Criteria For Good Test Questions (Analysis Of Writing Of Islamic Junior High School Exam Questions For Arabic Language Subjects At The Ta'mirul Islam Islamic Boarding School)

Luthfiyyah Nur Afifah
Postgraduate at UIN Raden Mas Said Surakarta
Email : luthfiyyah.nur19@gmail.com

Abstract
The aim of this research was to analyze the odd exam question in Arabic subject to the ninth-grade students at Boarding School of Ta'mirul Islam Surakarta in academic year 2020/2021 which include the validity, power of difficulty, discrimination power, and answer patterns. This method used in the research was a descriptive quantitative method. The result of the analysis showed that the quality which was seen from the 1) validity there were 11 or 27,5% questions which were valid and the rest were 29 or 72,5% of questions were invalid. 2) from the level of difficulty, there were 4 or 10% of questions which average level of difficulty, and there were 32 or 90% of questions which had difficult and easy level. 3) From the discrimination power there were 10 or 25% of questions which had good discriminations and 30 or 75% of questions which had low and negative. 4) from the answer patterns, it is known that in general all distributor in the question were unfunctional properly.

Keywords: *Good questions criteria, Analyze exam question*

INTRODUCTION

As we all know, teachers play a very important role in assessment. This is because the teacher's efforts to improve the quality of the questions they prepare are very important. However, this is often not implemented because of a person's tendency to think that the results of what they do are good. Likewise, teachers who have experience teaching and compiling test questions still find it difficult to realize that the test is still not at a perfect level. Therefore, it is very necessary to conduct an analysis of the writing of test items before they are tested on students.

This assessment is very important in learning evaluation. As Fachry Thaib said, assessment is an effective measuring tool and is very important in evaluating learning, in order to obtain a score on the learning outcomes of someone who is studying within a certain period of time (Fachry Thaib, 2003:20).

In assessment, the methods commonly used are demonstration, observation and tests. The use of this method is chosen based on the learning competency objectives to be measured. Competencies related to skills are measured using the demonstration method. Competencies related to attitudes are measured using the observation method. And competencies related to knowledge are measured using test methods

The test can be done orally or in writing. Oral tests are carried out using questions orally which are then answered by students orally. An oral test can be given at the end on each basic competency of the theme given by the teacher. Oral tests are given so that teachers can find out students' responses or answers directly.

Meanwhile, written tests are tests in written form, the answers to written tests are also in written form. The written test assessment aims to measure aspects of students' knowledge. Written tests can be objective tests, namely tests that require students to choose the answers that have been provided. Then the subjective test is a written test that asks students to provide long descriptive answers or sentences.

According to Suharsimi Arikunto (2008:207), a good test instrument must have the following requirements:

1. Validity

   A test can be said to be valid if the test is truly suitable for measuring what it is supposed to measure. A test can be said to have high validity if the text carries out its measuring function, or provides precise and accurate measuring results according to what is used in the test. The test is valid if the calculated R correlation value > R table (Sugiyono, 2008:248).

2. Reliability

   The reliability of a test is the level to which the test is able to demonstrate the consistency of the measurement results shown, in terms of the accuracy and accuracy of the results. A test is called reliable if the reliability figure is greater than 0.6 (Ghozali, 2005:129)

3. Objectivity

   The next requirement for a good test is that it is objective, meaning that there are no personal elements that influence the scoring system.

   According to Ngalim Purwanto, 1994, a test can be said to be objective if the test is prepared and implemented according to what it is. Viewed in terms of content or material, this means that the test material is taken or sourced from the material or learning material that has been provided which is used as a reference in compiling the results of the learning test.

4. Practicality

   A good test must also be practical, namely easy to carry out, easy to examine and equipped with clear instructions (Arikunto, 2008:62)

   According to Anas Sudijono 2006:97 practicality implies that learning outcomes tests must be able to be carried out easily because the tests are simple and complete. Simple in the sense that it doesn't require a lot of equipment or equipment that is difficult to procure. Complete in the sense that the test is equipped with instructions for taking it, an answer key, as well as guidelines for scoring and scoring.

5. Economical

   Apart from that, the test must also be economical, that is, its implementation does not require expensive costs, lots of energy and lasts a long time (Sugianto: 2016)

   Specifically for objective test instruments, apart from having these characteristics, they must also have the following requirements:
   1. Difficulty level is medium
   2. Good distinguishing power
   3. Good answer patterns (Arikunto,2008:57)

Based on the results of an interview with one of the teachers at the Ta'mirul Islam Islamic Boarding School, namely Ustadz Mohammad Hatta, on September 10 2021, the Islamic Junior High School exam questions were made based on the 2013 curriculum, while the Arabic language subjects at Ta'mirul Islamic Boarding School were not based on the 2013 curriculum but rather the Education Unit curriculum. which is based on Islamic boarding schools and so far, there has never been an analysis of the exam questions. The analysis in question is that the existing questions are then tested and corrected directly by the teacher in charge of the subject, then analyzed to find out whether the questions asked meet the standards.

Thus, the aim of this research is to examine the validity, error rate, differentiating power and answer patterns for Madrasah Exam questions for Arabic Language Subjects at the Ta'mirul Islam Islamic Boarding School in academic year 2020/2021.

METHOD

The method used in this research is a quantitative descriptive method. The sample in the research was 20 students from class IX of the Ta'mirul Islam Islamic Boarding School in Surakarta in academic year 2020/2021. Data collection techniques use test techniques and documentation techniques.

Data analysis as follows:

1. Validity of the test items, to be able to conclude whether the items are valid or not, analysis is carried out using the product moment correlation formula, both with deviation formulas and rough numerical formulas.

$$\text{rxy} = \frac{n\sum XY - (X)(\sum y)}{\sqrt{\{n\sum x^2 - ()\sum x)^2\}\{n\sum and^2 - (\sum y)^2\}}}$$

Information:

   rXY = correlation coefficient between variable X and variable Y, which in this case is considered as the item validity coefficient.

n = Number of Samples (number of students) X = Item score Y = Total score (Arikunto, 2008: 72)

2. Level of Difficulty of Question Items. To determine the level of difficulty of question items, the author uses the following formula:

$$P = \frac{B}{JS}$$

Information:

P = item difficulty index

B = the number of students who answered the question correctly regarding the item in question

JS = the total number of test takers who took the learning outcomes test (Anas Sudijono, 2001:372)

To determine the differentiating power of test items, researchers use the following formula:

$$D = \frac{NOT}{AND} - \frac{BB}{JB} = PA - PB$$

Information:

D = Daya Beda

JA = number of participants in the upper group

JB = number of lower group participants

BA = the number of participants in the upper group who answered the question correctly

BB = the number of lower group participants who answered the question correctly

PA = proportion of participants in the upper group who answered correctly (remember, P is the index of difficulty)

PB = proportion of lower group participants who answered correctly (Suharsimi Arikunto, 2010: 213-214)

3. The answer pattern is the distribution of test takers in terms of determining answer choices on questions in multiple choice form. The question answer pattern is obtained by counting the number of test takers who have answers a, b, c, d, e or who do not choose any answer.

Analyzing distractor function is often known by another term, namely analyzing the distribution pattern of item answers. What is meant by item answer distribution pattern is a pattern that can describe how test takers determine their answer choices based on the possible answers that have been paired with each item. A distractor is declared to function well if it is chosen by at least 5% of test participants (Sudijono 2001:411, Kadir: 2015)

RESULTS AND DISCUSSION
1. Validity of Question Items

Validity test is a test used to show the extent to which the measuring instrument used in measuring something is being measured. Ghozali (2009) stated that the validity test is used to measure whether a questionnaire is valid or not.

An item can be said to have high validity or can be declared valid, if the scores on the item in question have conformity or alignment with the total score or in statistical language there is a significant positive correlation between the item score and the total score (Anas Sudijono, 2001:184).

The following is an analysis of the validity of the Madrasah Subject Exam questions

Arabic language class IX at the Ta'mirul Islam Islamic Boarding School in Surakarta:

HASIL ANALISIS VALIDITAS BUTIR SOAL
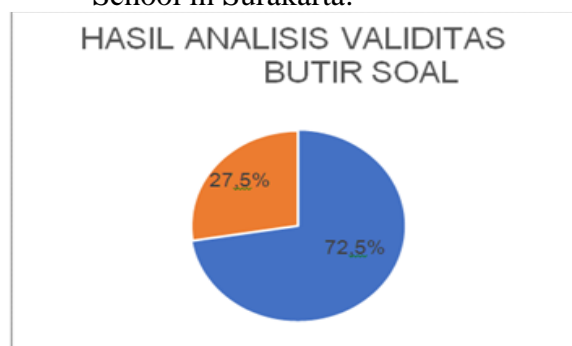
The results of the analysis of the validity of the question items show that: there are 11 or 27.5% of the question items that are declared valid. The valid question items are in questions no. 5, 7, 11, 15, 18, 19, 20, 27, 32, 33, and 38. Eleven of these question items are declared valid because they have been proven to have a significant correlation value with the total value. .

Meanwhile, 29 or 77.5% of the other questions declared invalid were questions number 1, 2, 3, 4, 6, 8, 9, 10, 12, 13, 14, 16, 17, 21, 22, 23, 24, 25, 26, 28, 19, 30,

31, 34, 35, 36, 37, 39, 40. These question items need to be revised if they are to be used again. This shows that when writing questions, trials and analysis are very necessary, then revisions are carried out before being tested on students.

2. Difficulty Level

According to Sukardi (2016:387), the level of difficulty is a number that shows the proportion of students who answered correctly in one question using an objective test.

The quality of the learning outcomes test items can be determined from the degree of difficulty or level of difficulty of each item. Learning outcome test items can be stated as good items if the items are not too difficult and not too easy, in other words the degree of difficulty of the questions is sufficient or moderate (Sudijono, 2009: 370).

Good questions are questions that are neither too easy nor too difficult. From this statement it can be interpreted that a good question must have a moderate or sufficient level of difficulty.

The following are the results of the analysis of the level of difficulty of the Madrasah Exam questions

Arabic language subject class IX at Pondok Ta'mirul Islam Surakarta.

HASIL ANALISIS TINGKAT KESUKARAN BUTIR SOAL
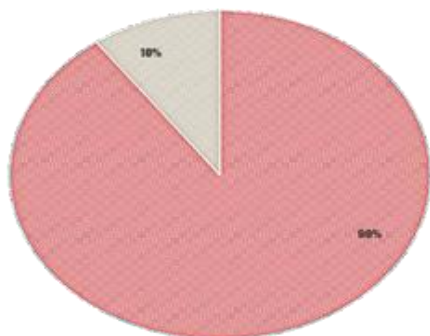
From the results of the analysis using manual calculations on the Arabic Language Madrasah Exam questions at the Ta'mirul Islam Islamic Boarding School in academic year 2020/2021, it is known that there is 1 question or 2.5% of the questions that have a difficult classification, namely question number 29.

For questions that are included in the difficult category, there are three possible follow-up actions, namely:
1. These items are discarded or dropped and will not be issued again in future learning outcomes tests.
2. It is re-examined, tracked and traced so that the factors that cause the item in question to be difficult for the test taker to answer can be identified. Are the sentences in the questions unclear, are the instructions on how to do (answer) the questions difficult to understand, or are there unclear terms in the questions?
3. Items that are too difficult can still be useful at any time, namely they can be used in tests (especially selection tests) which are very strict in nature.

It is said that good questions are items that have a medium level of difficulty. Based on the results of the analysis of the questions, there are 4 or 10% of the questions that have a medium classification. These questions are found in numbers 7, 14, 18, and

19. For items that are included in the medium category, it is best to immediately document them and then use them again in tests in the future.

Questions that are easy for students to answer correctly can also be included in the criteria for bad questions. In this case there are 35 or 87.5% of the questions that have an easy classification. Namely questions number 1, 2, 3, 4, 5, 6, 8, 9, 10, 11, 12, 13, 15, 16, 17, 20, 21, 22, 23, 24, 25, 26, 27, 28, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, and 40. Question items that are too easy can still be used for selection tests that are too loose in nature.

Based on the analysis of question items, there are 10% of question items that fall into the good category, and 90% fall into the bad category. Either because it's too hard or too easy.

3. Differentiating Power

The differentiating power of a question is the ability of a question item to differentiate between

students who have mastered the material being asked and students who have not? lack of mastery of the questions asked. Or in other words, it is an index of differences between high and low ability groups (Khaeruddin: 2017).

Differentiating Power is basically calculated on the basis of dividing training participants into two groups, namely the upper group and the lower group. The upper group is classified as an intelligent group, while the lower group is a group that is classified as less intelligent (Bagiono: 2017).

A test is said to have good discriminating power, if the questions can differentiate between students who are intelligent and students who are less intelligent, or to the extent to which the number of correct answers from students belonging to the upper group is different from students belonging to the lower group.

The criteria for the differential power index range from -1.00 to +1.00. The higher the discriminating power of a question, the better the question. If the differential power is negative (-) or has an index less than 0, it means that more of the lower group answered correctly than the upper group. Or in other words, the question items are not or are less functional (Gene, Glass, Julian 1970: 169; Koyan, 2012: 63). The following is an analysis of the differentiating power of the Islamic Junior High School exam questions

Class IX Arabic Language Subjects at Pondok Ta'mirul Islam:

Through analysis the differentiating power of the questions that have been carried out is known

Hasil Analisis Tingkat Daya
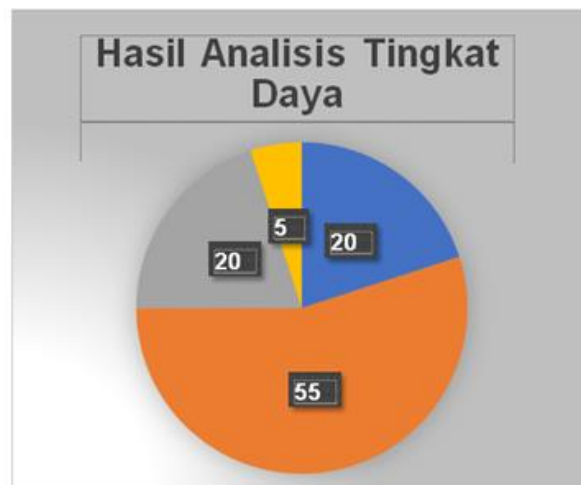
That there were 22 questions that had a bad classification. These questions are found in questions number 5, 6, 8, 9, 10, 11, 12, 13, 14, 16, 20, 21, 22, 23, 25, 26, 27, 28, 31, 37, 38, 40. Items that have poor classification can be traced back to be corrected and used again in learning tests, then the items are analyzed again to see whether the distinguishing power has increased or not.

Meanwhile, the items that have sufficient distinguishing power are classified as 20%, namely in numbers 4, 7, 19, 24, 32, 33, 34, and 36. The items that have a good classification are 2 items or 5% in number 15. and 18. Items that have adequate and good classification are items that have good differentiating power. This is because students who are in the clever category answer more correctly.

4. **Answer Pattern**

From the test analysis of question answer patterns, it can be determined whether the distractor is functioning properly or not. Distractors that are not selected at all by the testes are bad distractors. On the other hand, a distractor can be said to function well if the distractor has a great appeal to test takers who do not understand the concept or do not master the material (Arikunto, 2011: 233-238)

The multiple-choice objective test for each question item is equipped with several possible answers or what is often known as options or alternatives. There are 3 options or alternatives for the Odd Semester Exam questions for Class IX Students at the

Ta'mirul Islam Islamic Boarding School in academic year 2020/2021. Possible answers are attached to each question item, one of which is the correct answer (answer key) and the rest are incorrect answers. Wrong answers are what are commonly known as distractors. The number of distractors in this exam question is 120 options.

The purpose of installing distractors is to make the teste interested in choosing one of the available distractors. Distractors are said to be able to carry out their function well if students taking the test feel uncertain and choose one of the distractors as the answer.

The following are the results of the analysis of answer patterns from the Islamic Junior High School exam questions for class IX Arabic at the Ta'mirul Islam Islamic Boarding School in Surakarta:

Based on the analysis carried out, there were 75 or 62.5% distractors that were not chosen by students out of the 20 students. Because the student does not feel stimulated by the existing distractors. This shows that the distractor does not carry out its function as a distractor. The distractors are spread across questions number 1, 2, 3, 4, 5, 6, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, and 40.

Meanwhile, there are 45 or 37.5% of distractors that have carried out their function. These distractors are distributed in all the Arabic language exam questions for class IX at the Ta'mirul Islam Islamic boarding school in Surakarta in academic year 2020/2021 except for numbers 3, 5, 6, 8, 11, 14, 17, 21, 22, 25, 26, 28, 35, and 37

CONCLUSION

Based on the results of the research, data analysis and discussion, it can be concluded that in preparing the item analysis for the Arabic language exam for class IX at the Ta'mirul Islam Islamic boarding school in Surakarta in academic year 2020/2021, it is as follows:

Validity Most of the exam questions are invalid. There were 29 or 77.5% of questions declared invalid, namely questions number 1, 2, 3, 4, 6, 8, 9, 10, 12, 13, 14, 16, 17, 21, 22, 23, 24, 25, 26, 28, 19, 30, 31, 34, 35, 36, 37, 39, and 40.

The difficulty level indicates that the question has a poor level of difficulty. Items that have a poor level of difficulty are those that are classified as too easy and too difficult which are found in questions number 1, 2, 3, 4, 5, 6, 8, 9, 10, 11, 12, 13, 15, 16, 17, 20, 21, 22, 23, 24, 25, 26, 27, 28, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, and 40.

The answer pattern shows that the question has a bad answer pattern, this is because 75 or 62.5% of the distractors were not chosen by students out of the 20 students. The distractors are found in questions number 1, 2, 3, 4, 5, 6, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, and 40.

BIBLIOGRAPHY

Arikunto, S. 1993. Manajemen Pengajaran. PT Rineka Cipta. Jakarta Arikunto, S. 2008. Dasar-Dasar Evaluasi Pendidikan. Jakarta: Bumi Aksara.

Arikunto, S. 2012. Dasar-Dasar Evaluasi Pendidikan Edisi kedua. Jakarta : Bumi Aksara Arikunto, S. 2018. Dasar-Dasar Evaluasi Pendidikan Edisi 3. PT Bumi Aksara. Jakarta Bagiyono. 2017 Analisis Tingkat Kesukaran dan Daya Pembeda Butir Soal Ujian Pelatihan Radiografi Tingkat 1. Batan : Widyanuklida Vol 1.

Bumi Aksara.Thaib Fachri. 2003. Evaluasi Pengajaran Geografi (Buku Ajar). FKIP: Unisula

Gene, Glass, & Stanley, Julian C, 1970. Statistical Methods in Education and Psichology.New Jersey ; Prentice Hall, Inc.

Ghozali, Imam. 2005. Aplikasi Analisis Multivariate dengan SPSS. Semarang: Badan Penerbit UNDIP.

Ghozali, Imam. 2009. Aplikas Analisis Multivariete dengan Program SPSS. Semarang : UNDIP.

Kadir, Abdul. 2015 Menyusun dan Menganalisis Tes : Fakultas Tarbiyah dan Ilmu Keguruan IAIN Kendari.

Khaeruddin, Administrasi, Analisis, dan Kaidah Penulisan Tes (Jurnal Madaniyah Vol 1 Edisi XII Januari 2017)

Koyan I Wayan. 2012. Kontruksi Tes. Singaraja: Universitas Pendidikan Ganesha. Purwanto, M Ngalim, 1994 Prinsip-Prinsip dan Teknik Evaluasi Pendidikan.Bandung:Remaja Rosda Karya.

Sudijono Anas, 2009. Pengantar Evaluasi Pendidikan, Jakarta: Raja Grafindo.

Sudijono Anas. 2001. Pengantar Evaluasi Pendidikan. PT Raja Grafindo Persada.Jakaarta.

Sudijono, S. 2006. Pengantar Evaluasi Pendidikan. Jakarta: Rajawali Tekan

Sudjana Nana. 2008. Penilaian Hasil Proses Belajar Mengajar. PT Remaja Rosdakarya: Bandung.

Sugianto, Aris. 2016. Ciri-Ciri (Karakteristik) Tes yang Baik. IAIN Palangkaraya. Sugiyono. 2008. Metode Penelitian Kuantitatif Kualitatif dan R&D. Bandung : ALFABETA Sukardi. 2011. Evaluasi Pendidikan: Prinsip dan Operasionalnya. Jakarta Timur : PT.