

The Analysis on the Quality of Test Items of the Summative Test

Rabiatul Adawiah

Faculty of Teacher Training and Education
University of Lambung Mangkurat
Banjarmasin, Indonesia
rabiatuladawiah666@yahoo.com

Abstract—This study aimed to determine the quality of grains from the Final Exam of Junior High School, in Balangan Regency subjects Citizenship Education Year 2015/2016, which is indicated from the level of difficulty, discriminator index, and the effectiveness of the distractors. The subjects of the study were 200 students of junior high schools in Balangan Regency. The objects of the study were 50 summative tests items and the answer key. The data from documentation were analyzed by deploying the ANATES program version 4.1.0 to investigate the item difficulty, discriminator index, and the effectiveness of the distractors. The findings of the analysis revealed that: (1) based on the level of difficulty, there were 23 very easy items (46%), 11 easy items (22%), and 16 medium items (32%); (2) based on item discrimination, 21 items (42%) were categorized as poor discriminator, 16 items (32%) were categorized as fair discriminator, 11 items were categorized as good discriminator, and 2 items were categorized as very good discriminator; (3) based on the effectiveness of the distractor, functioned poorly was 12 items (24%), fair distractors are 16 items (32%), good distractors are 13 items (26%), and very good distractors are 9 points (18%).

Keywords—*Item Difficulty, Item Discriminator Index, Effectiveness Of The Distractors, Test Item*

I. INTRODUCTION

Teachers as the main component in education are required to balance the development of science and technology in society. It is because teachers have an important role in shaping the intelligence of the community, in the educational environment. Education must create qualified human resources, both scientifically and mentally. Therefore, it takes teachers who are professional in educating students, who are superior and qualified. Professional teachers are teachers who have a number of competencies that can support their duties.

Law of the Republic of Indonesia Number 14 Year 2005 on Teachers and Lecturers article 10 paragraph 2 mentions that there are four competencies that must be owned by a teacher, one of which is the pedagogic competence obtained through professional education. In pedagogic competence, teachers are required to conduct learning evaluation activities. Evaluation requires teachers to measure and assess the level of achievement of a program that has been implemented. This is in accordance with the Regulation of the Minister of National Education No. 41 of 2007 on Process Standards stating that

“Evaluation is done by educators on the learning outcomes to measure the level of achievement of competence learners, and it is used as the progress report of learning outcomes and to improve the learning process.” In the learning process, evaluation is one of the important components and steps that must be taken by the teacher to assess the effectiveness of learning [1].

Evaluation as a value-setting process relates to the performance and work of learners [2]. Evaluation is a measurement and assessment activity. Assessment is carried out after measurement and measurement is the basis of the assessment [3]. Measurement is defined as the scoring of student learning outcomes. Assessment is a systematic activity in collecting information about the process and learning outcomes of learners in order to make decisions based on certain criteria and considerations [1]. An evaluation tool commonly used in teaching and learning is a test. The test is a way or procedure that needs to be taken in the framework of measurement and assessment in education [4]. The test is as a task or a series of tasks [5]. Furthermore, it is said that the task can be in the form of assignment tasks or commands that must be done by students, so the value obtained symbolizes the behavior or achievement of learning outcomes of learners.

School final exams is part of the evaluation that aims to measure and assess the competence of students, so teachers can determine whether the students can pass or not. Therefore the questions used should be quality questions. One attempt to develop test questions with good quality is to analyze the items. The item analysis is the process of collecting, summarizing and using information from the learner's answer to make decisions about each assessment [2]. The item analysis aims to identify good, fair, and poor questions [6].

The activity of analyzing the items is one of the "obligations of every teacher". It is considered an obligation because each teacher in the end must be able to provide information to the institution or to the students themselves about how far the mastery and the ability that has been achieved by students to the materials and skills about the subjects that have been given. Reality on the ground often indicates that the grades obtained by the students from the test results are still low. The low test results of the students are not only caused by the low ability of students to answer the problem, but also because of the low ability of teachers in doing evaluation and compile the test tool/evaluation tool.

Evaluation tools in the form of written tests should have the characteristics or requirements as a good evaluation tool, among which must meet the requirements of the level of difficulty, discriminating power, and effectiveness of the distractor. Therefore, analyzing the items should be a series of learning activities that can not be left behind. So far, analysis of the item is rarely done, especially about the School Final Exam made by the teacher. Therefore this research was very important to do. Thus, later it can be obtained the picture of how the quality of exam questions in Balangan Regency, especially for the subject of Pendidikan Kewarganegaraan (civic education) in Junior High School.

II. METHODS

This research was an evaluation research, where the design and evaluation procedure in collecting and analyzing data are done systematically. This study aimed to determine the value or benefits of an educational practice. Evaluation in this research was conducted on civic education Final Test in state junior high school in Balangan Regency academic year 2015/2016.

Research subjects were junior high school students in Balangan Regency with a sample size of 200 people. The object of research was School Final Examination (UAS), which amounted to 50 questions, answer key, and answer examinees. Data were obtained by documentation technique and analyzed by computer to see level of difficulty, discrimination power, and effectiveness of distractor usage. The analysis of items with the computer is a quantitative item questionnaire with the calculations of computer programs [11]. This method was appropriately used because the calculation accuracy rate is higher than the manual processing with the calculator. In this research the computer program used was ANATES program version 4.1.0.

III. RESULT AND DISCUSSION

A. Result

1) Difficulty Level of Test Items

The difficulty level of the item is the opportunity to correctly answer a question at a certain level of learners' ability. The criteria to interpret the results of the item are as follows: 00.00 - 25.00 is very difficult category; 26.00 - 45.00 is a difficult category; 46.00 - 65.00 is included in the medium category, 66.00 - 85.00 is included in the easy category, and 86.00 - 100.00 is very easy.

TABLE I. DISTRIBUTION OF TEST ITEMS BASED ON THE DIFFICULTY LEVEL

No	difficulty level	Number of Questions	Items	%
1	Very easy	23	1,3,4,7,8,13,15,16,21,25,26,33,34,35,36,37,39,41,42,43,46,48,49	46
2	Easy	11	2,5,6,11,17,18,19,22,23,28,38	22
3	Medium	16	9,10,12,14,20,24,27,29,30,31,32	32
4	Difficult	0	-	0
5	Very difficult	0	-	0

Based on the results of the analysis of the final examination of the State Junior High School in Balangan Regency The academic year of 2015/2016 subjects of civic education, from 50 items, which is very easy about 23 items (46%), 11 items are easy items (22%), and 16 items are medium categories (32%). Distribution of test items based on difficulty can be seen in the Table I.

2) Discriminating Power

Discriminating Power is the ability of the problem in differentiating smart students with students who are less smart. The interpretation of discriminating power is as follows: 00.00 - 20.00 is poor discriminator; 21.00 - 40.00 is included in fair discriminator; 41.00 - 70.00 is a good discriminator, and 71.00 - 100.00 is very good discriminator.

The analysis results of Civic Education final exam for State Junior High School, in Balangan Regency, the academic year 2015/2016 as follows: 21 items (42%) were categorized as poor discriminator, 16 items (32%) were categorized as fair discriminator, 11 items were categorized as good discriminator (22%), and 2 items were categorized as very good discriminator (4%). The distribution can be seen in the Table II.

TABLE II. DISTRIBUTION OF ITEMS BASED ON DISCRIMINATING POWER

No	Discriminating Power	Number of Questions	Items	%
1	Poor	22	1,3,8,13,15,16,17,21,25,26,29,31,33,34,36,39,41,42,43,46,48,49	44
2	Fair	16	4,6,7,11,14,18,19,22,23,24,27,30,35,37,38,44	32
3	Good	10	2,5,9,10,20,28,40,45,47,50	20
4	Very good	2	12,32	4

3) Effectiveness of Distractor Use

The effectiveness of the distractor works best when it is chosen by at least 5% of test takers [3]. A distractor not chosen by testee at all means that the distractor is poor, too conspicuous and misleading. On the other hand, as a distractor it has a great appeal to test takers who lack understanding of concepts or lack of material control. The indicators used to determine the effectiveness of the use of the distractor areas follows:

TABLE III. CRITERIA FOR DETERMINING EFFECTIVENESS OF DISTRACTOR USE

No	Distractors that Work	Criteria
1	0	Poor
2	1	Fair
3	2	Good
4	3	Very good

The results of the analysis by using Anates program version 4.1.0 shows that 12 items (24%) were poor distractors, 16 item (32%) were fair distractor, 13 item (26%) were good distractors, and 9 item (18%) were very good distractors.

Distribution of test items based on the effectiveness of the distractor use can be seen in Table IV.

TABLE IV. DISTRIBUTION OF TEST ITEMS BASED ON DISTRACTOR EFFECTIVENESS

No	Distractor Effectiveness	Number of Questions	Items	%
1	Poor	12	1,11,16,29,37,40,41,42,45,46,47,50	24
2	Fair	16	2,4,8,9,12,15,17,19,21,22,26,27,31,33,38,48	32
3	Good	13	3, 5,6,7,10,13,18,25,30,34,36,43,44	26
4	Very Good	9	14,20,23,24,28,32,35,39,49	18

B. Discussion

1) Level of Difficulty

The difficulty level of the item is the opportunity to correctly answer a question at a certain level of ability. Problem level is a matter of how difficult the degree of difficulty problems [1]. Furthermore, it is said that a matter is said to be good if it has a problem level that is proportional). So, the problem is not too easy or too difficult. In other words, a good question is a matter that has medium difficulty. This is understandable because if the problem is too easy, it does not stimulate students in solving problems. Conversely, if the problem is too difficult, it causes students not to have a passion in working on the problem because it is beyond the reach of students' abilities.

The results of the difficulty level analysis of the Final School Exam for junior high school in Balangan Regency on the subject of Citizenship Education are 23 very easy items (46%), 11 easy items (22%), and 16 medium items (32%).

The test item is declared good if it has a level of difficulty in accordance with the purpose of the test [7]. For example, for the purposes of the semester exams, items with moderate difficulty are used, for selection items with high difficulty level are used, and for the purposes of diagnosis difficult items are used. The level of difficulty is related to the purpose of the test, and the items used for the semester exam should have a moderate level of difficulty [8].

If referring to some of the above opinions, that a good question is a matter that has a moderate degree of difficulty, then the final examination of State Junior High School in Balangan Regency for Civic Education showed that only 16 items (32%) that meet the criteria, while the rest items are easy and very easy.

Follow-up after the items were analyzed in terms of their difficulty are as follows [9]:

- Items that are categorized as good (medium difficulty level) can be directly recorded in the question bank.
- Categorized items are very difficult, there are three possible follow-up actions that can be done: (1) the items are discarded or dropped and are not released in the next learning result test; (2) researched, tracked and traced the cause of items difficult to answer by the testee. Afterwards it is done to make the item reusable in the learning result test; (3) the items are used in tests of very strict nature (selection tests) so that it can be stored in a question bank separately.

- There are also three possible follow-ups for items in the easy category: 1) the items are discarded or dropped and are no longer issued in the learning result test; (2) re-examined, traced to determine the factors that cause the item can be answered correctly by almost all the testees. Once known to be repaired, the item is to be re-issued to find out the level of difficulty of the item; (3) used in loose tests, in the sense that a large part of the testee pass in a selection test. Under these circumstances, it is wise if item items in the category are easily excluded in the selection test.

2) Discriminating Power

Discriminating power is the ability of a test item to distinguish between students who are high-ability and who are low-ability [6]. The higher the discriminator index of questions, then the problem is able to distinguish between students "who are smart" and "who are less smart". This analysis is to examine the items with the aim to know the ability of the test item in distinguishing students who are categorized as capable with students who are classified as less capable [10]. In this case, if an item is given to a capable student, the outcome will show high achievement, and if given to a weak student, the outcome will be low.

The results of analysis using ANATES version 4.1.0. are as follows. Items classified as poor discriminator are 21 items (42%). Items belonging to fair discriminator are 16 items (32%). The items belonging to the good discriminator are 11 items (22%), and the items classified as very good discriminator power are 2 items (4%).

It is very important to know the discriminating power of the item [4]. One of the bases in compiling the test item of the learning result is the assumption that the ability of the testee with each other is different and the test result test item must be able to give test results that reflect differences in ability among testees.

A good item is an item that can distinguish clever students and students who are less clever. In this case, test items can be answered correctly by students who are clever. The problem is said to have no discriminating power if the matter is tested to high achieving students, and the result is low [10]. However, if given to children with weaker ability, the results are higher. Thus, a test item that does not have a discriminating power will not produce a result that matches the actual students' ability. Therefore, if the problem does not have a good discriminating power, there will be peculiarities.

From the description, it can be said that the discriminator of School Final Exam for State Junior High School, Balangan Regency, 2015/2016 academic year, on Civic Education subjects, are 42% categorized as poor discriminator.

The follow-up after analyzing the test' discriminating power are as follows [9]:

- Items that have a good discriminator are stored in question bank. These items can be reused during upcoming learning test.
- Items with a low discriminator can followed up by using two possibilities: (1) traced for later repair and then reused in future learning test in order to know if the discriminating power is increased or not, (2) discarded.

- an item whose negative discriminating power should be discarded because of its low quality.

3) *The effectiveness of the distractor use*

Distractor analysis is conducted for multiple choice tests, which have answer options of 3 to 5 [8]. In multiple-choice question, there is one correct answer and some wrong answers or distractors. Arifin (2013: 279) states that a good item is the distractor chosen by the test participants evenly. Conversely, items that are poor distractors are not evenly selected. A distractor works well if the deceiver is at least selected by 5% of the test participants or more chosen by the lower group [6]. In this study, items are said to be good if at least one item has two good distractors.

The effectiveness of the distractors was analyzed by using ANATES version 4.1.0. The results indicated that the poor distractors are 12 items (24%), fair distractors are 16 items (32%), good distractors are 13 items (26%), and very good distractors are 9 items (18%).

Follow-up after the distractor effectiveness analysis are as follows [3]:

- The distractors are accepted because they are good. This means that all distractors on the item have been selected 5% of the test participants.
- The distractor is rewritten if it is not good. This means that the distractor has not performed its function properly (distractor chosen less than 5%).
- The distractors are rejected because they are not good. This means that the distractors are absolutely not selected by the test participants (0%).

Referring to the above opinion, it can be concluded that from the 50 items of the School Final Examination on Civic Education for junior high school in Balangan Regency, all of 12 items must be replaced.

4) *Item Quality Based on Level of Difficulty, Discriminating Power and Effectiveness of Distractor Use*

Determining the quality of the item is done by analyzing together the characteristics of the item assessment (difficulty level, discriminating power, and effectiveness of the distractor). The quality of the item is divided into three categories: good, fair, and poor. The criteria used are (1) The item is said to be good if the item meets at least two of the three criteria (difficulty level, discriminating power and effectiveness of distractor); (2) The item is said to be medium/fair if it meets only one of three criteria; (3) Item is said to be poor if it does not meet all criteria.

The analysis results of the School Final Exam on Civic Education for junior high school in Balangan Regency, 2015/2016 school year showed that good quality items are 12 items (24%), fair items are 18 items (36%), and poor items are 20 items (40%). Follow-up which should be done after knowing the quality of the items is as follows:

- Good quality items can be directly inserted into the question bank. So, it can be reused for future learning test.

- The items that are not good enough cannot enter the problem bank because they have not fulfilled the characteristics of the good question. In this case, the item can be fixed first according to the failure indicator.
- The poor items cannot enter the question bank and should be replaced with the new ones.

The item analysis provides the following benefits: (1) determining the defective or not properly functioning test items, (2) increasing the items' quality through the three components of analysis, namely level difficulty, discriminating power, and distractor, and (3) revising items that are not relevant to the taught material marked by the number of test takers who cannot answer a particular item [11].

IV. CONCLUSION

The analysis results of the final exam on Civic Education subject for junior high school in Balangan Regency academic year 2015/2016 showed that the test has 23 very easy items (46%), 11 easy items (22%), and 16 medium items (32%). The analysis results of the discriminator were that 21 items (42%) were categorized as poor discriminator, 16 items (32%) were categorized as fair discriminator, 11 items (22%) were categorized as good discriminator, and 2 items (4%) were categorized as very good discriminator. The results of the distractor effectiveness analysis showed that the distractors that functioned poorly were 12 items (24%), fair distractors were 16 items (32%), good distractors were 13 items (26%), and very good distractors were 9 items (18%). The results of a thorough analysis of the exam shows good quality items consisting of 12 items (24%), unfavorable items consisting of 18 items (36%), and bad items consisting of 20 items (40%).

REFERENCES

- [1] Z. Arifin, *Evaluasi Pembelajaran*, Bandung: Remaja Rosdakarya, 2013.
- [2] A. J. Nitko, and S. M. Brookhart, *Educational Assessment of Students*, 6nd Edition, Boston: Pearson Education, Inc, 2011.
- [3] S. Arikunto, *Dasar-dasar Evaluasi Pendidikan*, Jakarta: Bumi Aksara, 2013.
- [4] A. Sudijono, *Pengantar Evaluasi Pendidikan*, Jakarta: Raja Grafindo Persada, 2012.
- [5] G. Sax, *Principles of educational and psychological measurement and evaluation*, Belmont, California: Wadsworth Publishing Company, 1980.
- [6] Daryanto, *Evaluasi Pendidikan*, Jakarta: Rineka Cipta, 2012.
- [7] N. Sudjana, *Penilaian Hasil Proses Belajar Mengajar*, Bandung: Remaja Rosdakarya, 2013.
- [8] Sukiman, *Pengembangan Sistem Evaluasi*, Yogyakarta: Insan Madani, 2012.
- [9] Sugiyono, *Metode Penelitian Kuantitatif Kualitatif dan R&D*, Bandung: Alfabeta, 2012.
- [10] Mansyur, R. Harun, and Suratno, *Asesmen Pembelajaran di Sekolah: Panduan bagi Guru dan Calon Guru*, Yogyakarta: Pustaka Pelajar, 2015.
- [11] Kusaeri and Suprananto, *Pengukuran dan Penilaian Pendidikan*, Yogyakarta: Graha Ilmu, 2012.