PRELIMINARY PROCESS OF CLASSROOM TEST ITEMS CONSTRUCTION AND ANALYSIS

Siti Fairuz Binti Dalim, Nabilah Abdullah, Noor Shah Mohd Salleh, Mohammad Mubarrak Yusof & Siti Nur Fasihah Harun
Faculty of Education
Universiti Teknologi MARA, Shah Alam, Selangor
sitifairuz3325@salam.uitm.edu.my, mubarrak@salam.uitm.edu.my

ABSTRACT

This paper discusses the preliminary process of classroom test items development using six cognitive levels of the Bloom’s Taxonomy. The test, consisting of 23 multiple-choice items on the topic of Quadratic Expressions and Equations was distributed to 39 form 4 students from Sekolah Menengah Kebangsaan Mohd Khalid in Johor. The result of the analysis shows only 22% of the items can be kept and used in the future. The remaining 78% needs to be eliminated or revised. In conclusion, the test items analysis is a good method in evaluating the quality of the test items constructed.

Field of Research: Items Construction, Items Analysis, Bloom’s Taxonomy, Classroom Test.

1. Introduction

In everyday life, an assessment normally involves process determining the objectives, gathering and processing of information, and finally making conclusions and decisions (Azizi & Mohd Isha, 2008). When the process is systematically performed, the final decision made will be more accurate and the actions taken will be more meaningful. In educational field, Gronlund (1993) defined assessment as a systematic process to determine how far students have successfully achieved the learning objectives. Considering the many advantages that assessment can offer, it is important that teachers put in more efforts towards conducting effective assessments for their students. According to Postlethwaite (2005), one valid approach to carrying up assessment for students’ learning is by giving tests. This is supported by previous researcher such as Ismail Ali (1985) who claimed that test is the best way to assess the students’ achievement in learning. More importantly, the Malaysian Examination Syndicate (Lembaga Peperiksaan Malaysia, LPM) (2001), has also stated that test and examination are the activities to assess the effectiveness of teaching and learning.

1.1 Test Items Constructions

As claimed by Mohd Affandi (1998), constructing test items is not an easy task. Only those who are expert in the field and those who are aware of the current issues in the subject are able to construct good test items. However, there are several guidelines and procedures teachers can benefit from when constructing test items. The five general steps in test item construction agreed and used by most researchers involves the planning, writing, reviewing, pre-testing and validating processes (Farhady, 1986).

a) Test Planning

Test planning is very important and useful for the teachers to get accurate information about the test and to ensure that the test is reliable and valid. Valid and reliable test items that are well planned would eventually allow good interpretation about the student performance. During this stage, the teacher must first determine the objectives of the test. All the elements for those objectives should be listed as well as the number of items devoted for each element. After that, the teacher must decide on
b) Writing Test Items

Writing test items requires the knowledge of an expert and must be done very carefully (Farhady, 1986). There are several guidelines that can be applied during the test item writing process. These guidelines, as proposed by Farhady (1986) and Azizi and Mohd Isha (2008) are the general principles in writing the test items.

- Instructions should be simple and easy to understand.
- Use a simple and clear writing language that can help the students to understand the questions and provide good answers.
- Construct the test items which really test the students’ knowledge about the topics and not the students’ general knowledge.
- Avoid grammatical error for all the choices and the alternatives.
- Use a similar length, same contents and contexts for all the alternatives of the same item.

c) Reviewing Test Items

According to Farhady (1986), reviewing the test items that have been developed is very important to minimize the pitfalls occurred during the earlier stages of the test items constructions. This process will be conducted by the experts who will critically examine the correspondence between test contents and the table of specifications. The level of difficulty and the appropriateness of the items will also be considered at this stage after that, some comments to improve or modify the test items will be given. The first draft of the test will be produced after modifications and improvements have been performed.

d) Pre-testing

Pre-testing is the next stage in the steps of test items development. This process involves administering the test to a group of examinees who are similar in knowledge to the target group (Farhady, 1986). The main purpose of conducting the pre-testing is to determine the characteristics of individual items. This is achieved by conducting item analysis. Through this item analysis, the difficulty index and discrimination index for each item can be obtained. Based on the results, items that do not fulfill the accepted standards for the test items will be either modified or discarded.

e) Validation

Validation is the final step of the test items development. It is performed after some modifications are carried out on specific items based on analysis results. According to Farhady (1986), the modifications can include changing a distractor, stem, complete item or discarding the whole item. The pre-final draft of the test will be produced when all these steps are done. The validation stage is the last process that a test will go through when all the test items have met the necessary standard.

This report will be focused on the application of several steps mentioned above in the construction of a classroom test items. Detailed explanation will be given for each of the steps involved. This will include the specification of the learning objectives, the construction of the test items according to Bloom’s
1.2 Rational of the Research

This research is conducted considering the many advantages that it has not only for the teachers, but also for the students and schools. Firstly, this research informs the importance of assessments in educational field. This includes its importance to improving the students’ achievements as well as improving the teachers’ competencies in teaching skills.

Secondly, this research provides excellent exposure for teachers in terms of constructing the test items and analysing the items based on the carry out procedures. By having sound knowledge of test items construction, the teachers can conduct the assessments as part of their teaching strategies. As mentioned by Raminah (1991), a good teacher should know not only how to teach in class but also to know how to construct good test items. She also added that well-planned and systematic testing and evaluation procedures is extremely crucial in ensuring that the teaching and learning activities will yield the expected positive results.

Besides that, this research will also explore the step-by-step procedures involved in constructing test items. This includes the use of Table of Assessment Specification (Jadual Penentuan Ujian, JPU) as the first step in identifying the test objectives, the contents and the number of items for each of the objectives. At this stage, teachers will be exposed to the concept of developing good test items according to Bloom’s Taxonomy. By referring to the Bloom’s Taxonomy, the items constructed should cover all the six cognitive levels which are knowledge, comprehension, application, analysis, synthesis, and evaluation. By having this knowledge, teachers can enhance their skills in terms of providing good quality questions that can assess all the students’ cognitive skills (Kaur, 2008). The Bloom’s Taxonomy will also serve as a reference for teachers so that they do not only focus on students’ knowledge to memorize things but more importantly they can assess the students’ higher cognitive skills which are analysis, synthesis and evaluation (Mokthar, 1995).

Another important factor for conducting this research is to expose the teachers to the practical aspects of item analysis. According to Gronlund (1993), item analysis provides valuable information regarding the difficulty and discrimination index of the questions. Therefore, this analysis can serve as an indicator to show whether the items constructed are clear, easy or difficult and suitable for the groups to be tested. Teachers can also gain information and academic skills on developing better test items in the future.

a) Rationale of the Chosen Topic

The rational for choosing Quadratic Expressions and Equations for the purpose of test items constructions and analysis is because this chapter is one of the important topics in form four Mathematics provided by the Curriculum Development Centre (CDC). Based on the analysis of past years’ examination papers, many questions are selected from this chapter for the purpose of public examination (Soon & Quan, 2011). Therefore, it is important to include this topic as part of the test items construction and analysis. The advantage of developing the test items for this chapter is, it helps the students to improve their understanding and mastery for this topic. Since all questions from this chapter are developed in such a way that it follows all the cognitive skills of Bloom’s Taxonomy, therefore, these questions are able to assess the students’ cognitive skills. More importantly, the students’ higher cognitive skills which are analysis, synthesis and evaluation are also given greater emphasis. As supported by ETS (2003), the properly developed and interpreted classroom assessment can help the teachers to better understand about what their students are learning. Besides that, the teachers can also use the pre-test result of this topic to diagnose their students’ strengths and weaknesses regarding this chapter. This means that the teachers are able to identify in which part of
the topic that students are weak in and need further revision. They can also use the result to monitor the students’ progress and after that can further decide on the appropriate teaching and learning strategies to improve their students’ achievements (Sang, 2008).

b) Preparing the Table of Specification (Test Blue Print)

In Malaysia, the test blueprint used is known as Jadual Penentuan Ujian (JPU) (Kaur, 2008). It serves as a guide to help teachers prepare questions based on specified criteria needed to ensure the validity and reliability of the test. JPU is normally constructed after the topics and cognitive levels are identified (Sang, 2008). JPU is critical in ensuring equal weightage for all the topics and cognitive levels of students. There are several characteristics of Jadual Penentuan Ujian (JPU) as explained below (Kaur, 2008):

a) Topics to be tested
b) Cognitive level for questions of the test
c) Percentage weightage for questions
d) Total number of questions to be tested

The Table of Specification for the topic of Quadratic Expressions and Equations have been prepared by taking into consideration the six cognitive levels of Bloom’s Taxonomy and also all the criteria that the table should have. Therefore, the Table of Specification (JPU) constructed will provide the information about the topics to be tested, the learning objectives, the cognitive level as well as the number of questions for every topic. Table 1 shows the table of specification.

Table 1: Table of Specification

<table>
<thead>
<tr>
<th>Content</th>
<th>Level of Cognition</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Knowledge</td>
</tr>
<tr>
<td>A) Understand the concept of quadratic expressions.</td>
<td>* Identify quadratic expressions . [1, 2]</td>
</tr>
<tr>
<td></td>
<td>21.7%</td>
</tr>
<tr>
<td>B) Factorise quadratic expressions.</td>
<td>* Factorise quadratic expressions of the form ax^2+ bx+ c where b=0 or c=0. [6]</td>
</tr>
<tr>
<td></td>
<td>30.43%</td>
</tr>
<tr>
<td>C) Understand the concept of quadratic equations.</td>
<td>*Identify quadratic equations with one unknown</td>
</tr>
</tbody>
</table>
3. Methodology

3.1 Samples

Upon the completion of planning and writing of test item, the next stage is to select a sample of students who have learned about the Quadratic Expression and Equations topic for the purpose of conducting the pre-testing. The school selected is a secondary school in Johor which is Sekolah Menengah Kebangsaan Mohd Khalid. About 39 Form 4 students from 2 classes have been selected to sit for the test. The constructed test consists of 23 multiple-choice items. The questions included in the test have been designed in such a way that they cover all the four subtopics of the chapter which are quadratic expressions, factorization of quadratic expressions, quadratic equations and roots of quadratic equations.

3.2 Item Analysis

Item analysis is one of the methods that can be used in order to assess and evaluate the quality of the test items by collecting, summarizing as well as using the information gained from the students’ responses (Mitra et al., 2009). In this method, the students’ performance (score) will be used to calculate the level of difficulty (p) and the index of discrimination (D) of each of the test items (Sim & Rasiah, 2006). Item analysis can help to evaluate the standard of multiple choice examination questions as discussed by Mitra et al. (2009). Besides that, it can also be used to determine the discrimination of distractor index. All indexes are put under index card. The entire question has its own index card. Table 2 shows an example of question 1 index card for question 1.
Table 2: Index Card for Question 1

<table>
<thead>
<tr>
<th>Date:</th>
<th>31 Mac 2011</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item:</td>
<td>1. All of the following is a quadratic expression, EXCEPT</td>
</tr>
<tr>
<td></td>
<td>A. $\frac{3}{4}(3x^2 - 4x + 1/5)$</td>
</tr>
<tr>
<td></td>
<td>B. $2y^2 + 2$</td>
</tr>
<tr>
<td></td>
<td>C. $y(6x^2 + 3x + 1)$</td>
</tr>
<tr>
<td></td>
<td>D. $2m^2 + 4m + 7$</td>
</tr>
<tr>
<td>$p$ =</td>
<td>0.70</td>
</tr>
<tr>
<td>$D$ =</td>
<td>0.40</td>
</tr>
<tr>
<td>Answers</td>
<td>Upper Group</td>
</tr>
<tr>
<td>A</td>
<td>0</td>
</tr>
<tr>
<td>B</td>
<td>1</td>
</tr>
<tr>
<td>C</td>
<td>9</td>
</tr>
<tr>
<td>D</td>
<td>0</td>
</tr>
<tr>
<td>No Answer</td>
<td>0</td>
</tr>
<tr>
<td>Correct Answer</td>
<td></td>
</tr>
<tr>
<td>Comments:</td>
<td>This item can be kept and used in the future as it has an acceptable level of difficulty ($p=0.70$) and very good discrimination index ($D=0.40$). The best distractor is B with ($Dd=0.20$) because many students from the lower group who have the wrong answers chose B as their answers. The distractor A and D need to be modified and improved since they have low discrimination indexes which mean that they are not good in discriminating between the high and low achievers. Since this question is to test the students’ ability to really understand about the characteristics of quadratic expression, therefore the recommended options for A and D are:</td>
</tr>
<tr>
<td></td>
<td>A) $2x(3x + 2)$</td>
</tr>
<tr>
<td></td>
<td>D) $(2y + 1)(y - 5)$</td>
</tr>
</tbody>
</table>

4. Finding & Discussion

Figure 1 below shows a graph of the distribution of the difficulty index ($p$) and the discrimination index ($D$) for all 23 items constructed. Based on the analysis performed, the mean of the difficulty index, $p$ is 0.81 and the discrimination index, $D$ is 0.29. In general, the result of the analysis shows that the items have a high level of difficulty index which is more than 0.7. This means that the items constructed item are too easy for students (Sim & Rasiah, 2006). The discrimination index also shows a poor value which means that the items developed are not suitable in discriminating between good and weak students.

Based on the analysis performed on each of the item, only 22% of the items have difficulty index ($p$) between 0.60 to 0.70 which are considered as moderately difficult and are acceptable. The items, numbered 1, 7, 9, 12 and 22 can be kept and used in the future. According to Yap et al (1986), the items with difficulty index in the range of 30% to 70% are acceptable and considered as moderately difficult items. This is because, the items having the index of difficulty in between this range shows that most of the students have better chances in answering the questions. These items are also considered as good items since the discrimination index is more than 0.4. Only one item from this group has a discrimination index of 0.2 which is still acceptable but needs to be improved. Items with high
discrimination index means that they are effective in discriminating or differentiating between good and weak students. This is supported by Sim and Rasiah (2006) who claimed that the higher the index of discrimination, the better they are in discriminating the students who have high scores with the students who have lower scores. Therefore, these types of items should be kept by the teachers to be used in future.

![Figure 1: The Relationship between Difficulty Index (p) and Discrimination Index (D)](image)

*Figure 1: The Relationship between Difficulty Index (p) and Discrimination Index (D)*

About 78% of the multiple-choice items constructed should be eliminated or extensively revised since the difficulty index (p) is more than 0.70 which are considered as very easy items. From the 78% of these very easy items, about 44.4% of the items has very good discrimination index which is between 0.40 and 0.50. These items which include items 2, 6, 8, 14, 18, 20, 21 and 23 are very good in discriminating between the students from the upper and lower groups. Therefore, it is highly recommended that these items are revised and improved in terms of lowering the difficulty index. As for items 3 and 11, the discrimination index falls within the acceptable range in discriminating the students which is between 0.20 and 0.30. Therefore, it is also recommended that these items are revised and improved. Items 4, 5, 10, 13, 15, 16, 17 and 19 should be eliminated since the index difficulty is very high, exceeding the acceptable range of 0.70. Besides that, these items also have poor and negative discrimination index. Meaning that these items are very poor and are unable to discriminate the students.

In terms of the discrimination index of distractors, the item analysis performed shows that most of the distractors have poor discrimination index with the value ranges between -0.1 and 0.4. Therefore these distractors should be revised and improved. However, there are some items which have good discrimination of distractors such as items 2, 8, 14 and 18.

According Sim and Rasiah (2006), it is important to evaluate the multiple-choice items to know how effective they are in assessing the students’ knowledge about a topic. One of the important components in an item analysis is the discrimination index. The poor discrimination index will serve as indicator towards the weaknesses of the items which can be contributed by ambiguous wording or poorly constructed items. Therefore, the items with the poor discrimination index should be reviewed by the respective individuals. In terms of the level of difficulty index, the lower the p value means that
the item is more difficult. There are some possible reasons that can cause this to happen. For example, the topic tested is not appropriate at the stage of the students' learning or the topic is taught well by the teachers. To improve the standard of MCQ items, evaluation by experts are encouraged (Sim and Rasiah, 2006).

The results of the item analysis performed for the test items developed for this research project shows that most of the questions are high in the index of difficulty which means that the items are too easy to be answered by students. There are several possible reasons contributing to this situation such as the items are too easy or poorly constructed which enable the students to easily guess the answers (Sang, 2008).

5. Conclusion

Designing a test item construction is not an easy task. However, there are several guidelines that can be used in helping the construction of the test items. As explained by Farhady (1986), there are five general steps involved in constructing the test items. It begins with the planning stage, followed by item writing, reviewing, pre-testing and finally validation. At the planning stage, several important aspects must be considered such as identifying the learning objectives, specifying the test contents, developing the table of specification and determining the types of the test items to be used. This stage has also involved the use of the six cognitive levels of Bloom's Taxonomy which include the knowledge, comprehension, application, analysis, synthesis and evaluation. There are many important reasons for applying the Bloom’ Taxonomy cognitive levels in the test item construction process as explained in the previous section of this report.

After the test items construction has been done, the next important stage is to perform the item analysis. In general, the purpose of conducting the test item analysis is to ensure the reliability of the test item by assessing its quality (Mitra et al., 2009). They added that the item analysis will help in identifying the good and poor items. The items considered as good are the items that have an acceptable range of difficulty index (p) between 0.3 to 0.7 and and good discrimination index (D) which is greater than 0.4. Therefore, these item should be kept to be used in the future. In constrast, the items that are considered as poor should be eliminated or revised if it is be used in the future. These items have the difficulty index (p) that less than 0.3 or exceeds 0.7. In terms of the discrimination index (D), the value is less than 0.3 which means that, this items are not effective to discriminate or differentiate between the students who have high scores with the students who have low scores. For these types of items, investigations should be carried out to identify the factors that cause the poor discriminations before they are discarded.

As for this research project, there are 23 multiple-choice items on the topic of Quadratic Expressions and Equations have been constructed. During the test item construction stage, all the guidelines and criteria mentioned above are considered. The test item analysis has also been performed for the test items. The results obtained shows that only 22% of the items are acceptable and can be kept. The remaining 78% of the items have high difficulty index which exceeds 0.7. However, the discrimination index calculated are very good. Therefore, these questions should be reviewed before they are rejected so that the the quality of this test items can be improved.

Based on the test item construction that have been performed, it can be concluded that a well-structured assessment strategy is very important to ensure that all the objectives of the assessment can be successfully achieved. As mentioned by Sim and Rasiah (2006), an appropriate assessment strategy is a key part for an effective sustainable curriculum development. Therefore, it is very important to develop the test items based on the well-defined strategy that have been mentioned.

References


