Advances in Social Science, Education and Humanities Research, volume 399
International Conference on Educational Psychology and Pedagogy (ICEPP 2019)

An Analysis of the Eight-Grade Test's Items of the Islamic Junior High School in Yogyakarta

Nurul Aisyah

Universitas Muhammadiyah Yogyakarta
nurulaisyah@umy.ac.id

Abstract

This study is a survey intended to determine the level of difficulty and power of discrimination of a final test 2018/2019 academic year with the use of multiple choice applied in the subject of Aqidah-Akhlak for the Islamic Junior High School's Eighth Grade in Yogyakarta. Method of data collection use documentation review by analyzed 40 items of the multiple-choice questions provided by Ministry of Religion in Yogyakarta, which sample is MTs X by purpose sampling technique. This research was held by analyzing different levels of difficulty and power of discrimination of any item what used AnBuso versi 8.0 further interpreted into established criteria. The main source was the document of that multiple-choice instrument. The results showed that the level of difficulty was considered no proportional with 50% medium, 30% easy and 20% hard. Meanwhile, the power of discrimination was poor 24 items, fair 7 items, good with 6 items and 4 items satisfied. Therefore, it is not good assessment instrument to measure different capability of students.

Keywords: multiple-choice question, item analysis level of difficulty, power of discrimination

1. INTRODUCTION

The Summative Exam is conducted by sharing questions to students. Before the teacher prepares the test items, the teacher really needs to make a test of the construction of each question in each of the basic competencies previously set. An effort to improve the quality of learning processes and outcomes as part of improving the quality of education can be done through a system of assessment and evaluation. In assessing the process and student learning outcomes, the teacher provides an evaluation of the level of achieving student competence.

Construction of multiple-choice questions must match the guidelines for making good questions so that the questions can really measure the level of knowledge of students based on established competency standards. Efforts to know whether the questions made by the teacher are classified as feasible and good and provide maximum results in measuring and increasing the level of understanding of students, can be analysed on each item. Analysis of the quality of goods can be done by measuring the level of difficulty and the strength of various questions (Sudjana, 1990).

1.1 Item Statistics

Item statistics are used to assess the performance of individual test items on the assumption that the overall quality of a test derives from the quality of its items. The item statistics report provides the following item information such as item difficulty and index of discrimination.

Item difficulty, usually called P-value. P-values are proportional values, referring to the percentage of students who answer questions correctly. On the other hand, trouble items are numbers that indicate the ease and difficulty of items (Arikunto, 2009). Items range from 0 to 1.00, which indicates that the question is easier, the higher the value. A good test item is if the test item is not too difficult and not too easy or in other words the difficulty level of the test item is moderate. The index range for item difficulties is 0.00 to 1.00. The smaller the difficulty index, the harder it is. The classification used to interpret the results of the calculation of the difficulty level of items, namely 0,000-0,299 included in the category of difficult questions, 0,300-0,699 included in the medium category, and easy categories between 0,700-1,000 (Arikunto, 2009).

The test consists of objective tests and subjective tests, calculation of the difficulty of the problem in two different types of tests. Objective tests can find out the difficulty level of items in the following ways:

P = B / JS
P = item difficulty index number
B = number of students who answer the questions correctly
JS = total number of students in the test

Item discrimination refers to the ability to distinguish students about how well they be able to answer questions on an exam. It also shows how items distinguish between high-ability students from those with low abilities. These values range from negative scores to 1.00. Aggarwal's theory (1986)

that test items that have a discrimination index of negative value are reject. Aggarwal follows general guidelines which are said to have a satisfying discrimination index if above 0.30 is "good". If between 0.10 to 0.30 is "fair" and below 0.10 is “poor”. In addition, Naga (1992) states that if the discrimination index score is less than or equal to 0.19 it must be rejected (not used) and if between 0.20 to 0.29 must be fully revised. Scores of 0.30 to 0.39 are slightly revised and equal to 0.4 or more is satisfied category.

Analysis of items is a method of gauging the quality of test items on the examination (McAlpine, 2002). This is a process of measuring the quality of test items based on student responses in individual tests. In addition, item analysis can also be used to identify the ambiguity of test items to the components of the academic component that must be measured according to the syllabus.

1.2 Islamic Junior High School

In Yogyakarta, most of the Islamic school based on two rules for the learning process as well, Ministry of religion rule especially for Madrasah Tsanawiyah (MTs) and Muhammadiyah provincial board for Muhammadiyah Junior High School. Researcher did the research in MTs X is one of the private schools which based on both rules, because this is MTs Muhammadiyah. Especially on Aqidah Akhlaq Subject, that is following the rule of Ministry of religion on Bantul regency, so on the Aqidah Akhlaq examination students use the question provided by Ministry of religion, but in other subjects the item test is provided by Muhammadiyah provincial board of Yogyakarta.

2. METHOD

This type of research is survey research. Survey research examines a large population or small population by selecting and reviewing selected samples from that population, to find the incidence, distribution, and relative interrelations of the variables (Kerlinger, 2004). Survey research can be used for evaluation purposes (Singarimbun & Effendi, 1995). Survey research according to Soehartono (2000) is classified as having two objectives, firstly to provide an overview / explanation of something, secondly to conduct analysis. Survey research has contributed greatly to the methodology of social science. One of the most important contributions of survey research is careful and rigorous sampling procedures, the overall design and implementation of design studies, definitions that are clear and specific to research problems, and analysis and interpretation of data. Survey research uses a work chart to outline the design of the study. The work chart starts with survey targets, records each step that is to be taken, and ends with the final report (Kerlinger, 2004). According to Moehadjir (2002) one type of survey research, which is carried out to obtain basic data to obtain a general picture that is useful for planning and public policy.

The research was carried out at MTs X by analyzing the academic achievement test items on the aspects of difficulty level and the power of discrimination index about Aqidah Akhlaq in the Summative Evaluation of Odd Semester 2018/2019 academic year which was then interpreted into the criteria set. Sampling is done by purposive sampling technique, which is 62 eighth grade students. The object of this research is 40 multiple choice questions (MCQs) which have been compiled by the Ministry of Religion, Bantul Regency. Each analysis in this study aims to determine the size of the difficulty level and the multiple-choice discrimination index of the eighth grade Aqidah Akhlaq subjects on MTs X.

3. RESULTS AND DISCUSSION

3.1 An Analysis of Difficulty Index

Result of the analyse of difficulty index on this summative test on Aqidah Akhlaq Subject in MTs X is not proportional. It can be seen on this table and chart:

Table 1. Distribution of Question on Difficulty Index

Categories	Number of Question	Total of Question
Hard	8,17,21,23,25,33,36, 38	8 Questions
Medium	2,3,5,6,7,9,10,11,12, 13,15,16,18,20,29,31 ,34,37,39,40	20 Questions
Easy	1,4,14,19,22,24,26,2 7,28,30,32,35	12 Questions

Source: Primary data processed

The difficulty level of items in the hard category is only 20%, meaning that students have difficulty working on the questions with the following number questions: 8, 17, 21, 23, 25, 33, 36, and 38. Meanwhile, for the easy category it reaches 30% it means that many students answer correctly the following number of questions: 1, 4, 14, 19, 22, 24, 26, 27, 28, 30, 32, and 35. 50% are medium categories with numbers 2, 3, 5 , 6, 7, 9, 10, 11, 12, 13, 15, 16, 18, 20, 29, 31, 34, 37, 39, and 40.

Witherington's Theory in Arikunto (2009) that proportions between easy and hard categories must be balanced. If not, then the question must be corrected by revising or rejected. Thus, it can be concluded that the questions that have been made by the Bantul regency ministry of religion need to be improved in order to obtain satisfactory results of student academic achievement which can be done by balancing between the proportion of hard and easy categories.

3.2 An Analysis of Discrimination Index

Result of the analysis of discrimination index on this summative test on Aqidah Akhlaq Subject in MTs X is good, fair and poor. It can be seen on this table and chart:

Table 2. Distribution of Question on Discrimination Index

Categories	Number of Questions	Total of Questions
Good (0.330-0.490)	2,7,9,17,20,24,26,27,28	9 question
Fair (0.205 – 0.294)	1,10,13,19,22,32,35	7 question
Poor (-0.169-0.184)	3,4,5,6,8,11,12,14,15,16,18,21,23,25,29,30,31,33,34,36,37,38,39,40	24 question

Source: Primary data processed

The discrimination index in the good category reaches 40%, while the poor category is 60%. The 40% consisted of 16 questions which included 2, 7, 9, 17, 20, 24, 26, 27, 28, 1, 10, 13, 19, 22, 32, and 35. 60% consisted of 24 questions 3, 4, 5, 6, 8, 11, 12, 14, 15, 16, 18, 21, 23, 25, 29, 30, 31, 33, 34, 36, 37, 38, 39, and 40. Test items that have a good discrimination index so the item works well, that is able to measure the learning achievement of each student in the class by distinguishing between the upper group (students who can work correctly) and the lower group (students who cannot work), and vice versa if the test item has the category is not good then the test item does not work as expected, it is not able to group the upper group students with the lower group students.

The above is based on Aggarwal's theory (1986) that test items that have a discrimination index are negatively rejected. Aggarwal follows general guidelines, namely the satisfying discrimination index category when above 0.30 is "good". If between 0.10 to 0.30, the discrimination index is "fair" while "poor" is below 0.10. Besides that, Naga (1992) states that if the discrimination index score is less than or equal to 0.19, it must be rejected (not used) and if between 0.20 to 0.29 must be full revision. Scores of 0.30 to 0.39 are slightly revised then equal to 0.4 or more very satisfying. Thus, it can be concluded that the questions that have been made by the Bantul regency ministry of religion some of which must be discarded, full revised, a little revision and there are 4 items that are satisfactory. 24 about should have rejected 3, 4, 5, 6, 8, 11, 12, 14, 15, 16, 18, 21, 23, 25, 29, 30, 31, 33, 34, 36, 37, 38, 39, and 40 because the 24 test items had a discriminatory index of 0.169 to 0.184. 7 full revision items 1, 10, 13, 19, 22, 32, and 35 because they have a discrimination index of 0.205 to 0.294. Slightly revision on items 2, 9, 17, 24, and 27, while the very satisfied are items 7, 20, 26, and 28.

4. CONCLUSION

The findings of this paper are important for students, teachers and test developers for the Ministry of Religion. They must be careful when selecting items or making the test item. The quality of the questions will depend on the range of index difficulties and the discrimination index that has been designed which it able to measure academic capability of students on learning process. Indeed, this will also support the improvement of teacher pedagogical competencies.The results showed that the level of difficulty was considered no proportional with 50% moderate, 30% easy and 20% hard. Meanwhile, the power of discrimination was poor 24 items, fair 7 items, good with 6 items and 4 items satisfied. Therefore, it is not good assessment instrument to measure different capability of students.

5. REFERENCES

Aggarwal, Y. P. (1986). *Statistical methods, concepts, applications, and computations*. New Delhi: Sterling Publication.

Arikunto, S. (2009). *Dasar-dasar evaluasi pendidikan.* Jakarta: Bumi Aksara

Kerlinger, F. N. (2004). *Asas-asas penelitian behavioral*. Yogyakarta: UGM Press.

McAlpine, (2002). *A summary of methods of item analysis*. UK: Computer Assisted Assessment Centre.

Moehadjir, N. (2002). *Metodologi penelitian kualitati*f. Yogyakarta: Rake Sarasin.

Naga, D.S. (1992). *Pengantar teori skor pada pengukuran Pendidikan.* Jakarta: gunadarma.

Singarimbun, M., & Effendi, S. (1995). *Metode penelitian survei*. Jakarta: PT Pustaka LP3ES.

Sudjana, N (1990). *Penilaian hasil proses belajar mengajar*. Bandung: PT Remaja Rosdakarya.