

Validity of Teacher-Made Assessment: A Table of Specification Approach

Lei Mee Ing¹, Mohammed Borhandden Musah¹, Shafeeq Hussain Vazhathodi Al-Hudawi¹, Lokman Mohd Tahir¹
& Naail Mohammed Kamil²

¹ Department of Educational Foundation and Social Science, Faculty of Education, Universiti Teknologi Malaysia, Skudai, Malaysia

² School of Business, SEGI University, Taman Sains Selangor, Malaysia

Correspondence: Mohammed Borhandden Musah, Department of Educational Foundation and Social Science, Faculty of Education, Universiti Teknologi Malaysia, Malaysia. E-mail: borhandden@utm.my

Received: October 11, 2014 Accepted: November 18, 2014 Online Published: February 12, 2015

doi:10.5539/ass.v11n5p193

URL: <http://dx.doi.org/10.5539/ass.v11n5p193>

Abstract

The validity of teacher-made assessment remains debatable in the educational assessment process. This study investigates the content validity of teacher-made assessment in three Chinese Elementary Schools in Johor, Malaysia. It also examines teacher understanding of table of specification in the sampled schools. A questionnaire with 10 items was distributed to 30 teachers in order to collect the data on table of specification. Items 1 to 4 examine teacher understanding of the table of specification while items 5 to 10 test the content validity of teacher-made assessment. The results showed that teachers exhibited a low understanding of the table of specification. The analysis revealed that the majority of them never attended courses concerning table of specification and were unable to build a comprehensive table of specification for the subjects they teach. The findings also demonstrated that teacher-made assessment was valid in terms of content validity. However, most of the teachers did not refer to the table of specification while building instruments for assessment. This indicates that teachers lack basic knowledge in designing a standard table of specification and they lack awareness on the importance of the table of specification. Recommendations of the study for teacher-made assessment improvements were also addressed.

Keyword: table of specifications, content validity, teacher-made assessment, subject-matter expert, elementary schools

1. Introduction

It is undeniable fact that teachers often evaluate student performance based on students' test scores. Students' test scores are often cumulative of scores of different assessment on contents that they go through during a particular time frame. Since these assessments determine the career directions of different individual learners, teachers must ensure that the assessment instruments are valid and relevant (Griswold, 1990). Concerns about ensuring the validity of the assessments remain vibrant in educational development. Chan and Gurnam (2006) reported that there is a global move to decentralise assessments with the introduction of authentic alternative forms of assessments that are continuous and formative in nature. Malaysia among many countries has taken a transformative approach towards the matter.

The education system in Malaysia was more test-oriented in assessing the student performances until the mid-nineties. Recently, new directions are being taken by the Ministry of Education (MoE) Malaysia. The Prime Minister of Malaysia, Najib bin Tun Abdul Razak stated that "In order to meet our high aspirations amidst in an increasingly competitive global environment, we cannot stand still. Our country requires a transformation of its entire education system," (Preliminary Report Malaysia Education Blueprint 2012-2025, 2013, p. 52). The MoE thus implemented the School-Based Assessment (SBA) into the education system of Malaysia in order to replace the test-oriented assessment of students' performances with SBA.

SBA had been introduced into all elementary schools in Malaysia in the year of 2011 and all the secondary schools in 2012. The former Malaysian Minister of Education, Tan Sri Musa Mohamed once quoted to have said on May 7, 2003: "We need a fresh and new philosophy in our approach to exams ... we want to make the

education system less exam-oriented and [we] are looking at increasing SBA as it would be a better gauge of students' abilities" (Fook & Sidhu, 2006, p. 2).

In SBA, student performances are assessed in the cognitive, affective and psychomotor aspects in line with the National Philosophy of Education (NPE), Malaysia. Development of these aspects are key to the balanced and holistic growth of student potentials. However, a study by Al-Hudawi et al. (2014) indicate that "major absence across the NPE element was found in 'developing the potentials of individual' according to student perspectives" and in "education is an ongoing effort' with reference to both student and teacher perspectives" (p. 65). Therefore, they call for teachers to extend their focus and their role in facilitating for students' holistic development. This in turn indicates that the role of teachers in the new assessment system is vital, not only in assessing their performance, but also in effective instructional planning, where teacher-made assessment (TMA) plays a major role (Majid, 2011). This is particularly important since a research study by Mehrens & Lehmann (1987) states that more than half of the tests used in their classes are constructed by teachers. Since many decisions are made based on the student test scores, therefore, teachers must ensure that the tests are relevant and reliable (Philips, 1990). In other words, assessment should be a fairly precise measuring tool.

However, students raise complaints that sometimes results of tests are not so dependable that students obtain scores that are either higher or lower than they really ought to be (Frisbie, 1988). Learners often feel that assessment in some instances is completely or partially unrelated to the contents that they learned in the class.

To be more specific, there is a mismatch between the content examined in class and the material assessed at the end of unit test (Fives & DiDonato-Barnes, 2013). They further state that this may lead to the discredibility of assessment tools to provide evidence, based on which teachers make valid judgments about students' progress. Developing a Table of Specifications (TOS) is, thus, a strategy that educators can use to overcome this problem. However, findings of A. Majid (2011) reach the conclusion that teachers in Malaysia lack knowledge and skills in conducting SBA such as in the oral English assessment despite the availability of the guidelines and objectives concerning TOS.

With regards to the implementation of SBA in the education system, the Chinese Elementary Schools in Malaysia have given teachers authority to assessing their students' performance. However, whether the TMAs are reliable and whether they fulfil the content validity was not studied. Hence, this study investigated the content validity of TMAs and examined teachers' understanding of the TOS in three Chinese Elementary Schools in Malaysia.

1.1 Assessment

Assessment is defined as an opinion or judgement about somebody or something that has been thought about very carefully (Oxford Dictionary, 2010). It is a common practice in schools to determine whether a previously taught knowledge has been learned by the students during the instruction in the class. Given this definition, assessment is the systematic collection of data about what students know, understand and are able to do in relation to students' achievement of particular goals of learning. It is a matter of fact that instruments used to collect such data should meet clear criteria open to public scrutiny (Pratt, 1980). Brualdi (1998) stated that, performance based assessment represents a set of strategies for the application of knowledge, skills, and work habits through the performance of tasks that are meaningful and engaging students. It provides teachers with the information about how a child understands and applies knowledge (Brualdi, 1998).

1.2 Validity

Validity is an important criterion for TMA. According to Grimm and Yarnold (2006), Micheal and Burlingame (1996), Pedhazur and Schmelkin (1991), validity can be defined as the extent to which an instrument measures what it supposed to measure. Validity consists of content validity, face validity, criterion-related validity (or predictive validity) and construct validity (Micheal & Burlingame, 1996). Micheal and Burlingame further posited that the content validity is the degree to which the instrument fully assesses the construct of interest, face validity measures the characteristic or trait of interest, criterion-related validity is assessed when one is interested in determining the relationship between scores and a test with reference to specific criterion while the construct validity is the degree to which an instrument measures the trait or theoretical construct that it is intended to be measured. Of the types of validity mentioned, this study employs content validity with particular reference to TMA.

1.3 Content Validity

Content validity commonly has been held to be the most important type of validity that is needed for criterion-referenced measures (Linn, 1980). It involves two major concepts that are the content relevance and

content coverage (Bachman, 1990). According to Yen-Fen (2004), content relevance refers to the extent to which the aspects of the ability to be assessed are actually tested by the task, indicating the requirement to specify the ability domain and the test method facets. Content coverage concerns with the extent to which the test tasks adequately demonstrate the performance in the target context, which may be achieved by randomly selecting representative samples (Yen-Fen, 2004).

1.4 Table of Specification

A TOS is defined as a test blueprint which helps teachers to align objectives, instruction, activity and assessment (Fives & DiDonato-Barnes, 2013; Notar, Zuelke, Wilson, & Yunker, 2004). Elsewhere Fives and DiDonato-Barnes stated that the TOS can be used in conjunction with lesson and unit planning to help teachers to make a clear connection between planning, instruction, activity and assessment. The primary goal of a TOS is to improve the validity of a teacher's evaluation in relation to a particular assessment (Fives & DiDonato-Barnes, 2013). Notar et al. (2004) suggested that teachers should make use of the TOS since it identifies not only the content areas covered in class, but also the performance objectives at each level of the cognitive domain of Bloom's Taxonomy.

In other words, TOS projects that teachers should bear two major elements; evidence based on the validity of test content and response process evidence when devising an assessment instrument for decision-making (Fives & DiDonato-Barnes, 2013). The former encompasses the extent to which a test assesses the ability it is designed to assess (Wolming & Wikstrom, 2010). It enables teachers to see whether or not the assessed objectives actually reflect the ability which the subject is designed to cater (Wolming & Wikstrom, 2010). The latter underscores the kind of skill a learner needs to exhibit in the process of instruction and assessment activities. It aligns the level of thinking required for a given assessment and the content delivered in the instructional process and activities (Wolming & Wikstrom, 2010).

It is very important for a teacher to prepare his/her own TOS, due to the fact that the TOS acts as the conjunction with lesson and unit planning to help the teachers to make a clear connection between planning, instruction and assessment (Fives & DiDonato-Barnes, 2013).

Taken altogether, the literature on teacher-made assessment and the use of TOS in making content relevant and inclusive, the following research questions were put forward:

1. Do the teachers in the selected Chinese Elementary Schools know about the TOS?
2. Does the teacher-made assessment in the selected Chinese Elementary Schools fulfil the content validity based on the TOS?

2. Methodology

2.1 Sample

The respondents of this study were 30 teachers conveniently selected from three elementary Chinese schools in Johor Bahru, Malaysia. The teachers were conveniently selected, across class groups. There were no limitations on the gender, age, and work experiences in selecting the respondents.

2.2 Instrument

The study used a survey questionnaire in the process of data collection on the variables under study. The questionnaire comprises two major sections. Section one, formed three items, which requested respondents to provide general demographic information such as gender, age, and work experience. Section two, comprises ten items on the TOS (Fives & DiDonato-Barnes, 2013; Notar et al., 2004). Items 1 through 4 were intended to examine the understanding of the teachers on TOS while items 5 through 10 intended to test the content validity of the TMA. A four-point Likert scale was employed for data collection. The scale is interpreted as: 1=strongly disagree, 2=disagree, 3=agree, and 4=strongly agree.

Two stages of mean calculations were used to interpret the results. The mean score of each item was first used to define the responses of the respondents (Ghafar, 2012). The middle point of the scales in the questionnaire was then used to calculate the cumulative mean of the responses (Ghafar, 2012). According to Glass and Hopkins (1996), mean scores range from 1.50-2.50 are labelled as low (in terms of validity in the case of this study). Mean scores range from 2.51-3.50 are identified as average, and mean scores range from 3.51-5.00 or greater are classified as high.

2.3 Content Validity

Prior to data collection, the instrument was content validated. The content validity of the survey instrument in this study was obtained by distributing the ten-item survey instrument to the subject matter experts (SME); two

senior assistant principals and a principal selected from different Chinese elementary schools in Johor, Malaysia to seek their judgments on the instrument. The selected experts were requested to comment if the items are clear, comprehensively captured the core elements of the TOS, and could be easily understood by the teachers. However, if the contrary is observed, they are requested to underline the ambiguous items or words for further modification or exclusion.

The feedback received from the three SMEs seemed to come to the conclusion that the survey instrument is clear on the scale of the TOS, and relevant in answering the research questions for this study. However, one of the senior assistants raised a concern that one of the items which sounded “*I am clear with the time allocated for each topic of the subject which I involved in*” should be related to the TMA and not the content of a subject itself. After discussion, the statement for this item was revised and modified into “*I am clear with the time allocated for answering the assessment of the subject which I involved in*”.

3. Results and Discussion

The data gathered from the completed questionnaires by 30 respondents who were teachers from the three selected Chinese Elementary Schools was analysed using the predictive analytic software (PASW) version 22.0. Cronbach's alpha coefficient of reliability which acts as an effective tool to ensure reliability of the instrument was conducted (Santos, 1999). The number of test items, item inter-relatedness and dimensionality affect the value of alpha where the acceptable values ranging from 0.70 to 0.95 (Tavakol & Dennick, 2011).

The result showed that the questionnaire used as the instrument for this study was reliable since the Cronbach's Alpha is 0.791 (Santos, 1999). Hence, the result indicated that the items used in the questionnaire for this study were adequate in number and interrelated with each other.

3.1 Respondents Characteristics

The results of the frequencies indicate that 11 or 36.7% of the respondents were male teachers, while 19 or 63.3% of respondents were female teachers. The distribution of the respondents according to age revealed that the majority of the respondents were aged between 21-30 years old with 14 or 46.7%. Followed by the age group of 31-40 years old with 8 or 26.7%. Furthermore, respondents aged between 41-50 years old were 6 or 20.0%. In addition, teachers who were above 51 years of age were considered as the least number in terms of teachers' participation with only two respondents or 6.7%. The results also indicated that the majority of respondents with 17 or 56.7% have being teaching for the period of 1-5 years. This was followed by the respondents who had more than 16 years of teaching experience with 5 or 16.7%. While respondents of 6-10 years and 11-15 of teaching experience were the least with 4 or 13.3% respectively according to the participation scale. Table 1 presents the details.

Table 1. Frequency and percentages of respondents' gender, age, and work experience

Variable	Frequency	%
Gender		
Male	11	36.7
Female	19	63.3
Total	30	100
Age		
21-30	14	46.7
31-40	8	26.7
41-50	6	20.0
51 and above	2	6.7
Total	30	100
Work experience		
1-5	17	56.7
6-10	4	13.3
11-15	4	13.3
16 and above	5	16.7
Total	30	100

3.2 Teachers' Understanding of the Table of Specification

The results of the individual item means showed that Item 2 “I can build a TOS for the subject that I involved in” and Item 4 “I attended courses related to the building of instruments for an assessment” had a low mean score

which was lower than the middle point. This indicated that the majority of respondents never attended courses related to the TOS. Consequently, they were unable to build a TOS due to the lack of knowledge on it. This finding corresponded with Cooper, Pittman, and Womack (n. d.) who stated that many teachers are unclear of how to use assessments due to inability to use the TOS.

Interestingly, the overall results also showed that the teachers' understanding on TOS was slightly lower with the cumulative mean of 2.485 which was below the middle point, 2.50. Table 2 demonstrates the details.

Table 2. Mean, cumulative mean and standard deviation of the items examining the teachers' understanding on TOS

	N	Minimum	Maximum	Mean	SD
1. I know about the Table of Specification (TOS).	30	3	3	3.00	.000
2. I can build a TOS for the subject that I involved in.	30	2	3	2.17	.379
3. I discuss with my colleagues in building the instruments for assessment.	30	2	4	3.20	.484
4. I attended courses related to the building of the instruments for an assessment.	30	1	2	1.57	.504
Cumulative mean: 2.485					

3.3 The Validity of Teacher-Made Assessment

The results demonstrated that almost all items related to the validity of TMA had exceeded the middle point of the mean score. However, Item 5, "I refer to the TOS while building the instruments for assessment" had a low mean score which was lower than the middle point. This indicated that the majority of the respondents did not refer to the TOS while building assessment instruments. This finding did not align with Notar et al. (2004) who suggested that teachers should make use of the TOS since it identifies not only the content areas covered in class, but also the performance objectives at each level of the cognitive domain of Bloom's Taxonomy. Teachers who do not use the standard construction guidelines for a test development are likely to fall short of assessing student achievement well and are likely to have poor content validity (Notar et al., 2004).

The results also showed that the overall mean score of the TMA was higher than the average since it exceeded the middle point of 2.50 with a cumulative mean of 3.202. The validity of the TMA obtained while teachers are not referring to the TOS in developing assessment instruments raised a concern that the assessments made by the teachers might lack evidence of test content alignment. This is so, because there is no specification for the test alignment with the specific content, and that it is hardly claimed that the assessed objectives actually reflect what should be assessed in the absence of the TOS. Based on the results, the TMA in the selected Chinese elementary schools exhibited only response process evidence and lacks evidence based on test content. That is, while the TMA addresses the content taught in the classroom given a certain level of meaning (i.e students' responses), the content is not aligned with instruction or guideline driven from the TOS guidelines. Table 3 presents the details.

Table 3. Mean, cumulative mean and standard deviation of the items examining the content validity of the teacher-made assessment

Items	N	Minimum	Maximum	Mean	SD
5. I refer to the TOS while building the instruments for an assessment.	30	1	3	2.10	.403
6. I am clear with the contents of the subject which I involved in.	30	3	4	3.43	.504
7. I am clear with the format of assessment for the subject which I involved in.	30	3	4	3.47	.507
8. I am clear with the time allocated for answering the assessment of the subject which I involved in.	30	3	4	3.47	.507
9. I am able to categorise the items of assessment based on their level of difficulties.	30	3	4	3.17	.379
10. I am clear with the level of understanding of my students' for the subject which I involved in.	30	3	4	3.57	.504
Cumulative mean: 3.202					

4. Conclusions

The findings in this study concluded that the teachers in the selected Chinese Elementary Schools in the state of Johor Bahru, Malaysia had a low understanding on TOS. This was no surprise since the majority of the respondents never attended courses related to TOS and consequently were unable to build a TOS for the subjects which they were involved in.

The study, however, reached the conclusion that the TMA was observed to be average in terms of the content validity. It is worth noting that the average validity obtained was not based on teachers' alignments of their assessment with TOS, due to the fact that the very item that referred TMA to TOS exhibited a lower mean score. This is a clear indication that most of the teachers did not refer to the TOS while building the instruments for an assessment. The average validity obtained while teachers do not refer to the TOS, is a clear indication that they refer to their students' abilities and responses in constructing assessment instruments rather than referring to the TOS. The main reasons for these unsatisfactory mean scores were due to the lack of knowledge in designing a standard TOS as well as the low awareness among teachers about the importance of the TOS in the sampled Chinese elementary schools. Several recommendations such as creating and promoting awareness of the importance of the TOS, training on designing a TOS and the usage of multiple assessment methods were driven from the findings to help the teachers having a better understanding of the TOS.

5. Recommendations

Only an average validity, which does not refer to the TOS was observed among for the assessments made by the teachers from three elementary schools in Pontian. Hence, the study puts forward some recommendations for improving the validity of the TMA.

5.1 Enhancing Teacher Awareness about the Importance of Table of Specification

From the result, it is clear that most of the teachers did not refer to the TOS while building assessment instruments. This was due to low awareness amongst teachers on the importance of the TOS. Thus, it is important for the administrators or school principals to build a high awareness among the teachers on the importance of TOS and conduct trainings on using TOS in order to improve teacher teaching strategies. This can be done with certain programmes such as a talk by an experienced person in designing TOS, in-school training and peer coaching especially between the experienced and novice teachers.

5.2 Training on Designing a Table of Specification

The findings clearly showed that the majority of teachers never attended courses related to TOS and thus they were unable to build a TOS. Lack of adequate training in developing and designing TOS leads to lower TMA, and affects its validity (Cooper, Pittman & Womack, n.d.; Micheal & Burlingame, 1996). Training or workshop will be the best ways for achieving this objective. Well-developed survey, and in the case of student assessment, instruments will provide quality data which are pre-requisite evidences to deciding and making judgments about to student achievement of particular goals of learning.

5.3 Multiple Assessment Methods

Teachers and administrators should understand that there is always an error in all classrooms and standardised assessments (McMillan, 2000). According to McMillan, multiple methods of assessment should be employed to ensure that the assessments are fair, leading to a valid inference with a minimum of error. Hence, the teachers need to use various approaches for the assessments that they gather quality data on pupils' performances.

However, it is important for a teacher to best modify a TOS to meet the needs in order to improve the validity of a teacher's evaluation based on a given assessment (Fives & DiDonato-Barnes, 2013). Although considerable time and effort may need to be spent when developing the TOS, it would worth effort, not just because it will ease the preparation tests accordingly, but for its significance to getting the most relevant and telling data on students' performance, once the plan is developed (Notar et al., 2004).

6. Limitations and Future Research

This study used a cross-sectional survey method to sample elementary school teachers participated in the study. However, a mixed-method using strength of both qualitative and quantitative means would address more adequately the depth and breadth of the TMA across the sampled schools. The sample size and the use of the convenience sampling technique were other limitations associated with this study. The study sampled only 30 teachers from three Chinese elementary schools in Johor, and thus the findings should be should not be generalised to all Chinese schools nationwide. Therefore, future studies should sample more schools from different states; and in case of the convenience sample, there is a likelihood that the sample might not be

representative of the actual population. Therefore, future studies should render considerable attention to the characteristics of the respondents. More specifically, a random sampling technique is recommended in future studies. In addition, the study exclusively sampled only Chinese elementary school teachers. Thus, there was no inclusion for Indian and Malay elementary school teachers, which might be considered if the findings were to be generalised across the elementary schools and to capture comprehensive information on the TMA across Malaysia.

References

- Alade, O. M., & Omoruyi, I. V. (2014). Table of Specification and its Relevance in Educational Development Assessment. *Journal of Educational and Developmental Psychology*, 2(1), 1-17. Retrieved June 2, 2014, from <http://www.eajournals.org/wp-content/uploads/Table-of-Specification-and-its-Relevance-in-education-al-development-Assessment.pdf>
- Al-Hudawi, S. H. V., Fong, R. L. S., Musah, M. B., & Tahir, L. M. (2014). The Actualization of the Malaysian National Philosophy of Education in Secondary Schools: Student and Teacher Perspectives. *International Education Studies*, 7(4), 57-68. <http://dx.doi.org/10.5539/ies.v7n4p57>.
- Bachman, L. F. (1990). *Fundamental Considerations in Language Testing*. New York: Oxford University Press.
- Brualdi, A. (1998). Implementing Performance Assessment in the Classroom. *Journal of Practical Assessment, Research & Evaluation*, 6(2).
- Cooper, T., Ashley, P., & Simona, W. (n. d.). *Using Reliability, Validity, and Item Analysis to Evaluate a Teacher-Developed Test in International Business*. Retrieved June 21, 2014, from http://stu.westga.edu/~ahinson1/abed6106/evaluationandtesting_research_article.pdf
- Fives, H., & DiDonato-Barnes, N. (2013). Classroom Test Construction: The Power of a Table of Specifications. *Practical Assessment, Research & Evaluation*, 18(3), 1-7. Retrieved April 27, 2014, from <http://pareonline.net/pdf/v18n3.pdf>
- Fook, C. Y., & Sidhu, G. K. (2006). *School-Based Assessment among ESL Teachers in Malaysian Secondary Schools*, Shah Alam, Selangor. Faculty of Education, Universiti Teknologi MARA.
- Frisbie, D. A. (1988). Reliability of Scores from Teacher-Made Tests, *Article of NCME Instructional Module, University of Iowa*.
- Ghafar, M. N. A. (2012). *Reka Bentuk Tnajuan Soal Selidik Pendidikan*. Johor, Malaysia: UTM Press.
- Glass, G., & Hopkins, K. (1996). *Statistical Methods in Education and Psychology*. New Jersey: Prentice -Hall, Inc. Inglewood Cliffs.
- Grimm, L. G., & Yarnold, P. R. (2006). *Reading and Understanding Multivariate Statistic*. Washington, DC. American Psychological Association.
- Griswold, P. A. (1990). Assessing Relevance and Reliability to improve the Quality of Teacher-Made Tests. *Journal of NASSP Bulletin*, 74, 18-24. <http://dx.doi.org/10.1177/019263659007452305>.
- Linn, R. L. (1980). Issues of Validity for Criterion Referenced Measures. *Article of University of Illinois, Urbana-Champaign*.
- Majid, F. (2011). School-Based Assessment in Malaysian Schools: The Concerns of the English. *Journal of US-China Education Review*, 8(10), 1-15. Retrieved September 20, 2014, from <http://education.uitm.edu.my/v1/images/stories/publication/faizah/article7.pdf>
- McMillan, J. H. (2000). Fundamental Assessment Principles for Teachers and School Sdministrators. *Journal of PAREonline*, 7(8). Retrieved June 2, 2014, from <http://pareonline.net/getvn.asp?v=7&n=8>
- Mehrens, W. A., & Lehmann, I. J. (1987). Using Teacher-Made Measurement Devices. *Journal of NASSP Bulletin*. <http://dx.doi.org/10.1177/019263658707149605>
- Micheal, J. M., & Burlingame, G. M. (1996). The Reliability and Validity of the Outcome Questionnaire. *Clinical Psychology and Psychotherapy*, 3(4), 249-258. [http://dx.doi.org/10.1002/\(SICI\)1099-](http://dx.doi.org/10.1002/(SICI)1099-)
- Notar, C. E., Zuelke, D. C., Wilson, J. D., & Yunker, B. D. (2004). Table of Specifications: Insuring Accountability in Teacher-Made Tests. *Journal of Psychology*, 31(2). Retrieved June 2, 2014, from <http://www.freepatentsonline.com/article/Journal-Instructional-Psychology/119611686.html>
- Pratt, D. (1980). *Curriculum: Design and development*. New York: Harcourt Brace Jovanovich.

- Preliminary Report, Malaysia Education Blueprint. (2013-2025). Retrieved April 20, 2014, from <http://www.moe.gov.my/userfiles/file/PPP/Preliminary-Blueprint-Eng.pdf>
- Santos, R. A. (1999). Cronbach's Alpha: A Tool for Assessing the Reliability of Scales. *Journal of Extension*, 37(2). Retrieved June 2, 2014, from <http://www.joe.org/joe/1999april/tt3.php>
- Tavakol, M., & Dennick, R. (2011). Making Sense of Cronbach's Alpha. *Journal of Medical Education*, 53-55. <http://dx.doi.org/10.5116/ijme.4dfb.8dfd>
- Tracie, C., Ashley, P., & Simona, W. (n. d.). *Using Reliability, Validity, and Item Analysis to evaluate a Teacher-Developed Test in International Business*. Retrieved July 7, 2014, from <http://stu.westga.edu/~ahinson1/abed6106/evaluationandtestingresearcharticle.pdf>
- Wolming, S., & Wikstrom, C. (2010). The Concept of Validity in Theory and Practice. *Assessment in Education: Principles, Policy & Practice*, 17(2), 117-132. <http://dx.doi.org/10.1080/09695941003693856>
- Yen-Fen, L. (2004). Issues of Validity and Reliability in Second Language Performance Assessment. *Journal of Teachers College, Columbia University*, 4(2), 1-4.

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).