

Development and Validation of Assessment Practice Inventory for Teacher Educators

Siti Eshah Mokshein¹, Othman Lebar¹, Jamal@Nordin Yunus¹, Azali Rahmat¹, Mohd Uzi Dollah¹, Azliza Muhammad, Nor Azah Mansor¹, Azma Mahmood² & Norain Md Noor³

Universiti Pendidikan Sultan Idris, Tanjong Malim, Perak¹

Institut Perguruan Guru Malaysia, Kementerian Pendidikan Malaysia, Cyberjaya, Malaysia²

Institut Perguruan Guru Kampus Sultan Abdul Halim, Kedah, Malaysia³

email: eshah@fppm.upsi.edu.my, othman@fppm.upsi.edu.my

Abstract

The purpose of the study was to develop and validate an assessment practice inventory for teacher educators adapted from Zhang and Burry-Stock (1994, copyright 2003) to be used to evaluate assessment practice among teacher educators. The initial instrument contained 70 items (excluding background information) that were divided into three important sections – practice related to assessment literacy standards (Section B); belief about assessment principles (Section C); and frequency of carrying out described items (Section D). The instrument was administered to 254 teacher educators from a teacher education university and a teacher training institute. Exploratory factor analyses (EFA) and reliability tests were performed on the data. Results showed that the instrument developed yielded high values of internal consistency as reflected by the Cronbach alpha values. Results of EFA suggest that 21 items need to be removed due to their non-dimensionality as they have more or less equal loadings on several factors. Thus, the final draft of the instrument contained 49 items. Even though the reliability and validity of the instrument are within the acceptable range, more data need to be gathered using bigger sample size so that further analysis using Item Response Theory can be used to explore deeper into the psychometric characteristics of the items before the instrument can be finalized.

INTRODUCTION

In its strive to achieve Vision 2020 that is to become a fully industrialized country in its own mould by the year 2020, Malaysia foresees the importance of k-economy that is based on highly knowledgeable and skilful workers who will be able to accelerate the development of the country. For this, the country acknowledges the importance of education in producing the kind of manpower needed to achieve the national goals. This poses a set of new challenges to the education system to provide the most appropriate and high quality curriculum, quality teachers, quality delivery system, quality infrastructure, and quality assessment systems that are able to meet the present and future demands.

A research conducted by the Academy of Leadership for Higher Studies Higher or *Akademi Kepimpinan Pengajian Tinggi (AKEPT)*, however, found that 50 percent of the teachers observed failed to deliver their lessons effectively, particularly, their inability to inculcate higher order thinking skills (Ministry of Education, 2013). The National Education Transformation Plan and the Education Blueprint 2013-2025, in line with the Government Transformation Plan (GTP), which emphasise the need for the 11 shifts have made it imperative for teacher education to focus on transformational change rather than incremental change. Therefore, it is crucial to critically review the existing teacher education programme in order to transform the Teacher Education Model so that it will be able to develop quality teachers for the knowledge era and beyond. Teachers of today and for the future need to be more creative, innovative and able to integrate knowledge across disciplines and domains. Current teacher education model in Malaysia has not been reviewed and researched comprehensively. Thus, a niche research grant study (NRGS) ‘Development of a Teacher Education Model for Preparing Quality Teachers for the Future’ was launched in 2013.

The main aim of the NRGS -Teacher Education Model is to conduct a critical analysis of the existing teacher education models in the context of best practices, national needs and global trends and subsequently develop and validate a new teacher education model. This model will encompass teacher knowledge, skills, and values within the context of reasoning and minds for quality teachers for the future who are capable of making innovative and creative pedagogical decisions in varying contexts. As a starting point, Gardner's model of the Five Minds for the Future, namely Disciplined Mind, Synthesizing Mind, Creating Mind, Respectful Mind and Ethical Mind (Gardner, 2007) will be referred to in constructing the provisional model of teacher education. The provisional model will be a synthesis of the five provisional frameworks. This will serve as the basis for developing the generic attributes of future teachers across levels, disciplines in the context of curriculum content, teaching and learning, assessment, teacher leadership and teacher clinical experience and induction. These attributes of quality teachers for the future would include deep subject matter knowledge, pedagogical content knowledge, and ability to integrate knowledge and engage in creative and innovative teaching in challenging contexts. This will be materialised through the integration of five frameworks developed and validated through five projects. The objectives are to analyze the existing teacher education models in the context of policy, theory and practices of selected local and international teacher education programmes; benchmark the profile of knowledge, skills and values that will form the foundations for constructing a curriculum framework for preparing quality teachers for the future; validate a curriculum framework which encompasses guiding principles for teaching and learning, assessment, teacher leadership, clinical experience and teacher induction; develop curriculum for selected teacher education programmes at the pre-school, primary and secondary levels; validate and refine curriculum for selected teacher education programmes at the pre-school, primary and secondary levels and produce a teacher education model for preparing quality teachers; conduct an impact study to gauge the effectiveness of the newly developed teacher education model for preparing quality teachers; and refine and finalise the new teacher education model. There are five projects that run concurrently under this The NRGS – Teacher Education Model:

- (i) Curriculum structure and content,
- (ii) Teaching and learning,
- (iii) Assessment,
- (iv) Teacher leadership, and
- (v) Clinical experience and teacher induction.

The NRGS longitudinal study covers eight selected programs, namely B. Ed. (Pre-School), B. Ed. (Primary School) and B. Ed. Secondary (Mathematics, Science, TESL, History, Physical and Health Education, Moral Education). The study is divided into four phases over five years:

Phase 1: Preliminary Research, Gap Analysis and Benchmarking - Critical and comparative analysis of selected local and international education models and programmes in the context of policy, theory and practice. It will also involve the benchmarking and profiling of knowledge, skills and values of teacher education that will form the foundation of constructing a framework for a model of teacher education to develop quality teachers for the future.

Phase 2: Monitoring, Validating and Refining Frameworks – Validating and refining the main curriculum framework (including structure and content) and the four frameworks for teaching and learning, assessment and accountability, teacher leadership, clinical experience and teacher induction frameworks. The validation will involve the validating of the frameworks by local experts, as well as local stakeholders and collaborators, plus process-based validation as the framework is being implemented.

Phase 3: Refining and Validating the Teacher Education Model for Preparing Quality Teachers for the Future – This phase will run concurrently with Phase 2. Findings from the research conducted by the five projects will be critically studied and synthesised in order to develop the new model to be referred to as the Teacher Education Model for Preparing

Quality Teachers for the Future. The model will be validated by international and local experts in teacher education and through round-table meetings with the different stakeholders. The reports and findings will serve to validate and refine the final version of the model. The output of this phase is the refined teacher education model validated by international and local experts.

Phase 4: Impact Study on the Effectiveness of the Teacher Education Model for Preparing Quality Teachers for the Future - Investigation of the performance of student teachers who have been trained based on the new Teacher Education Model in clinical experience and performance of beginning teachers during induction period. This will also involve the collection and examination of evidence of school students' work and student teachers performance in teaching in relation to the principles of knowledge, skills and values that have been developed through the different frameworks. A set of criteria will be developed to be used as a basis for measuring the performance of teachers. The output of this phase will be the final Teacher Education Model and attributes of quality teachers for the future.

Background and Rationale

Assessment plays a vital role in any reform. It is conducted to evaluate the needs for a reform, identify the kind of reforms needed, monitor the implementation of the reform, measure the effects and assess the effectiveness of the whole reform to ensure that the investment is worthwhile. In education, assessment is often used for three different purposes – assessment for learning, assessment as learning, and assessment for learning.

While the role of assessment in improving education is vital, there are several critical aspects of education that matter most for student learning. These key factors include well-prepared teachers, well-designed and coherent curriculum, skilful instruction that is adapted to students' needs, and personalized learning environments in which students are well-known by their teachers (Darling-Hammond, 1996). Providing these key features of a sound education is a major foundation of an accountability system.

The center of this accountability system is highly qualified teachers, as teachers are responsible for developing a sound curriculum, implementing successful pedagogies, and designing more personalized schools. Research has found that effective teachers have strong content knowledge and pedagogical training in the field in which they teach, as well as an understanding of students and how they learn (for a review, see Darling-Hammond, 2000). Studies demonstrate that students taught by under-qualified teachers have significantly lower achievement on state reading and mathematics tests after controlling for factors such as poverty levels and the language background of students (Betts, Rueben & Danenberg, 2000; Darling-Hammond, 2000).

Researchers have advocated that classroom assessment should support instruction and enhance students' learning (Shepard, 2000). However, studies show that teachers have consistently used a variety of factors in their assessment practices and consequently make erroneous decisions. Even more disturbing is that most teachers lack effective assessment knowledge and skills; that is, when evaluating student academic achievement, teachers exhibited misconceptions about assessment practices (McMillan, 2001). In short, while many seem to understand assessment, more seem to misunderstand it instead.

In the past, definitions of accountability focused primarily on the interaction of goals, indicators, decision rules, and consequences. Although those components are still central to any accountability system, the latest movement has been to focus more on capacity building and providing appropriate supports. That is, the purpose of accountability is not simply to identify and punish ineffective educational institutions, but to provide appropriate supports to ensure that all educational institutions are effective.

Accountability occurs when policies and practices work together to provide good education and to correct problems as they occur. Accountable systems increase the probability of high-quality practice, leading to positive outcomes. They reduce the probability of malpractice or educational harm, and they call attention to problems and needs. Furthermore, accountability must be two-way: If

students are accountable for learning to certain standards, schools, districts, and states must be accountable for providing them with the necessary resources for learning.

Teacher education program that are accountable to students will ensure that they have well-qualified teacher educators in adequately resourced institutions that are designed to support teacher educators in providing good instruction. Assessments of learning and other indicators of institutions conditions can help evaluate the extent to which educational goals are accomplished. But the focus of accountability must be kept on what is needed actually to improve achievement as well as on how progress is to be measured.

The above discussion above raised several questions about assessment such as to what extent teacher educators used constructive alignment approach in planning and designing assessment of student learning, to what extent current practice of teacher education (internal and international) engaging students in meaningful learning via effective assessment practice, and what assessment and accountability model and system is effective in assessing student teachers comprehensively in a variety of domains including cognitive, affective and psychomotor.

The research project 'Developing and validating an assessment and accountability framework for preparing quality teachers for the future' under the national NREGS – Teacher Education Model has the following objectives:

- i. To analyze the existing assessment and accountability framework in the context of policy, theory and practices of selected local and international Teacher Education Programmes.
- ii. To benchmark the profile of knowledge, skills and values that will form the foundations for constructing an assessment and accountability framework for preparing quality teachers for the future.
- iii. To validate the assessment and accountability framework with international experts, local stakeholders in the context of practice and contribute to the development of curriculum framework
- iv. To determine assessment and accountability strategies for selected teacher education programmes at the pre-school, primary and secondary levels.
- v. To validate and refine the assessment and accountability strategies for selected teacher education programmes at the pre-school, primary and secondary levels.
- vi. To conduct an impact study to gauge the effectiveness of the assessment and accountability framework.
- vii. To refine and finalise the assessment and accountability framework of assessing teachers for the future and contribute to the development of the Teacher Education Model.

To meet the first objective of the project, an instrument has to be developed to gather data on the current assessment practices of teacher educators, their beliefs about assessment, and the assessment principles they uphold when conducting assessments.

Research Objectives

The purpose of this study was to develop and validate an assessment practice inventory for teacher educators in Malaysia to be used in the project for developing and validating an assessment and accountability framework for preparing quality teachers for the future. It sought to explore teachers' knowledge about assessment, their beliefs about assessment, their assessment practice, and competency in conducting assessment. Specifically, this study sought to:

- i. develop an assessment practice inventory for teacher educators;
- ii. establish the validity of the instrument; and
- iii. establish the reliability of the instrument.

METHODOLOGY

This survey was piloted to 254 teacher educators from a teacher education institute (IPG) and an education university in the country. The instrument developed was tried out to 20 teacher educators to check for the appropriateness of the language and content. Analysis showed that all the items in the instrument developed seen to be appropriate by the respondents. However, several phrases in the instrument were found to be ambiguous and needed rephrasing. The final draft of the instrument was administered to the 254 respondents from the eight teacher education programs that were selected through stratified random sampling technique.

Development of the Instrument

The instrument was developed based on the nine principles of best practices in educational assessment (American Association of Higher Education, 1996), Classroom Assessment Practices and Teachers' Self-Perceived Assessment Skills (Zhang & Burry-Stock, 2003) and Assessment of Students' Learning: Practice among Malaysian Teachers (Suah See Ling, Ong Saw Lan & Shuki Osman, 2009).

Literature from best practices in educational assessment highlighted nine important principles (American Federation of Teachers, 2009), of which seven were picked up and further developed in the assessment project to be matched with the essential values, skills, and knowledge. The seven principles and related attributes to be developed in the project are shown in Table 1.

Zhang and Burry-Stock (1994, copyright 2003) developed an instrument to investigate teachers' assessment practices across education levels and content areas, as well as teachers' self-perceived assessment skills as a function of teaching experience and measurement training. Data from 297 teachers on the Assessment Practices Inventory were analyzed in a MANOVA design. They found that as grade level increases, teachers rely more on objective tests in classroom assessment and show an increased concern for assessment quality ($p < .001$). Across content areas, teachers' involvement in assessment activities reflects the nature and importance of the subjects they teach ($p < .001$).

Table 1 Assessment Guiding Principles and Accountability

	GUIDING PRINCIPLES	KNOWLEDGE	VALUES	SKILLS
1.	The assessment of student learning begins with educational values.	<ul style="list-style-type: none"> ○ The teacher understands the assessment policy and practice ○ The teacher understands relevant learning theories ○ The teacher understands relevant assessment theories and its relationship to learning outcomes 	<ul style="list-style-type: none"> ○ The teacher acknowledges assessment as a tool to measure educational goals ○ The teacher recognizes assessment as a vehicle for educational improvement – policy, teaching, learning, innovations, reforms 	<ul style="list-style-type: none"> ○ The teacher utilizes multiple assessment methods (eg. test, performance assessment) to measure multiple domain of learning ○ The teacher applies methods of assessment that cater for student diversity and differences
2.	Assessment is most effective when it reflects an understanding of learning as	<ul style="list-style-type: none"> ○ The teacher understands learning as multidimensional ○ The teacher analyzes multiple assessment 	<ul style="list-style-type: none"> ○ The teacher acknowledges multiple approaches and methods in teaching 	<ul style="list-style-type: none"> ○ The teacher utilizes multiple assessment methods (eg. test, performance assessment) to

	GUIDING PRINCIPLES	KNOWLEDGE	VALUES	SKILLS
	multidimensional, integrated, and revealed in performance over time	<p>methods and techniques to measure learning</p> <ul style="list-style-type: none"> ○ The teacher understands constructive alignment between learning outcomes, teaching and learning and assessment 	<ul style="list-style-type: none"> ○ The teacher appreciates multiple assessment in assessing students holistically 	<p>measure multiple domain of learning</p> <ul style="list-style-type: none"> ○ The teacher applies methods of assessment that cater for student diversity and differences
3.	Assessment works best when the programs it seeks to improve have clear, explicitly stated purposes	<ul style="list-style-type: none"> ○ The teacher understands levels of learning taxonomy (Cognitive, Affective and Psychomotor) ○ The teacher understands multiple purpose of assessment ○ The teacher understands how to select and develop assessment methods appropriate to the level of students ability ○ The teacher understands constructive alignment between learning outcomes, teaching and learning and assessment 	<ul style="list-style-type: none"> ○ The teacher recognises the importance of accurate assessment method and interpretation ○ The teacher acknowledges the multiple purpose of assessment ○ The teacher realises the importance of alignment between learning outcomes, teaching and learning and assessment 	<ul style="list-style-type: none"> ○ The teacher applies constructive alignment in designing assessment tasks ○ The teacher uses feedback to improve own teaching and students learning and achievement ○ The teacher selects and constructs multiple methods of assessment appropriate for decision making
4.	Assessment requires attention to outcomes but also equally to the experiences that lead to those outcomes.	<ul style="list-style-type: none"> ○ The teacher understands the difference between formative and summative assessment ○ The teacher understands authentic and performance assessment ○ The teacher understands assessment for Learning and Assessment of Learning 	<ul style="list-style-type: none"> ○ The teacher recognises the importance of continuous assessment ○ The teacher appreciates the importance of feedback in student learning ○ The teacher acknowledges the appropriate statistics in measurement of student learning 	<ul style="list-style-type: none"> ○ The teacher conducts formative and summative assessment in assessing students performance ○ The teacher performs authentic assessment to measure higher order thinking skills ○ The teacher utilizes appropriate scoring rubrics in assessment task

	GUIDING PRINCIPLES	KNOWLEDGE	VALUES	SKILLS
		<ul style="list-style-type: none"> ○ The teacher understands basic statistics in testing and measurement ○ The teacher understands and apply data analysis and interpretations ○ The teacher understands the scoring rubrics 		
5.	Assessment works best when it is ongoing not episodic	<ul style="list-style-type: none"> ○ The teacher understands the difference between formative and summative assessment ○ The teacher understands assessment for learning and assessment of learning 	<ul style="list-style-type: none"> ○ The teacher recognises the importance of continuous assessment ○ The teacher appreciates the importance of feedback in student learning 	<ul style="list-style-type: none"> ○ The teacher conducts formative and summative assessment in assessing students performance ○ The teacher monitors students learning through continuous assessment
6.	Assessment fosters wider improvement when representatives from across the educational community are involved	<ul style="list-style-type: none"> ○ The teacher understands multiple purpose of assessment ○ The teacher uses a variety of reporting techniques ○ The teacher understands how to communicate assessment results to various group of people 	<ul style="list-style-type: none"> ○ The teacher recognises the purpose of assessment results ○ The teacher realises the importance of assessment results to various group of people 	<ul style="list-style-type: none"> ○ The teacher performs multiple strategies in communicating assessment results ○ The teacher prepares assessment reports appropriate to the demand of various audiences
7.	Through assessment, educators meet responsibilities to students and to the public	<ul style="list-style-type: none"> ○ The teacher understands the concept of assessment and accountability ○ The teacher understands the basic principles of quality assessment ○ The teacher understands the 	<ul style="list-style-type: none"> ○ The teacher recognises the importance of accurate reporting of assessment results ○ The teacher acknowledges unethical practice in assessment ○ The teacher beliefs in 	<ul style="list-style-type: none"> ○ The teacher uses assessment results to inform students and publics on students performance as accurate as possible ○ The teacher utilizes assessment results to improve students learning

	GUIDING PRINCIPLES	KNOWLEDGE	VALUES	SKILLS
		mechanism of conducting quality assessment (monitoring, moderation, calibration) ○ The teacher understands unethical practice in assessment	valid and reliable assessment results	○ The teacher conducts assessment according to the appropriate principles of quality assessment (valid, reliable, not biased)

Regardless of their teaching experiences, teachers with measurement training report a higher level of self-perceived assessment skills in using performance measures; standardized testing, test revision, and instructional improvement; as well as in communicating assessment results ($p < .05$) than those without measurement training. The implications of the results for measurement training are also discussed.

Zhang and Burry-Stocks' (1994, copyright 2003) inventory contains 67 items that address issues in classroom assessment of student learning. The instrument used two types of scaling:

- a) To indicate how frequently respondents use the assessment practice described by the item, the following scales were used:

1 = not at all used,
 2 = seldom used,
 3 = used occasionally,
 4 = used often, and
 5 = used very often

- b) To indicate respondent's competence in carrying out the practice described by the items, the following scales were used:

1 = not at all skilled,
 2 = a little skilled,
 3 = somewhat skilled,
 4 = skilled, and
 5 = very skilled

The 67 items in Zhang and Burry-Stocks' Teacher Assessment Practice Inventory (2003) cover a wide range of assessment practices as follow:

Components in Zhang's Assessment Practice & Skills Inventory

1. Choosing appropriate assessment methods for instructional decisions.
2. Selecting textbook-provided test items for classroom assessment.
3. Revising previously produced teacher-made tests to match current instructional emphasis.

4. Administering announced quizzes.
5. Administering unannounced quizzes.
6. Evaluating oral questions from students.
7. Assessing students through observation.
8. Determining if a standardized achievement test is valid for classroom assessment.
9. Using a table of specifications to plan assessments.
10. Developing assessments based on clearly defined course objectives.
11. Matching assessments with instruction.
12. Writing paper-pencil tests.
13. Writing multiple-choice questions.
14. Writing matching questions.
15. Writing true/false questions.
16. Writing fill-in-the-blank or short answer questions.
17. Writing essay questions.
18. Writing test items for higher cognitive levels.
19. Constructing a model answer for scoring essay questions.
20. Ensuring adequate content sampling for a test.
21. Matching performance tasks to instruction and course objectives.
22. Defining a rating scale for performance criteria in advance.
23. Communicating performance assessment criteria to students in advance.
24. Recording assessment result on the rating scale/checklist while observing a student's performance.
25. Using concept mapping to assess student learning.
26. Assessing individual class participation.
27. Assessing group class participation.
28. Assessing individual hands-on activities.
29. Assessing group hands-on activities.
30. Assessing individual class participation.
31. Using portfolios to assess student progress.
32. Following required procedures (time limit, no hints, no interpretation) when administering standardized tests.
33. Interpreting standardized test scores (e.g., stanine, percentile rank) to students and parents.
34. Interpreting Percentile Band to students and parents.
35. Calculating and interpreting central tendency and variability for teacher-made tests.
36. Conducting item analysis (i.e., difficulty and discrimination indices) for teacher-made tests.
37. Revising a test based on item analysis.
38. Obtaining diagnostic information from standardized tests.
39. Using assessment results when planning teaching.
40. Using assessment results when developing curriculum.
41. Using assessment results when making decisions (e.g., placement, promotion) about individual students.
42. Using assessment results when evaluating class improvement.
43. Using assessment results when evaluating school improvement.
44. Developing systematic grading procedures.
45. Developing a grading philosophy.
46. Using norm-referenced grading model.
47. Using criteria-referenced grading model.
48. Using systematic procedures to determine borderline grades.
49. Informing students in advance how grades are to be assigned.
50. Establishing student expectations for determining grades for special education students.
51. Weighing differently projects, exams, homework, etc. when assigning semester grades.
52. Incorporating extra credit activities in the calculation of grades.
53. Incorporating ability in the calculation of grades.
54. Incorporating classroom behavior in the calculation of grades.
55. Incorporating improvement in the calculation of grades.

56. Incorporating effort in the calculation of grades.
57. Incorporating attendance in the calculation of grades.
58. Assigning grades.
59. Providing oral feedback to students.
60. Providing written feedback to students.
61. Communicating classroom assessment results to students.
62. Communicating classroom assessment results to parents.
63. Communicating classroom assessment results to other educators.
64. Avoiding teaching to the test when preparing students for tests.
65. Protecting students' confidentiality with regard to test scores.
66. Recognizing unethical, illegal, or otherwise inappropriate assessment methods.
67. Recognizing unethical, illegal, or otherwise inappropriate uses of assessment information.

Suah See Ling, Ong Saw Lan & Shuki Osman (2009) in obtaining information on school teachers' assessment practice in Malaysia developed an instrument adapted from Zhang and Burry-Stock. The inventory consists of three sub-sections namely information about teachers' background, training and knowledge on assessment, and assessment practices implemented by school teachers. The sample for the study was 602 teachers from the northern region of Malaysia. Data from the study were analyzed by calculating the mean values of the responses and percentage of respondent's practices. The results showed that the form of assessment frequently used by school teachers was multiple-choice objective test. There were significant differences among teachers from different school levels for aspects like developing marking scheme, giving feedbacks of evaluation results and the use of written test and the use of other strategies. Comparison among teachers teaching different subject area showed significant difference only in written test.

Items in Suah's (Year) instrument were organized into several components as below:

- (1) Test Development
 - a) Test Procedure (5 Items)
 - b) Sources of Test Development (6 Items)
 - c) Higher Order Thinking (6 Items)
- (2) Types of Assessment
 - a) Written Test (6 Items)
 - b) Performance Assessment (5 Items)
 - c) Other Strategies (5 Items)
- (3) Use of Assessment Results
 - a) Formative Assessment (7 Items)
 - b) Summative Assessment (3 Items)
- (4) Scoring and Grading (10 Items)
- (5) Scoring Scheme (6 item)
- (6) Providing feedback of assessment results (4 Items)

The newly developed assessment practice inventory in this study consists of 78 items (including demographic information) that were organized in five sections as below:

- A. Demography (8 Items)
- B. Assessment literacy standards (10 Items)
- C. Beliefs about assessment (9 Items)
- D. DU. Frequency in conducting described items (51 Items)
- E. DS. Competence in conducting described items (51 Items)

Sections D and E contain similar items but different responses, one in terms of frequency while the other one on competence in carrying out those described items based on their perception.

The assessment literacy standards covered in this instrument were developed based on the assessment literacy standards developed by the American Federation of Teachers (2009) as below:

- Standard 1: Teachers should be skilled in choosing assessment methods appropriate for instructional decisions.
- Standard 2: Teachers should be skilled in developing assessment methods appropriate for instructional decisions.
- Standard 3: The teacher should be skilled in administering, scoring and interpreting the results of both externally-produced and teacher-produced assessment methods.
- Standard 4: Teachers should be skilled in using assessment results when making decisions about individual students, planning teaching, developing curriculum, and school improvement.
- Standard 5: Teachers should be skilled in developing valid pupil grading procedures which use pupil assessments.
- Standard 6: Teachers should be skilled in communicating assessment results to students, parents, other lay audiences, and other educators.
- Standard 7: Teachers should be skilled in recognizing unethical, illegal, and otherwise inappropriate assessment methods and uses of assessment information.

Since this instrument is not an achievement test or speeded test, time is not a critical element in the administration of the questionnaire. Respondents were given sufficient time to complete the questionnaire until they were satisfied with the responses given. The questionnaires were then gathered and data were entered into spreadsheets using statistical package for social science (SPSS). Data analysis involved computation of descriptive statistics and examination of psychometric characteristics of the items using internal consistency measures and factor analysis.

FINDINGS AND DISCUSSIONS

The findings discussed in this paper are organized around two important aspects of instrument development, namely reliability and validity of the instruments.

Validity of the Instrument

Factor analysis was used to establish the validity of the instrument. At the initial stage, KMO Bartlett test was performed to check on the adequacy of the data and a statistic value of 0.947 was obtained. This indicates that 94.7 percent of the variable properties are explained by the data and factor analysis would be meaningful. Next, exploratory factor analysis (EFA) was performed on the data by section to identify the number of constructs and grouping of items for each construct. EFA was performed on all sections of the instrument except on section A (Demography).

EFA on Section B (Assessment Literacy Standards) has yielded two factors. Items related to selecting appropriate assessment methods; developing assessment items and/or tasks; and administering assessment tasks were group together as one factor, while items related to the using assessment results; reporting assessment results; analysis of items; and interpretation of assessment results fall into the second factor. Two items (Item B5 and B10), however were grouped into both constructs with more or less equal loadings and thus would be removed from the final instrument. Factor loadings for each factor were greater than 0.6. Detailed results of EFA for Section B are presented in Table 2.

Table 2 Results of EFA on Practice Related to Assessment Literacy Standards

Rotated Component Matrix ^a			Reliability (Cronbach alpha)
SECTION B: PRACTICE RELATED TO ASSESSMENT LITERACY STANDARDS (8/10 items selected)	Component		
		1	2
B.C1: (Choosing, Constructing, Administering Assessment)			.863
Choosing assessment method for instructional decision (b1)	.867		
Developing assessment method for instructional decision (b2)	.862		
Choosing appropriate scoring method with forms of assessment used (b4)	.717		
Administering multiple forms of assessment (b3)	.694		
Determining grading methods with form of assessments used (b5)	.644	.501	
B.C2: (Using assessment results, reporting results, analyzing and interpreting)			.786
Using assessment results in making decision on individual student.(b8)		.819	
Communicating assessment results to students and parents. (b7)		.762	
Using assessment results in improving teaching and learning. (b9)		.641	
Analyzing and interpreting test score and grade. (b6)		.616	
Recognizing unethical and illegal use of assessment results (b10)	0.466	.526	

Notes: Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
a. Rotation converged in 3 iterations.

The two constructs extracted from section B seem to suggest assessment literacy standards practiced by teacher educators can be grouped into two categories. One category is related to the processes before and during administering assessment, which include items related to selecting appropriate assessment methods; developing assessment items and/or tasks; and administering assessment tasks. Another category is related to the processes after assessment task which include using assessment results; reporting assessment results; analysis of items; and interpretation of assessment results. These constructs seem logical and items to fall nicely under the constructs.

For the nine items in Section C (Beliefs about assessment principles), EFA has yielded one single construct and all items obtained good loadings. The internal consistency as shown by the Cronbach alpha was 0.900. Details of the results are shown in Table 3.

Table 3 Results of EFA on Section C (Beliefs about Assessment)

Component Matrix ^a	
SECTION C: BELIEFS ABOUT ASSESSMENT PRINCIPLES (9/9 Items selected)	Component
	1
Assessment is most effective when it reflects an understanding of learning as multidimensional, integrated, and revealed in performance over time	.814

Pattern Matrix ^a										Reliability
SECTION D – FREQUENCY OF IMPLEMENTING DESCRIBED ITEMS (42/51 Items selected)	Component									
	1	2	3	4	5	6	7	8	9	
Developing systematic grading procedures (d34).	.320									
(D.C2 – Constructing Item (principles, alignment); choosing methods)										0.866
Matching assessments with instruction (d7).		.713								
Developing assessments based on clearly defined course objectives (d6).		.677								
Choosing appropriate assessment methods for instructional decisions (d1)		.642								
Matching performance tasks to instruction and course objectives (d13)		.532								
Writing various forms of objective test (d8)		.498								
Communicating performance assessment criteria to students in advance (d15)		.465								
Recording assessment result on the rating scale/checklist while observing a student's performance (d16)		.465								
Revising previously produced teacher-made tests to match current instructional emphasis (d3).		.453							-.387	
Ensuring adequate content sampling for a test (d12)		.372			-.312					
(D.C3 – Incorporating behaviour, effort, attendance, ability and improvement in assessment)										0.890
Incorporating classroom behavior in the calculation of grades (d41)			-.775							
Incorporating effort in the calculation of grades (d43)			-.770							
Incorporating attendance in the calculation of grades (d44)			-.766							
Incorporating improvement in the calculation of grades (d42)			-.729							
Incorporating ability in the calculation of grades (d40)			-.612							

Pattern Matrix ^a										Reliability
SECTION D – FREQUENCY OF IMPLEMENTING DESCRIBED ITEMS (42/51 Items selected)	Component									
	1	2	3	4	5	6	7	8	9	
Weighing differently projects, exams, homework, etc. when assigning semester grades (d39)			-.354							
(D.C4 - Inclusion of hands-on activities, student participation in assessment)										0.925
Assessing individual hands-on activities (d20)				.733						
Assessing group hands-on activities (d21)				.704						
Assessing group class participation (d19)				.654						
Assessing individual class participation (d18)				.613						
Using portfolios to assess student progress (d22)	.348			.391	-.306					
(DU.C5 - Preparing rubrics, writing essay questions, use test specification table in developing test)										0.765
Constructing a model answer for scoring essay questions (d11)					-.749					
Writing essay questions (d9)					-.662					
Writing test items for higher cognitive levels (d10)		.349			-.476					
Defining a rating scale for performance criteria in advance (d14)					-.376					
Using a table of specifications to plan assessments (d5)		.343			-.365					
(D.C6 – Assessment ethics)										0.807
Recognizing unethical, illegal, or otherwise inappropriate assessment methods (d50)						-.915				
Recognizing unethical, illegal, or otherwise inappropriate uses of assessment information (d51)						-.883				
Protecting students' confidentiality with regard to test scores (d49)						-.645				
Following required procedures (time limit, no hints, no interpretation) when administering standardized tests (d23)						-.308				

Pattern Matrix ^a										Reliability		
SECTION D – FREQUENCY OF IMPLEMENTING DESCRIBED ITEMS (42/51 Items selected)	Component											
	1	2	3	4	5	6	7	8	9			
(D.C7 – Dissemination of assessment results, provide feedback, the use of norm-referenced grading)											-0.045	
Communicating classroom assessment results to other educators (d48)							.653					
Communicating classroom assessment results to students (d47)												
Providing written feedback to students (d46)												
Using norm-referenced grading model. Use norm reference model in grading students (d35)												
Informing students in advance how grades are to be assigned (d38)		.323										
Using criteria-referenced grading model (d36)	.358											
Providing oral feedback to students (d45)												
(D.C8 – Selecting items from textbooks, assessing through observation)												
Selecting items from textbooks (d2)												
Assessing students through observation (d4)				.439								
(D.C 9 – Use assessment results for evaluating improvement, use systematic procedure)											0.781	
Using assessment results when evaluating class improvement (d32)												
Using systematic procedures to determine borderline grades (d37)												
Using assessment results when evaluating school improvement (d33)												

Notes: Extraction Method: Principal Component Analysis.

Rotation Method: Oblimin with Kaiser Normalization.

a. Rotation converged in 14 iterations.

Correlation between items i and j is a function of loadings of items i and j on the factors underlying the items. This is shown in the following equation:

$$\rho_{ij} = \sum_{k=1}^m a_{ik}a_{jk} ; \text{ where}$$

a_{ik} = loading of item i on factor k

a_{jk} = loading of item j on factor k

For items i and j with two factors for instance,

$$\rho_{ij} = a_{i1}a_{j1} + a_{i2}a_{j2} ; \text{ where}$$

a_{i1} = loading of item i on factor 1

a_{j2} = loading of item j on factor 2

Factor loadings indicate the strength and magnitude of the relationship between the items and the factors. Positive loadings will yield positive correlation between items i and j , while negative loadings may yield negative correlation between the two items. Thus, both negative and positive loadings are equally important because they reflect the nature of the relationship.

(D.C3 – Incorporating behaviour, effort, attendance, ability and improvement in assessment)

Reliability of the Instrument

Analyses of validity using exploratory factor analysis by construct for all sections in the instrument have yielded high reliability measures except for construct 7 section D (Frequency in carrying out describe items related to dissemination of assessment results, provide feedback, use norm-referenced grading). The reliability values were greater than 0.76 for all constructs in the instrument except for construct 7 section D which value was 0.045. This indicates inconsistency of the responses to the items under this construct. Thus, the other four items under this construct have to be re-examined and might be removed as well. This makes the total number of items to be removed 21. Details of reliability index by construct are shown in Table 5.

Table 5 Reliability of the Instruments by Construct

Constructs	Items	Reliability (Cronbach alpha)
B. Practice related to assessment literacy standards B.C1: Choosing, Constructing, Administering Assessment (4 Items) B.C2: Using assessment results, reporting results, analyzing and interpreting results (5 Items)	b1, b2, b3, b4 b6, b7, b8, b9, b10	0.863 0.786
C. Beliefs about assessment (9 Items)	c1, c2, c3, c4, c5, c6, c7, c8, c9	0.900
D. Frequency in carrying out described items D.C1: Item analysis, use of assessment results, grading procedures (10 Items)	d26u, d27u, d25u, d24u,d28u,d30u,d29u,d17u, d31u,d34u	0.929
D.C2: Constructing Item (principles, alignment); choosing methods (7 Items)	d7u, d6u, d1u, d13u, d8u, d15u, d16u	0.866
D.C3: Incorporating behaviour, effort, attendance, ability and improvement in	d41u, d43u, d44u, d42u, d40u, d39u, d20u	0.890

Constructs	Items	Reliability (Cronbach alpha)
assessment (7 Items)		
D.C4: Beliefs about assessment principles (3 Items)	d21u, d19u, d18u	0.925
D.C5: Preparing rubrics, writing essay questions, use JSU in developing test (3 Items)	d11u, d9u, d14u	0.765
D.C6: Assessment ethics (4 Items)	d50u, d51u, d49u, d23u	0.807
D.C7: Dissemination of assessment results, provide feedback, use norm-referenced grading (4 Items)	d48u, d47u, d45u, d35u	-0.045
D.C8: Selecting items from textbooks (1 Item)	d2u	-
D.C9: Use assessment results for evaluating improvement, use systematic procedure (3 Items)	d32u, d37u, d33u	0.781

CONCLUSIONS AND RECOMMENDATIONS

Analysis of validity by exploratory factor analysis and internal consistency on the data showed that the instrument seem to be sound and can be used to measure assessment practice related to assessment literacy standards, beliefs about assessment principles, and frequency of carrying out described items related to assessment. However, results of the analyses suggest that 19 items should be removed from the instrument. They were two (2) items from section B and fifteen (15) items from section D. Further examination of reliability indices showed that Construct 7 from section D (Dissemination of assessment results, provide feedback, use norm-referenced grading) yielded very low reliability index. Thus this construct and the four items underneath might be deleted from the final instrument. From the analysis conducted, 21 items will be removed, leaving 49 items in the final draft of the instrument.

Even though the reliability and validity of the final draft of the instrument are within the acceptable range, some of the items suggested to be removed in this analysis could be important for the assessment project to evaluate assessment practice among teacher educators. Therefore, a more detailed analysis with a larger sample (>1000) using Item Response Theory (IRT) model need to be conducted before the instrument can be finalized. The use of IRT will allow the researchers to explore deeper into the psychometric characteristics of each item and thus provide them with a higher sense of confidence to keep only important items in the instrument.

REFERENCES

- Alexander W. Astin, Trudy W. Banta, K. Patricia Cross, Elaine El-Khawas, Peter T. Ewell, Pat Hutchings, Theodore J. Marchese, Kay M. McClenney, Marcia Mentkowski, Margaret A. Miller, E. Thomas Moran, Barbara D. Wright. (2003). 9 principles of good practice for assessing student learning. American Association for Higher Education (AAHE) Assessment Forum.
- American Federation of Teachers, National Council on Measurement in Education, National Education Association. (2009). Assessment Literacy Standard.
- Betts, J. R., Rueben, K. S., & Danenberg, A. (2000). Equal resources, equal outcomes? The distribution of school resources and student achievement in California. San Francisco: Public Policy Institute of California.
- Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Education Policy Analysis Archives*, 8(1).
- Gardner, H. (2013). *Five minds for the future*. Boston: Harvard Business School Press.
- McMillan, J. H. (2001). Secondary teachers' classroom assessment and grading practices. *Educational Measurement: Issues and Practices*, 20, 20-32.

- Ministry of Education Malaysia. (2013). *Malaysia Education Blueprint 2013-2025 (Preschool to Post-Secondary Education)*. Putrajaya: Kementerian Pendidikan Malaysia.
- Shepard, L. A. (2000). *The role of classroom assessment in teaching and learning* (CSE Tech. Report 517). Los Angeles, CA: University of California.
- Suah See Ling, Ong Saw Lan & Shuki Osman. (2009). Pentaksiran Pembelajaran Pelajar: Amalan Guru-guru di Malaysia. *Majlis Dekan Pendidikan Malaysia*, 5(6).
- Zhang, Z., & Burry-Stock, J. (2003). Classroom assessment practices and teachers' self-perceived assessment skills. *Applied Measurement in Education*, 16(4), 323-342.