessment in HE: a theoretical model and seven principles of good feedback practice.* Available at: http://www-new1.heacademy.ac.uk/assets/documents/ assessment/web0015 _rethinking_formative_assessment_in_he.pdf (Accessed: 1 February 2015).

Nicol, D. J. and Macfarlene-Dick, D. (2006) 'Formative assessment and self-regulated learning: A model and seven principles of good feedback practice', *Studies in Higher Education*, 31(2), pp. 199-218.

NIER (National Institute for Educational Research) (1999) *An International Comparative Study of School Curriculums. Tokyo.* Available at: http://www. hurights.or.jp/pub/hreas/3/16national_institute.pdf (Accessed: 8 March 2012).

Nik Azis, N. P. and Noraini, I. (2008) *Perjuangan Memperkasakan pendidikan di Malaysia.* Kuala Lumpur: Utusan Pubs and Distributor Sdn Bhd.

NOAA (2009) *Designing Education Projects*. Available on line at: http://www .oesd.noaa.gov/leadership/DEP_Manual_2ndEdt_Final.pdf (Accessed: 21 January 2015).

Nor Azni, A. Z., Foo, S. F., Soaib, A. and Aminudin, H. (2014) 'Kepimpinan instruksional pengetua dalam perlaksanaan pentaksiran berasaskan sekolah' *Proceeding of the Global Summit on Education GSE 2014*. Kuala Lumpur, Malaysia, 4-5 March. Fakulti Pengajian Pendidikan, UPM. Available at: http://worldconferences.net/proceedings/gse2014/toc/papers_ gse2014/G171%20-%20NOR%20AZNI%20-%20%20KEPIMPINAN%20 INSTRUKSIONAL %20 PENGETUA%20DALAM%20PELAKSANAAN %20PENTAKSIRAN %20BERASASKAN%20SEKOLAH.pdf (Accessed: 27 December 2014).

Nora, A., and Barlow, E. A. (2006) 'Reporting structural equation modeling and confirmatory factor analysis results: A review', *The Journal of Educational Research*, 99(6), pp. 323-337.

Norizan, A. M. (2012) *Utusan Online*. Available at: http://www.utusan.com .my /utusan /Pendidikan/20121029/pe_01/Pelaksanaan-PBS-tidak-bebankan-guru pdf (Accessed: 9 July 2013).

Noruzi, M. R. and Rahimi, G. R. (2010) 'Multiple Intelligences, A New Look to Organizational Effectiveness', *Journal of Measurement Research*, 2(2), pp. 1-15.

Norzila, M. Y. (2013) *Cambridge Horizons.* Available at: http://www.cambridge assessment .org.uk/images/139719-sba-seminar-papers.pdf (Accessed: 14 November 2013).

Nor Hasnida M. G., Baharin Y. and Afian A. M. (2012) 'Why Do We Need to Change?. Teachers' Attitude Towards School-Based Assessment System' *SCR London's First International Conference on Social Science and Humanities in the Islamic World*, London, 28-30 May. Available at: http://bu.univ-ouargla.dz /production%20scientifique /international/2012/PAPER%20LONDON %20CONFERENCE1.pdf (Accessed: 8 March 2012).

NSW Department of Education and Training (2007) *Assessment in Primary Schools.* Available at: http://www.curriculumsupport.education .nsw.gov .au/timetoteach/report/doc/assessinpri.pdf (Accessed: 3 November 2011).

Nunnally, J.O. (1978) *Psychometric theory.* New York: McGraw-Hill.

Nuthanap, G. (2007) *Gender Analysis of Academic Achievement Among High School Students.* Thesis Degree of Master of Home Science in Human Development. University of Agricultural Science, Dharwal.

Nur, A. (2013) *Harvard Graduate School of Education.* Available at: http://isites.harvard.edu/fs/docs /icb.topic 1203150 .files/Panel%203%20-%20Activating%20the%20Greatest%20Power/AzizNur_Malaysia _RuralTeacher Retention.pdf (Accessed: 15 November 2014).

NWEA (2013) *Keeping Learning on Track*. Available at: http://www.nwea. org/node/12150 (Accessed: 7 July 2013).

Juliana, O. (2008) *Hubungan antara pengetahuan dengan persepsi amalan pentaksiran formatif dalam perlaksanaan penilaian kemajuan berasaskan sekolah.* Master Thesis. UMS. Available at: http://eprints.ums.edu.my /1991/

(Accessed: 9 November 2013).

Hashim, O. (2009) 'Pemantapan pentaksiran Kemahiran Menulis Bahasa Melayu Peringkat Sekolah Menengah', *Educationist,* III (2), pp. 90-103.

Omuga, T. O. (2013) *Constraints to implementation of the school based teacher development (SbTD) program in primary schools in Siaya District, Kenya*. Master Thesis. Kenyatta University. Available at: http://ir-library.ku.ac.ke/bitstream/handle/123456789/6340/Omuga%20Thomas%20Odhiambo.pdf?sequence=3&isAllowed=y (Accessed: 9 November 2013).

Ong, S. L. (2010) 'Assessment profile of Malaysia: high stakes external examinations dominate', *Assessment in Education: Principles, Policy & Practice,* 17(1), pp. 91-103.

Oppenheim, A. N. (1996) Questionnaire Design, Interviewing and Attitude Measurement, London: Pinter Publishers.

Othman, M. and Muijs, D. (2013) 'Educational quality differences in a middle income country: the urban-rural gap in Malaysian primary schools', *School Effectiveness and School Improvement*, 24(1), pp. 104-122.

Othman, I., Md Saleh, N. and Md Norani, N. A. (2013) 'The Implementation of School Based Assessment In Primary School Standard Curriculum', *International Journal of Education and Research*, 1(7), pp. 1-10.

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

Page, H. (2003) *The value of Likert scales in measuring attitudes of online learners.* Available at: http://www.hkadesigns.co.uk/websites /msc/ reme/ likert.htm (Accessed: 9 November 2011).

Pallant, J. (2007) *SPSS Survival Manual: A step by step guide to data analysis using SPSS for Windows, 3rd ed.* New South Wales: Allen and Unwin.

Palm, T. (2008) 'Performance Assessment and Authentic Assessment: A Conceptual Analysis of the Literature', *Practical Assessment Research & Evaluation*, 13(4).

Palomba, C. A. and Banta, T. W. (1999) *Assessing Essentials: Planning, Implementing and Improving Assessment in Higher Education.* San Francisco: Jossey-Bass.

Partners in Learning (2009) *A Systems Framework for Guiding Innovative School Reform*. Available at: http://www.is-toolkit.com/knowledge_library/kl_files/Innovation_Framework_whitepaper_final.pdf (Accessed: 2 November 2013).

Patton, M. Q. (1990) *Qualitative Evaluation and Research Methods*. Newbury Park: Sage Publications, Inc.

Patton, M. Q. (2000) 'Utilised-focused evaluation', in Stufflebeam D.L., Madaus G.F. and Kellaghan, T. (ed.) *Evaluation Models*. Boston: Kluwer Academic Publishers.

Pemandu (2010) *Improving Student Outcomes.* [Online]. Available at: http://www. pemandu.gov.my/gtp/upload (Accessed: 4 August 2013).

PEMANDU (2011a) *NKRA Pendidikan*. Available at: http://www.slideshare .net/mohdjamil/nkra-pendidikan (Accessed: 6 November 2011).

PEMANDU (2011b). *Literacy and Numeracy*. Available at: http://pelajaranperak.gov.my /v2/modules/mastop_publish/files/files_4b2ed3402c1e0pdf (Accessed: 8 November 2011).

Pengurusan Akademik PPD Hulu Selangor (2011) *KSSR.* Available at: http://www.slideshare.net/LatifahRazali/kssr-2011 (Accessed: 9 November 2011).

Penilaian dan Peperiksaan (2011) *Buku Pengurusan dan Pengendalian Pentaksiran Berasaskan Sekolah (PBS) Di Sekolah Rendah.* Available at: http://www.moe.gov.my/lp/files/penerbitan/buku_pbs/Kandungan.pdf (Accessed: 23 November 2011).

Perrie, M., Marion, S. and Gong, B. (2007) *The Role of Interim Assessments in a Comprehensive Assessment System: A Policy Brief, Centre for Assessment, The Aspen Institute.* Available at: http://www.achieve.org/files/ TheRole ofInterim Assessments.pdf (Accessed: 8 August 2013).

PIPP (2012) *Laporan Awal Pelan Pembangunan Pendidikan Malaysia 2012-2025*. Available at: http://www.moe.gov.my/userfiles/file /PPP / Preliminary-Blueprint-BM.pdf (Accessed: 8 August 2013).

Popham, W. J. (2008) *Transformative Assessment*. Alexandria, Virginia: ASCD.

Portal PKG Sg Pari (2012). *Penerangan Tentang Pentaksiran Berasaskan Sekolah*. Available at: http://pkgsp.org/v11/index.php?option= com_content &view=article&id=195:penerangan-tentang-pentaksiran-berasaskan-sekolah&catid=36:artikel&Itemid=92 (Accessed: 22 March 2012).

PPD Barat Daya (2010) *Sistem Pentaksiran Pendidikan Kebangsaan*. Available at: http://www.scribd.com/doc/56231571/tentang-penilaian-KSSR (Accessed: 8 March 2012).

PPD Hulu Selangor (2011) *Kurikulum Standard Sekolah Rendah (KSSR).* Available at: http://skkayang.edu.my/kssr2011-110521112300-phpapp02.pdf (Accessed:26 March 2012).

PPD Maran (2013) *PROGRAM "OUTREACH PBS - PUBLIC ENGAGEMENT PIBG" PERINGKAT DAERAH MARAN 2013.* Available at: http://gerbangkualiti .blogspot.co.uk/2013/03/program-outreach-pbs-public-engagement.html (Accessed: 9 March 2012).

Pramanik, T. K. (1988) *A study of some affective variables related to learning English as a second language at the post-secondary stage in Orissa.* Unpublished PhD thesis. Hydrabad, Central Institute of English and Foreign Languages.

Psbtpnsarawak (2011) *Ministry bent on improving academic performance*. Available at: http://psbtpnsarawak.wordpress.com/2011/08/22/ministry-bent-on-improving-academic-performance (Accessed: 9 January 2013).

QSA (Queensland Studies Authority) (2010) *SBA The Queensland system*. Available at: http://www.qsa.qld.edu.au/downloads/approach/school-based_assess _qld_sys.pdf (Accessed: 8 November 2013).

Qualifications and Curriculum Authority (2003) *School-Based Assessment*. Available at: http://www.traffordlearning.org/trafford /sections /public_html /teaching/curriculum/ict/documents/QCAreseachintopracticeAssforLrng.pdf (Accessed: 8 July 2013).

Queen's University Belfast (2013) *Summative assessment has a positive impact on learning*. Available at: http://www.qub.ac.uk/directorates / Academic StudentAffairs/CentreforEducationalDevelopment/AssessmentFeedback/GoodassessmentandfeedbackpracticeSummativeassessmenthasapositiveimpactonlearning/ (Accessed: 7 December 2013).

Quintana, S. M. and Maxwell, S. E. (1999) 'Implications of recent developments in SEM for counselling psychology', *The Counselling Psychologist*, 27(4), pp. 485-527.

Race, P. (2001) *The Lecturers' Toolkit*. Kogan Page: London.

Race, P. (2006) *The Lecturer's Toolkit – A Practical Guide to Assessment*, Routledge: London.

Raman, K. and Yamat, H. (2014) 'English Teachers' Voices on the challenges of SBA', *Frontiers of Language and Teaching*, 5(1), pp. 66-74.

Rahim, M. (2012) *Perkhidmatan Pelajaran*. Available at: http://perkhidmatan pelajaran .blogspot.co.uk/ (Accessed: 25 October 2014).

Ramlee, I. (2009) *The Impact of Schooling Reform on Returns to Education in Malaysia. MPRA Paper, University of Munich.* Available at: http://mpra.ub.uni-muenchen.de/15021/ (Accessed: 23 March 2012).

Randall, D. M. and Gibson, A. M. (1990) 'Methodology in business ethics research: A review and critical assessment', *Journal of Business Ethics*, 9, pp. 457-471.

Rayment, T. (2006) *101 Essential Lists on Assessment*. Kuala Lumpur: Institut Terjemahan Negara Malaysia Bhd.

Razali, A. (1987) *Model-model penilaian di dalam pendidikan: Satu penjelasan dan perbandingan.* Bangi: Universiti Kebangsaan Malaysia.

Ready, D., Lee, V. and Welner, G. (2004) 'Educational Equity and School Structure: School Size, Overcrowding, and Schools-Within-Schools', *Teachers College Record,* 106(10), pp. 1989-2014.

Richardson, V. (2003). 'Constructivist Pedagogy', *Teachers College Record*, 105(9), pp. 1623-1640.

Reyneke, M., Meyer, L. and Nel, C. (2010) 'SBA: the leash needed to keep the poetic 'unruly pack of hounds' effectively in the hunt for learning outcomes', *South African Journal of Education,* 30, pp. 277-292.

Rinaldi, C. (2006) *In dialogue with Reggio Emilia: Listening, researching and learning.* London: Routledge.

Robinson, B. (2002) *The CIPP approach to evaluation. COLLIT Project*. Available at: http://www.fivehokies.com/Evaluation/Evaluation% 20Approaches /Management%20Oriented/CIPP%20Approach%20to%20Evalaution.pdf (Accessed: 9 November 2011).

Rohaya, T. (2009) *Pembinaan dan Pengesahan Instrumen Ujian Literasi Pentaksiran.* PhD thesis. Universiti Teknologi Malaysia. Available at: http://eprints.utm.my/13641/7/RohayaTalibPFP2009CHAP1.pdf (Accessed: 22 November 2013).

Rossi, P. H., Lipsey, M. W. and Freeman, H. E. (2004) *Evaluation: A Systematic Approach*. 7th edn. Thousand Oaks: Sage Publications.

Rust, C. (2002) *Purposes and principles of assessment. OCSLD, Learning and teaching briefing papers series. Oxford Brookes University*. Available at: http://www.brookes.ac.uk/services/ocsd/2learntch/briefing_papers/p_p_assessment.pdf (Accessed: 6 November 2011).

Rust, C., ODonnovan, B. and 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), pp. 231-240.

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

Said, H., Badru, B. and Shahid, M. (2011) 'Confirmatory Factor Analysis (CFA) for Testing Validity And Reliability Instrument in the Study of Education', *Australian Journal of Basic and Applied Sciences*, 5(12), pp. 1098-1103.

Salinah, J. (2002) *Kertas Kerja 31, UPSI.* Available at: http://pustaka2.upsi.edu.my/eprints/304/1 /Salinah%20Ja'%20afar-Kesantunan% 20Bahasa%20 Pertembungan %20Antara%20Budaya%20Melayu-Bukan%20Melayu%20dan %20Islam-Bukan%20Islam.pdf (Accessed: 23 November 2014).

Salmiah, J. (2013) 'Acceptance towards SBA among agricultural integrated living skills teachers: challenges in implementing a holistic assessment', *Journal of Technical Education and Training (JTET),* 5(1), pp. 44-51.

Sang, M. S. (2006) *Nota Pengurusan Pembelajaran*. Kuala Lumpur: Multimedia-ES Resources Sdn Bhd.

Sarouphim, K. M. (2007) 'DISCOVER in Labenon: A Pilot Study' *88th Annual Meeting of the American Educational Research Association*, Chicago, April 2007. Available at: http://www.eric.ed.gov/ERICWebPortal /search/ detailmini.jsp?_nfpb=true&_&ERICExtSearch_SearchValue_0=ED496155&ERICExtSearch_SearchType_0=no&accno=ED496155 (Accessed: 8 March 2012).

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A. and King, J. (2006) 'Reporting Structural Equation Modeling and Confirmatory Factor Analysis Results: A Review', *Heldref Publications*, 99(6), pp. 323-337.

Scriven, M. (1967) 'The methodology of evaluation', in Stake, R. E. (Ed.), *Curriculum Evaluation: AERA monograph series on evaluation* (pp. 39-85). Chicago: Rand McNally.

SEAMEO Secretariat (2001) *THE SECONDARY SCHOOL EXAMINATION SYSTEMS IN COUNTRIES IN THE SEAMEO REGION.* Available at: http://www.seameo .org /vl/library/dlwelcome/publications/ebook /exam/2ndintro.htm# (Accessed: 8 November 2011).

Sekaran, U. (1992) *Research Methods for Business – A skill building approach. 2nd ed.* United States of America: John Wiley & Sons, Inc.

Schneider, B. (1990). Organisational Climate and Culture. San Francisco: Jossey Bass

Shabait (2010) *What Motivation is and Why It is Important in Classroom Situation, Ertrea, Ministry of Information.* Available at: http://www.shabait.com/categoryblog/3200-what-motivation-is-and-why-it-is-important-in-classroom-situation (Accessed: 8 July 2013).

Shaharuddin, I. (no date) *Kaji semula ruang Bandar dan luar bandar*. Available at: http://www.ukm.my/lestari/doc/popular/document%2823%29.pdf (Accessed: 8 February 2015).

Shepard, L. A. (2005) 'Linking Formative Assessment to Scaffolding', *Educational Leadership,* 63(3), pp. 66-70.

Shimura, K. (2003) *Issues of Program Evaluation in Social Work Education: A Review from the Perspectives of Evaluation Framework.* Available at: http://hrr.ul.hirosaki-u.ac.jp/dspace/bitstream /10634/5929/1 /HiroGaku ShakaiFukushi_3_51.pdf (Accessed: 8 November 2013).

Sidek, R. and M Saleh, S. (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).

Silin, S., Zarins, S. and Mulford, B. (2002) 'What characteristics and processes define a school as a learning organisation? Is this a useful concept to apply to schools?', *International Education Journal*, 3(1), pp. 24-32.

Silver, H. (2004) *Evaluation research in Education*. Available at: www.edu. plymouth.ac.uk/resined/evaluation/ (Accessed: 9 July 2013).

Simatwa, E. M. (2010) 'Piaget's theory of intellectual development and its implication for instructional management at pre-secondary school level', *Educational Research and Reviews*, 5(7), pp. 366-371.

Sinar Harian (2014) *Kementerian beri kefahaman kepada guru fahami PBS*. Available at: http://www.sinarharian.com.my/kementerian-beri-latihan-kepada-guru-fahami-pbs-1.268776 (Accessed: 5 September 2014).

Siti Rohayah, A., T. Subahan, M. M., Norasmah, O. and Ibrahim, M. Z. (2004) 'Analisis kecerdasan pelbagai di kalangan pensyarah di sebuah IPT', *Jurnal Teknologi*, 41(E), pp. 33-42.

Skinner, B. F. (1971) *Beyond Freedom and Dignity*. New York: Vintage Books.

SMK Tunku Bendahara (2012) *Manual Pentaksiran Berasaskan Sekolah (PBS) 2012.* Available at: http://akademik.kedah.edu.my/index. php?option=com

_content&view=article&id=111:kssr&catid=1:latest-news (Accessed: 22 November 2011).

Snijders, T. A. B. and Bosker, R. J. (2011) *Multilevel Analysis: An Introduction to basic and advanced multilevel modelling*. London: SAGE Pubications.

SQA (2008) *The Assessment Systems of Finland and Queensland*. Available at: http://www.sqa.org.uk/files_ccc/PNP_ResearchReport1_TheAssessmentSystemsFinlandQueensland.pdf (Accessed: 29 November 2013).

Stake, R. E. (1973) *Conference on New Trends in Evaluation, Institute of Education at Goteberg University.* Available at: http://education.illinois.edu/circe/Publications/Responsive_Eval.pdf (Accessed: 8 March 2012).

Stanford SRN (2008) *What is Performance-based Assessment? Stanford University School of Education.* Available at: http://edpolicy.stanford.edu/sites/default/files/events/materials/2011-06-linked-learning-performance-based-assessment.pdf (Accessed: 9 March 2012).

Steenhuis, H-J. and Bruijn de, E. J. (2006) *Empirical Research in OM: three paradigms.* Available at: http://doc.utwente.nl/73618/ (Accessed: 6 November 2013).

Stevahn, L., King, J. A., Ghere, G and Minnema, J. (2005) 'Establishing Essential Competencies for Program Evaluators', *American Journal of Evaluation*, 26, pp. 43-59.

Stiggins, R. (2005) 'From Formative Assessment to Assessment for Learning: A Path to Success in Standards-Based Schools', *Phi Delta Kappan*. 87(4), pp. 324-328.

Stiggins, R. J. (2002) 'Assessment Crisis: The Absence of Assessment For Learning', *Phi Delta Kappan,* 83(10), pp. 758-765.

Stiggins, R. J. and Conklin, N. F. (1992) *In Teachers Hands: Investigating the Practices of Classroom Assessment.* Albany, N. Y.: State University of New York.

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).

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

Stufflebeam, D. L. (1983) 'The CIPP Model for Program Evaluation' in G. F. Madaus, M. Scriven and D. L. Stufflebeam (ed), *Evaluation Models:*

Viewpoints on Educational and Human Services Evaluation. Boston: Kluwer Nijhof.

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

Stufflebeam, D. L. (2003a) 'The CIPP Model for Evaluation' in Kellaghan, T and Stufflebeam, D. L., *International Handbook of Educational Evaluation*, pp. 31-62, Dordrecht: Kluwer Academic Publishers.

Stufflebeam, D.L. (2003b).'The CIPP Model for evaluation'. *2003 Annual Conference of the Oregan Program Evaluators Network* (OPEN), Oregan.

Stufflebeam, D. L. (2007) *CIPP Evaluation Model Checklist*. [Online]. Available at: http://www.wmich.edu/evalctr/archive_checklists /cippchecklist_ mar07.pdf (Accessed: 4 January 2014).

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

Stufflebeam, D. L. and Shinkfield, A. J. (2007) *Evaluation Theory, Models and Applications*. San Francisco: CA, Jossey- Boss.

Stufflebeam, D. L., Madaus, G. F. and Kellaghan, T. (2000) *Evaluation Models: Viewpoints on educational and human services evaluation.* 2nd edn. Boston: Kluwer Academic Publishers.

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

Suzieleez, S., Venville, G. and Chapman, A. (2009) 'Classroom Assessment: Juxtaposing teachers' beliefs with classroom practices'. *Australian Association for Research in Education: International Education Research Conference*, 29 November to 3 December. Available at: http://www.academia.edu/421177 /Classroom_Assessment_Juxtaposing_Teachers_Beliefs_with_Classroom_Practices (Accessed: 8 November 2013).

Swaffield, S. (2011) 'Getting to the heart of authentic assessment for learning', *Assessment in education: Principles, policy & practice*, 18(4), pp. 433-449.

Tabachnick, B. G. and Fidell, L. S. (2001) *Using multivariate statistics.* 4th edn. Boston: Allyn and Bacon.

Talib, R., Kamsah, M., Naim, H. and Latif, A. (2014) 'From Principle to Practice: Assessment for Learning in Malaysian SBA Classroom', *International J Soc. Sci. & Education*, Vol. 4 Issue 4.

Tan, A. M. (2010) *Pentaksiran Berasaskan Sekolah (PBS) di Malaysia. Kesediaan Guru, Isu dan Panduan Perlaksanaan.* Gerakbudaya Enterprise: Petaling Jaya.

Tan, H. Y. (2004) *Using Formative Assessment Strategies to aid students' learning in adult English teaching as foreign language – An Experimental Study*. Beijing: Captial Normal University Press.

Underhill, N. (1987) *Testing spoken language: A handbook of oral testing techniques.* London: Cambridge University Press.

Tashakkori, A. and Teddlie, C. (2003) *Handbook of mixed methods in social and behavioural research.* Thousand Oaks, CA: Sage.

Tayyari, F. (2004) *The Science Creative Quarterly.* Available at: http://www.scq.ubc.ca/the-genetic-basis-of-intelligence/ (Accessed: 25 January 2013).

Teaching-Assisstants.co.uk. (2013) *Essential Job Information.* Available at: http://www.teaching-assistants.co.uk/job-information-for-teaching-assistants.htm (Accessed: 29 August 2013).

The Free Dictionary (2015) *The Free Dictionary*. Available at: http://www.the freedictionary.com/evaluation (Accessed: 7 February 2015).

The Northern Ireland Curriculum (2009) *AfL, Assessment for Learning A Practical Guide, A CCEA Publication*. Available at: http://www.nicurriculum .org.uk/docs/assessment_for_learning/AfL_A%20Practical%20Guide.pdf (Accessed: 7 November 2013).

Then, S. (2011) *The Star online.* Available at: http://www.thestar.com.my/story /?file=%2F2011%2F1%2F26%2Fnation%2F20110126132520 (Accessed: 21 January 2015).

Thimmappa, P. R. and Sharma, S. (2003) 'Implications of School Based Assessment for the Pacific', *Journal of Educational Studies,* 25(1), pp. 40-55.

Thomas, C., Brit, P., Blackbourn, J. M. and Blackbourn, R. (2004) 'Portfolio Assessment: A Guide for Teachers and Administrators', *National Forum of Educational Administration and Supervision Journal-Electronic,* 23(4E), pp. 1-8.

Tittle, C. and Hill, R. (1967) 'Attitude Measurement and Prediction of Behaviour: An Evaluation of Conditions and Measurement Techniques', *Sociometry*, Vol. 30.

Tomic, W. (1993) 'Behavior and Cognitive in education Psychology', *A Journal of Human Behavior,* 30(314), pp. 38-46.

Troy, T. (2011) *Comprehensive Assessment Systems: Purposes and Implementation, WAKE Country Public School System, E & R Report No 11. 10.* Available at: http://www.wcpss.net/results/reports/2011/1110cas.pdf (Accessed: 9 November 2013).

Tseng, K., Diez, C. R., Lou, S. and 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), pp. 256-261.

Tuckman, B. W. (1985) *Evaluating instructional program*. Boston: Allyn and Bacon.

Tunc, F. (2010) *Evaluation of an English language teaching program at a public university using CIPP Model.* Available at: http://etd.lib.metu.edu.tr/ upload/12611570/index.pdf (Accessed: 8 Nov 2013).

Tyler, R. W. (1950) *Basic principles of curriculum and instruction*. Chicago: The University of Chicago Press.

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).

UNESCO-IBE (2011) *World Data on education*. Available at: www.ibe.unesco .org /fileadmin/user_upload/.../pdf.../Malaysia.pdf (Accessed: 9 November 2011).

Underhill, A. F. (2006) 'Theories of Learning and their Implications for on-line assessment', *Turkish Online Journal of Distance Education-TOJDE*, 7(1), pp. 165-174.

Unicef (2008) *Education is human right.* Available at: http://www.unicef.org/ malaysia/Malaysia_Education_Actions.pdf (Accessed: 22 November 2011).

USM (2007) *National Philosophy of Education. 50 Years Higher Education Development in Malaysia (1957-2007).* Available at: http://jpt.mohe .gov. my/eng/DASAR/falsafah.php (Accessed: 26 June 2013).

Utusan Online (2011) *Guru tidak terbeban kerja*. 8 February. [Online]. Available at: http://ww1.utusan.com.my/utusan/info.asp?y=2011&dt =0208&sec =Forum&pg=fo_01.htm (Accessed: 9 November 2011).

Utusan Online (2012) Kerajaan tambahbaik PBS. Available at: http://ww1.utusan.com.my/utusan/Dalam_Negeri/20121104/dn_01/Kerajaan-tambah-baik-PBS (Accessed: 9 November 2014).

Utusan Online (2014a) *PBS, Utamakan Guru, Pelajar*. Available at: http://ww1.utusan.com.my/utusan/Rencana/20140302/re_07/PBS-Utamakan-pelajar-guru (Accessed: 22 November 2014).

Utusan Online (2014b) *Guru terima latihan penambahbaikan PBS.* Available at: http://ww1.utusan.com.my/utusan/Parlimen/20140410/pa_06/Guru-terima-latihan-penambah-baikan-PBS (Accessed: 22 November 2014).

Utusan Online (2014c) *80% beban guru berkurangan*. Available at: http://ww1.utusan.com.my/utusan/Dalam_Negeri/20140319/dn_06/80-beban-guru-berkurangan (Accessed: 2 November 2014).

Valencia, S. W. (2002) *Understanding Assessment: Putting together the puzzle*, Available at: http://www.eduplace.com/state/author/valencia.pdf (Accessed: 9 November 2011).

Vegas, E. (2007) 'Teacher labour markets in developing countries', *Future of Children*, 17(1), pp. 219-232.

Walinga, J. (2008) 'Toward a Theory of Change Readiness', *The Journal of Applied Behavioral Science,* XX(X), pp. 1-33.

Wan Zah, W. A., Fatimah, H. and Ismail, S. A. and Tamby Subahan, M. M. (2005) 'Kesan Program NILAM terhadap tabiat membaca menurut pandangan murid dan guru', *Jurnal Pendidikan,* pp. 39-55.

Wang, X. (2008) 'Teachers' Views on Conducting Formative Assessment in Chinese Context', *Engineering Letters*, 16(2).

Wawasan Pengarang (2014) *Utusan Sarawak.* Available on line at: http://www.utusansarawakonline.com/opinion/371/INFRASTRUKTUR-LENGKAP-MAMPU-LAHIR-PELAJAR-CEMERLANG/ (Accessed: 24 November 2014).

Weaver-Hart, A. (1988) 'Framing an innocent concept and getting away with it', *UCEA Review,* 24(2), pp. 11–12.

Webster, A., McNeish, D., Scott, S., Maynard, L. and Haywood, S. (2012) *What influences teachers to change their practice? A rapid research review.* Available at: http://www.bristol.ac.uk/cubec/portal/pr7.pdf (Accessed: 3 November 2014).

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

Weir, C. J. (1994) Understanding and developing language tests. London: Prentice Hall.

Welch, G. W. (2011) *Children youths families and schools*. Available at: http://r2ed.unl.edu/presentations/2011/RMS/022511_Welch/022511_Welch.pdf (Accessed: 3 November 2012).

West, A. and Ylonen, A. (2009) 'Market-oriented school reform in England and Finland: school choice, finance and governance', *Educational Studies*, 36(1), pp. 1-12.

Wholey, J. S. (1979) *Evaluation: Promise and Performance*. Washington, DC: Urban Institute.

Wikipedia (2013) *Evaluation*. Available at: http://en.wikipedia.org /wiki / Evaluation (Accessed: 3 November 2012).

Wiliam, D. (1998) 'The validity of teachers' assessment', *22nd Annual Conference of the International Group for the Psychology of Mathematics Education,* South Africa, July. Available at: www.dylanwiliam.org / Dylan... /PME%2022%20WG6%20paper.doc (Accessed: 10 Jan 2013)

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

Wiliam, D. (2006) 'Formative Assessment: Getting the focus right', *Educational Assessment,* 11(3), pp. 283-289.

Wiliam, D. (2007a) *National Council of Teachers of Mathematics*. Available at: http://www. nctm.org/uploadedFiles/Research_News_and_Advocacy /Research /Clips_and_Briefs/Research_brief_05_-_Formative_Assessment.pdf (Accessed: 5 October 2014).

Wiliam, D. (2007b) 'Content then process: Teacher learning communities in the service of formative assessment' in D. B. Reeves (edn). *Ahead of the curve: The power of assessment to transform teaching and learning* (pp. 183-204). Bloomington, IN: Solution Tree.

Wiliam, D. (2008) 'Changing Classroom Practice', *Educational Leadership*, 65(4), pp. 36-42.

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

Wiliam, D. (2011) 'What is assessment for learning?' *Studies in educational evaluation*, 37, pp. 3-14.

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

Wilson, T. D. (2012) *The New York Times*. Available at: http://www.nytimes. com /2012/03/11/books/review/the-power-of-habit-by-charles-duhigg. html?pagewanted=all&_r=0 (Accessed: 8 November 2011).

Woolfolk, A. (2004) *Educational Psychology*. 9th edn. Boston: Allyn and Bacon.

Worthen, B. R., Sanders, J. R. and Fitzpatrick, J. L. (1997) *Program Evaluation: Alternatives Approaches and Practical*. New York: Addison Wesley Longman, Inc.

Wren, D.G. (2009) 'Performance assessment: A key component of a balanced assessment system', *Research Brief*, 2, pp. 1-12.

Wylie, E. C. (2008) *Tight but loose: Scaling up Teacher Professional Development in Diverse Contexts, Educational Testing Service.* Available at: http://www.ets.org/Media/Research/pdf/RR-08-29.pdf (Accessed November 2013).

Yip, D. Y. and Cheung, D. (2005) 'Teachers' Concern on SBA of practical work', *Journal of Biological Education*, 39(4), pp. 156-162.

Yong, H. T. and Lim, C. S. (2008) *Buku Koleksi Bahan Seminar Inovasi Pedagogi IPBL.* Available at: http://www.ipbl.edu.my/portal /penyelidikan/ seminar papers /2008/5a%20Hwa.pdf (Accessed: 27 November 2011).

Young, S. and Giebelhaus, C. (2005) *Formative Assessment and Its Uses for Improving Student Achievement. Education Data Management Solutions, STI.* Available at: www.cbohm.com/news/STI/STI_White_Paper.pdf (Accessed Nov 2011).

Zainuddin, A. (2012) *A Handbook on SEM Structural Equation Modelling*. 6th edn. Kampus Kota Bharu: UiTM Kelantan.

Zalika, A., Faridah, K. and Mohd Johdi, S. (2009) 'Memperkasakan Pedidikan Luar Bandar' *Prosiding Persidangan Kebangsaan Pendidikan Luar Bandar 2009*, Hotel Baverly, Kota Kinabalu, Sabah, 3–5 Februari. University Malaysia Sabah: Sekolah Pembangunan dan Pendidikan Sosial. Available at: http://irep. iium.edu.my/11670/1/Procedng_-_Memperkasakan_ Pend_Luar_Bandar.pdf (Accessed: 3 November 2014).

Zhang, J. (2004) *China's Skill Assessment System. Institute of Population and Labour Economics.* Available at: http://siteresources.worldbank .org/

EDUCATION/Resources/278200-1126210664195/1636971-1126210694253/China_Skills_Assessment.pdf (Accessed: 3 March 2012).

Zhang, G., Zeller, N., Griffith, R., Metcalf, D., Williams, J., Shea, C. and 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), pp. 57-84.

Zimbardo, P. and Leippe, M. (1991). The Psychology of Attitude Change and Social Influence. Philadelphia, PA: Temple University Press.

Zinovieff, M. A. (2008) *UN Task Force*. Available at: http://esa.un.org/techcoop/fellowships/SFOMeeting/ParticipantArea/BackgroundDocuments/6_REVIEW%20report%20FINAL%20.pdf (Accessed: 3 February 2015).

Bibliography