

# An Analysis of Summative Test on the 8<sup>th</sup> Grade of Muhammadiyah Junior High School of Yogyakarta

Nurul Aisyah

Departement of Islamic Education  
Universitas Muhammadiyah Yogyakarta  
Yogyakarta, Indonesia  
[nurulaisyah@umy.ac.id](mailto:nurulaisyah@umy.ac.id),  
[orcid.org/0000-0002-3199-3231](http://orcid.org/0000-0002-3199-3231)

**Abstract-** This piece of work was aimed to survey intended to determine the level of difficulty and power of discrimination of a final test with the use of multiple choice applied in the subject of Aqidah-Akhlak for the Muhammadiyah Junior High School's Eighth Grade in Yogyakarta. Sample of this survey is 200 pupils of the Muhammadiyah Junior High School's 8<sup>th</sup> Grade in Yogyakarta. It analysed 40 items of the multiple-choice questions provided by a group of teachers provided by Muhammadiyah provincial board. This research was held by analysing different levels of difficulty and discrimination index of any items further interpreted into established criteria. The main source was the document of that multiple-choice instrument. The results showed that the difficulty index was considered proportional with 60% moderate, 27% easy and 13% hard. Meanwhile, the power of discrimination was good with 25% good, 62% fair and 13% poor. Therefore, it is not good assessment instrument to measure capability of pupils. Enhancing the quality of test especially on administering and developing Multiple Choice Questions (MCQs) will make easier for teacher assessing the quality of learning and more able to is measuring the capability pupils into learning competences.

**Keywords:** *Multiple-choice; summative item analysis; difficulty index of item; discrimination index of item*

## I. INTRODUCTION

Curriculum improvement is closely related to measuring student learning outcomes. The better curriculum must be balanced with 2013 curriculum training (K13), the learning method that is Higher Order Thinking Skills (HOTS) to the preparation of questions based on HOTS as a measure of student achievement. In this era teachers to education policy makers are also required to follow the development of the world where education must be directed to the era of Industry Revolution 4.0 (Era IR 4.0), of course it is inseparable from the evaluation of education. The increasing rapidly development of education is offset by the demands of the times that have entered the era of IR 4.0, hence the need for evaluation is increasingly needed.

Educational evaluation as a measure of student progress and measurement of existing systems in schools is

usually called a summative Test. Regulation of the Minister of Education and Culture of the Republic of Indonesia number 4 2018 which is about Learning Outcomes Evaluation. In Chapter IV, it concerns exam materials, namely in point 5 which states that school examination texts are prepared by the unit of educational institutions, then by referring to Chapter III article 8 concerning Implementation of school examinations can be through paper-based exams or computer and paper-based exams. Thus, all educational units in conducting school examinations can provide their own questions through the teacher working group and through certain teams such as the Muhammadiyah educational institutions, namely through the Muhammadiyah primary and secondary education assemblies. The exam questions are provided questions about the final assessment of semester which are held twice a year, namely in the odd and even semesters.

Daryanto (2008: 28) stated that evaluation techniques are divided into two, test and non-test. Non-test techniques are techniques for evaluating student learning outcomes without testing students, but observing, interviewing, questionnaires and checking documentation. While the test technique, student learning outcomes are measured by test items. In addition, teacher' competences are also be needed on this era. Based on the Law of the Republic of Indonesia number 14 of 2005 concerning teachers and lecturers that there are four competences which should be understood by them, like professional, pedagogical, social and personality competences. Evaluation of learning outcomes is an aspect contained in pedagogic competence where a teacher must masterful for it, in pedagogical competencies of teachers are required to evaluate learning outcomes. Evaluation of learning outcomes to be held with the aim that a teacher can measure the learning objectives it has been achieved by students. The teacher will measure students' competencies towards the learning material then evaluate the learning outcomes of students. The results of these survey are very useful for teachers, students and stakeholders because they can be taken into consideration to improve their learning activities.

For a teacher, evaluating learning is a compulsory activity. The evaluation activity will make it easier for the teacher to observe students' abilities. Evaluation of learning outcomes must be carried out continuously so that it can monitor the process, progress, and results (Juliyanto et al.,

2016: 54). In addition, it can also be used as a benchmark for how to teach. For students, it will be used as a benchmark for learning, what extent the learning achievements that they have achieved. While, for stakeholder, especially for parents, it can be used as a reference in controlling student learning.

Analysis of test items is an activity carried out to assess the level of quality of the question. If there is an item that has good quality, it can be reused, if there are test items that are not good then it needs to be removed or revised so that it can be tested to students. The quality of test items is very influential for teachers in knowing the success of learning of students if the test items used are still far from good quality, so the evaluation results will inform the wrong results related to the success of learning students (Hasna, 2017:4).

Aqidah Akhlak learning is one of the Al-Islam Kemuhammadiyah (ISMUBA) subjects which emphasizes pupils to understand and maintain faith, as well as strong faith, in which also learn how to interact with humans (*habluminannas*) and how to interact with Allah (*habluminallah*). Aqidah Akhlak Learning teaches about the values that govern the relationship between humans and God, regulates relationships with fellow humans, regulates relationships with the environment and regulates itself (Yanti, 2017: 5). Aqidah Akhlak is one of the subjects in which discuss the pillars of faith that are associated with appreciation of Al-Asma 'ul-Husna, as well as the creation of exemplary and habituation in practicing commendable morals and Islamic customs in daily life.

The learning objectives of Aqidah Akhlak include: a) Growing Aqeedah through giving, fertilizing, and developing knowledge, appreciation, practice, habituation, and experience of students about Islamic aqeedah so that they become Muslims who continue to develop their faith and devotion to Allah; b) Manifesting human beings with noble character and avoiding despicable morality in daily life, both in individual and social life, as a manifestation of the teachings and values of Islamic faith.

Substantially the *Aqidah Akhlak* subjects have contributed in giving motivation to students to practice *Akhlakul Karimah* and Islamic adab in daily life as manifestations of his faith in Allah, His angels, His books, His messengers, *Qada'* and *Qadar*. *Al-akhlak al-karimah* is very important to be practiced and accustomed to early on by students in everyday life, especially in anticipating the negative impact of the era of globalization and the multidimensional crisis that has plagued the nation and country of Indonesia.

Muhammadiyah as one of the leading Islamic organizations that emphasizes improving religious life by activating Muhammadiyah business charities in the fields of education, social and health. The process and efforts of modernization of Muhammadiyah are establishing educational institutions in learning activities to integrate and interconnect science relations with general science. Then, to make it happen, Muhammadiyah established the Mu'allimat madrasa in 1924, its initial name was Kweekschool Muhammadiyah which was divided into two parts, namely, for the boys, named madrasah Mu'allimin and for the girls

named Mu'allimat, who has now expanded to become many Muhammadiyah schools starting from early childhood education, elementary school, junior and senior high school and higher education.

The pattern and dynamic process of education carried out by Muhammadiyah, namely by implementing the first two education systems, follow the existing government school system by adding spirits and religious studies. In the following stage, Muhammadiyah builds education institution marked by its own style for the quality, nature, and manner of teaching. The education system applied by Muhammadiyah is its strength base in managing and operating educational institutions, even until its development. At the present, in Muhammadiyah, two education systems are still used as a basis for the development movement and progress of Muhammadiyah education.

Based on the results of observations and preliminary interviews, the analysis of test items was only conducted once so far, and it is unclear whether the test items prepared and provided by the Muhammadiyah elementary and secondary education assemblies in the Yogyakarta have been classified as good or not in terms of difficulty and discrimination index. A good test is a test that is carried out continuously on test analysis items, to measure whether the student's competency has been measured properly or not. The purpose of this study was to analyse level of difficulty and discrimination index of any test items on summative test of Aqidah Akhlaq subject.

## II. LITERATURE REVIEW

Literature review is an explanation of the importance of the study conducted by the researcher. Based on several literature that has been studied, the following various related discussions:

The first stage is Item Analysis of Multiple-Choice Questions at the Department of Paediatrics, Arabian Gulf University, Manama, Bahrain written by Sultan Qaboos in University Medical Journal 2018. His research aims to analyse items post-validation of MCQs in medical terms, and to evaluate the relationship between item difficulties, item discrimination and distractor functions. Whether or not they produce a good category of questions. If it is not, then it needs to be modified, but if it is good then it can be reused without modification. The method used is cross-sectional study conducted in the Department. Subsequently, Analysis of one-best MCQs consist of the difficulty index, discrimination index and distractor efficiency.

The study aims to investigate the connection of items that have a good index of difficulty and discrimination with the efficient use of the distractor function to find out how 'ideal questions'. The method used is cross-sectional study conducted at Fatima Jinnah Dental College, Karachi.

The aim is to assess the quality of MCQs which comprises 40 items of test. Lastly, the study aims to analyse the items of test on difficulty and discrimination level in education test for research. This piece of work was aimed to analyse items of test of a researcher who made test in the

subject of Research in Education for the pupil-teachers of Master of Education (M. Ed). It involves the difficulty index and discrimination index. A test of MCQs was used as a data collection instrument in different Colleges of Education to 200 pupil-teachers taken randomly.

From the three things above, what distinguishes this research from the previous one is that this study only focuses on the difficulty index and discrimination index. Moreover, this analysis item can be used in various items of examination such as health, education and other scopes. On the other hand, this piece of work is very useful for all, like teacher, team who conducted the test, pupils and parents of pupils.

### III. METODOLOGY

Type of this study is survey. Survey research examines a small or large population by selecting and reviewing the samples to find the incidence, distribution and relative interrelation of variables (Kerlinger, 2014: 660). Survey research used to solve observed problems, to assess needs and set goals, and most importantly to determine the specific goals or objectives that have been achieved, to establish policies, and generally to describe what is and in what context (Isaac & Michael, 1997: 136). In addition, Muhadjir (2002: 63) said one type of survey research is to get basic data to get a general overview that is useful for planning and public policy. On the other hand Kraemer (1991) mentions three characteristics of survey research. Firstly, this research is used to describe quantitatively specific aspects of a population. Second, the data needs for research is collected from subjective perspective. Third, survey research using a portion of the population can then be generalized. Moreover, survey research can be used for evaluation objectives.

This study belongs to the third category of Kraemer’s identification which use a part of big population then generalize and be used to purpose evaluation. Final of this study have purpose to make recommendation for Muhammadiyah elementary and secondary education assembly policy can be better.

This research was conducted in Muhammadiyah Junior High Schools in Yogyakarta by analyse the difficulty level and discrimination index of academic summative test items about the material of Aqidah Akhlaq Odd semester, 2018/2019 academic year is then identified into the criteria that have been set. Sampling is done by purposive sampling technique, with surveying 200 8th grade students. The object of this study was 40 multiple choice questions (MCQs), 8th grade Aqidah Akhlaq subjects provided by the Muhammadiyah elementary and secondary education assemblies. The data obtained is analysed using the Microsoft excel application, the *Anbusho* tool version 8.0 and had to use key answer of the exam to make clear the analyse the item.

Table 1. Key Answer of all item test

1	C	11	D	21	C	31	C
2	B	12	C	22	D	32	B
3	A	13	A	23	B	33	B
4	B	14	B	24	C	34	D
5	D	15	A	25	A	35	D
6	C	16	D	26	A	36	C
7	C	17	C	27	D	37	B
8	D	18	D	28	A	38	C
9	A	19	A	29	C	39	A
10	B	20	B	30	B	40	B

Data organization is done by entering the summative test results of students into Microsoft Excel using *Anbusho* version 8.0. The items in the test should be proportional (not too easy and not too hard), like table 2 below:

Table 2. Coefficient ranges of difficulty index

Coefficient Ranges	Categories
0.000–0.299	Easy
0.300–0.699	Moderate
0.700–1.000	Hard

A good discrimination index has ranges like table 3 below:

Table 3. Coefficient ranges of Discrimination Index

$0.4 \leq x \leq 0.69$	good
$0.2 \leq x \leq 0.39$	fair
$x < 0.2$	poor

### IV. RESULTS & ANALYSIS

Item statistics are used to assess whether test items work well or not. It provides information about the level of difficulty and discrimination index. Item Analysis is very much needed in learning evaluation to increase effectiveness. Each item contributes on analyse and assess. In the current study, the result like the following figure:

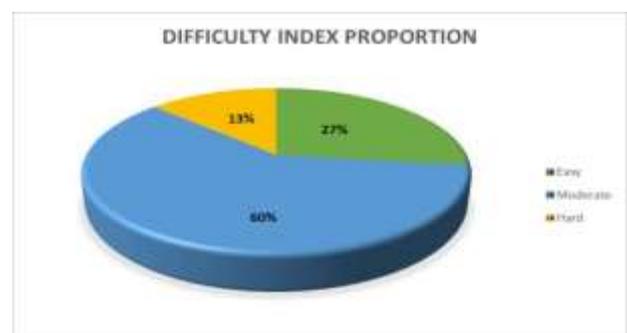


Figure 1. Difficulty Index Proportion

Item difficulty index is the percentage of pupils who answering the items test correctly. It is the relative frequency

examinees choose the correct answer (Thorndike, Cunningham, Thorndike, & Hagen, 1991). Items ranges from a low of 0 to a high of +1.00. The higher ranges, it indicates easier items, and vice versa.

The quality of test items is very influential for teachers in knowing the success of learning of students if the test items used are still far from good quality, so the evaluation results will inform the wrong results related to the learning success of students (Hasna, 2017: 4). According to Arikunto (2009) numbers that show easy and difficult items are called difficulty index. The higher the difficulty index, the harder the item of test, and vice versa. Analysis of the difficulty index of difficulty is the activity of reviewing test items from the difficulty factor so that it can be known which test items are classified as easy, moderate, and hard.

Comparisons between easy-to-moderate and hard questions can be made with 3-4-3 ratio, meaning that in the question there are 30% questions with easy categories, 40% with moderate categories, and 30% questions with hard categories. Another comparison that can be used is the ratio of 3-5-2, which means that 30% is easy, 50% is moderate, and 20% is hard.

The result of this analyse, it can be seen from the pie chart above which the proportion not balance like the theory 3-4-3 ratio, moreover 3-5-2 ratio. So, it can be said that need to modify or reject some numbers.

According to Witherington in Arikunto (2009: 207) the range of index numbers in the difficulty of question items is 0.00 to 1.00. The smaller the index of difficulty, the more difficult the problem. The classification used to interpret the results of the calculation of the difficulty level of the item questions, namely 0,000-0,299 included in the hard category, 0,300-0,699 included in the moderate category, and 0,700-1,000 included in the easy category.

Table 4. Ranges of Coefficient and Classification each number

Coefficients	Categories	Test Items
0 – 0.299	easy	1, 2, 10, 15, 23, 26, 31, 35, 36, 38, 39 (11 items)
0.3 – 0.699	moderate	3, 8, 9, 11, 12, 13, 14, 16, 17, 18, 19, 20, 21, 22, 24, 25, 27, 28, 29, 30, 33, 34, 37, 40 (24 items)
0.7 - 1	hard	4, 5, 6, 7, 32 (5 items)

The figure below is the distribution of item of test as whole is included easy, moderate and hard categories.



Figure 2. The Distribution of Difficulty Index of Items

On the result of analyse discrimination index used application of *Anbusho* verse 8.0 is like the following figure, nevertheless before that we need to pay attention on this table:

Table 5. Ranges of Coefficient and Classification each numbers of discrimination index

Coefficients	Categories	Test Items
$0.4 \leq x \leq 0.69$	good	8, 9, 16, 17, 18, 21, 28, 30, 34, 38
$0.2 \leq x \leq 0.39$	fair	1, 2, 3, 6, 7, 10, 13, 14, 15, 19, 20, 22, 23, 24, 25, 26, 27, 29, 31, 33, 35, 36, 37, 39, 40
$x < 0.2$	poor	4, 5, 11, 12, 32

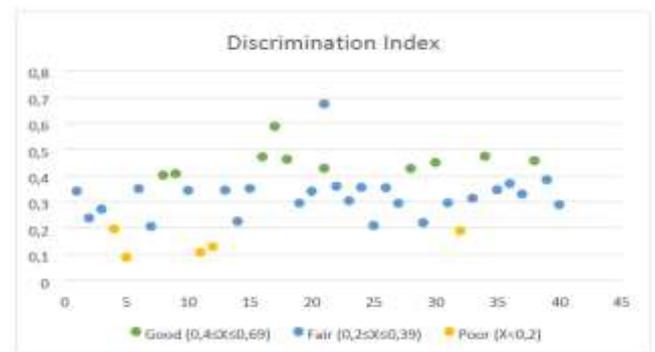


Figure 3. Distribution of discrimination index on each item

The calculation of discrimination index is a measurement of the extent to which an item able to distinguish between smart students and those who cannot do the test. The higher of the distinguishing coefficient, the more able to distinguish between the two. Items ranges from a low of -1.00 to a high of +1.00.

Ngalim (Hasna, 2017: 60) states that differentiation is a matter of ability to differentiate between students who belong to not smart group and students belonging to the smart group. In another definition that the problem differentiating index is the ability of the item in distinguishing students who have mastered the material and students who have not mastered the material. Septiana (2016) a test is said to be good if it has good differential power, if given a smart student then the results are good and if given to students who are weak then the results are low, if not so the test is said not to have good differential power. Good test with different power can produce a picture of results that are relevant to the ability of the actual learners.

Sudijono (2015) pointed out that for the follow-up of the results of discrimination strengths are: initially, the test object which has the power of good discrimination can be used again in the next test; in the following stage, the test items with the power of discrimination that are categorized as bad can be corrected and can be used in subsequent tests or discarded; later, test items with the power of negative discrimination are discarded. The benefits of the power of discrimination on each question is to improve the quality of question items based on the strength index. Based on the

empirical data, each item can be known whether it is feasible to be used, revised, or rejected. In the same case, it can be found out about how pupils able to understand the materials.

If there are items that cannot distinguish students' abilities, there are possibilities for several things to happen: 1) There are 2 correct answers in one question; 2) Incorrect answers key; 3) The competency that becomes the reference is not clear; 4) Questions with material that are very difficult, so students only guess, and 5) Distractors that do not function properly.

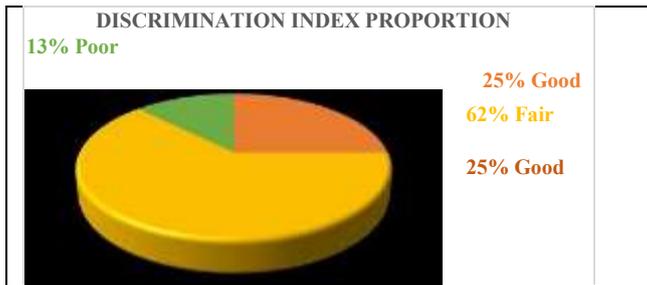


Figure 4. Discrimination index proportion

The result of this analyse, it can be seen from the pie chart above which the proportion 25 % (10 item of test) and 62 % feasible to be used that are number 1, 2, 3, 6, 7, 8, 9, 10, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 33, 34, 35, 36, 37, 38, 39, and 40. 13 % for be revised, that are number 4, 5, 11, 12, and 32. There is nothing to be rejected, because reject category just has negative value 1.

#### V. CONCLUSION & RECOMMENDATION

In sum up, we can say that result of this study significance for student, teachers and test developer. Especially for teacher or test developer, they should be very careful while selecting items. The size of a proportional item will depend on the quality of questions, the range of difficulty index and the purposes for which the test has been designed. The difficulty index should be in balance, not too easy and not too hard (moderate) namely proportional and having good proportion like theory. The discrimination index still needs to develop or to be feasible more to make good items. Enhancing the quality of test especially on administering and developing Multiple Choice Questions (MCQs) will make the teachers easier in assessing the quality of learning and more able to measure the capability pupils into learning competences.

#### REFERENCES

- Arikunto. Suharsmi (2009). *Dasar-Dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara.
- Boopathiraj, C., Chellamani, K. (2013). An Analysis of Test Items on Difficulty and Discrimination Index in The Test for Researcher in Education. *International Journal of Social Science & Interdisciplinary Research* Vol 2 (2).
- Firman. (2017). [skripsi]. Jakarta: Universitas Islam Negeri Syarif Hidayatullah.
- Hasna,A. (2017). [Skripsi]. Yogyakarta: Universitas Negeri Yogyakarta.
- Hingorjo, MR., Jaleel F. (2012). Analysis of one-best MCQs: the difficulty index, discrimination index and distractor efficiency. *J Pak Med Assoc* 2012 Feb;62(2):142-7.
- Isaac, S., & Michael, W. B. (1997). *Handbook in research and evaluation: A collection of principles, methods, and strategies useful in the planning, design, and evaluation of studies in education and the behavioral sciences*. (3rd Ed.). San Diego: Educational and Industrial Testing Services.
- Juliyanto, J., Murtinugraha, R. E., & Arthur, R. (2016). Kualitas Butir Tes Pilihan Ganda Pada Program Keahlian Teknik Gambar Bangunan. *Jurnal PenSil*, 5(1).
- Kheyami, D., Jaradat, A., Al-Shibani, T., & Ali, F. A. (2018). Item Analysis of Multiple-Choice Questions at the Department of Paediatrics, Arabian Gulf University, Manama, Bahrain. *Sultan Qaboos University medical journal*, 18(1), e68–e74.
- Kraemer, K. L. (1991). Introduction. Paper presented at The Information Systems Research Challenge: Survey Research Methods.
- Kurniawati, F. E., & Miftah, M. (2015). Pengembangan bahan ajar aqidah ahklak di madrasah ibtidaiyah. *Jurnal Penelitian*, 9(2), 367-388. *Akuntansi Indonesia*, 12(2).
- Muhadjir, Noeng. 2002. *Metodologi Penelitian Kualitatif*. Yogyakarta: Rake Sarasin.
- Purwanti, M. (2014). Analisis Butir Soal Ujian Akhir Mata Pelajaran Akuntansi Keuangan Menggunakan
- Rao C, Kishan Prasad H L, Sajitha K, Permi H, Shetty J. (2016) Item analysis of multiple choices questions: Assessing an assessment tool in medical students. *Int J Educ Psychol Res* 2016; 2:201-4.
- Septiana, N. (2016). Analisis Butir Soal Ulangan Akhir Semester (UAS) Biologi Tahun Pelajaran 2015/2016 Kelas X dan XI pada Man Sampit. *Edu Sains: Jurnal Pendidikan Sains & Matematika*, 4(2).
- Singarimbun, Masri dan Effendi, Sofian. 1995. *Metode Penelitian Survey*. Jakarta: PT Pustaka LP3ES.
- Thorndike, R. M., Cunningham, G. K., Thorndike, R. L., & Hagen, E. P. (1991). *Measurement and evaluation in psychology and education* (5th Ed.). New York: MacMillan.
- Undang-Undang Republic Indonesia Nomor 14 Tahun 2005 Tentang Guru dan Dosen
- Yanti, S. F., & Tantoro, S. (2017). Pengaruh Pembelajaran Aqidah Akhlak Terhadap Perilaku Siswa Di Madrasah Aliyah Negeri Kampar Timur. *Jurnal Online Mahasiswa (JOM) Bidang Ilmu Sosial dan Ilmu Politik*, 4(1), 1-12. Microsoft Office Excel 2010. *Jurnal Pendidikan*