

## THE APPLICATION-BASED ANALYSIS OF QUESTIONS ITEM QUALITY IN JUNIOR HIGH SCHOOL

**Rabatul Adawiah**

Civics Education Department Faculty of  
Teacher Training and Education  
Lambung Mangkurat University

**Address for Correspondence:**

[rabiatuladawiah@ulm.ac.id](mailto:rabiatuladawiah@ulm.ac.id)

### ABSTRACT

The activity of analyzing the items is one of the obligations for each teacher in an effort to improve the quality of the questions. However, for Civics Education teachers, this has never been done especially for questions created by the Subject Teachers' Consultation which are used for the end-semester assessment. The purpose of this study was to determine the quality of the questions based on distinguishing feature, level of difficulty, and effectiveness of the distractor. This study is an evaluation study of 50 items of Civics Education Subject test in Banjarmasin, totaling 50 questions in the form of multiple choice questions at the end-semester test, academic year 2019/2020. The data collected are in the form of: (1) final exams question sheet, (2) question answer key sheet, and (3) students' answer sheet. All data is obtained by documentation techniques. Data analysis uses the AnBuso version 8.0 application. The criteria for determining the quality of items are: (a) questions are considered good if the distinguishing feature is good/good enough, the level of difficulty is medium and all alternative answers are effective, (b) revision of alternative answers, if the distinguishing feature is good/good enough and the level of difficulty is medium, but the alternative there are ineffective answers, (c) good enough, if the distinguishing feature is good/good enough but the level of difficulty is easy/difficult, and (d) not good, if the distinguishing feature is not good. The results of this study indicate that the questions used for the end-semester test at Junior High School in Banjarmasin are 50 % of poor quality.

**Keywords:** Distinguishing Feature, Effectiveness Of Distractor, Level Of Difficulty.

### INTRODUCTION

In the era of globalization, almost all countries strive to improve the quality of education. Various efforts have been made by the Government to improve the quality of education, and among them is through improving the quality of learning and the quality of the assessment system. The quality of learning and the quality of the assessment system are two interrelated things. A good learning system will produce good quality of learning, and the quality of learning can be seen from the results of the assessment.

Assessment is defined as the activity of interpreting the measurement data according to certain criteria or rules (Widoyoko, 2014: 30; Arifin, 2013: 4). Assessment includes all the ways used in assessing individual performance (Mardapi, 2008: 5). Another opinion says that assessment is taking a decision on something by referring to certain measures such as good or bad, smart or not smart, high or low (Supardi, 2015: 11).

From some of the definitions above, it can be said that the assessment emphasizes the effort made by the teacher or student in order to obtain information in connection with the learning that has been done. The information obtained can be used as feedback for a better learning process. Assessment aims to plan and implement learning, maintain classroom atmosphere, provide feedback and appreciation, students placement, diagnose students' learning problems and assess the level of academic progress (Russell & Airasian, 2012: 5-8). Another opinion says that the assessment aims to determine the level of progress and students' development in a certain period (Popham & Eva L. Baker, 2008: 151). Whereas the assessment function is (1) diagnostic, to identify student performance, (2) formative, to help student learning, (3) summative, to review, transfer and certification, and (4) evaluative, to see about teacher and institutional performance (Weeden, Winter and Broadfoot, 2002: 19). Therefore, assessment is a very important part of learning (Russel & Airasian, 2012: 2; Mansyur, Harun Rasyid and Suratno, 2015: 22), and has a strong influence in improving the learning process (Raymond, et al., 2012: 1-6; Bers, 2008: 31-39). Based on this, teachers are required to have sufficient ability to conduct assessments, because good judgment can motivate educators to teach better and encourage students to be better in learning (Mardapi, 2012: 4).

To conduct an assessment, one of the tools commonly used in the teaching and learning process is a test. Test is a tool or procedure used to find out or measure something in a situation by means or rules that have been determined (Arikunto, 2013: 67). Another definition says that the test is a series of questions that have right or wrong answers, questions that require answers or responses to measure a person's ability level in certain aspects (Wening, 2012: 4).

As a form of learning achievement test, it is very important to maintain the quality of the questions. Assessment will produce a right information if the tool or instrument used to carry out the measurement meets several criteria such as validity, reliability and objectivity (Anderson, 2003: 10; Kubiszyn & Borich, 2013: 326). Item analysis becomes an important part in guaranteeing item validity (Nunnally & Bernstein, 1994: 304). Item analysis is an attempt to test the quality of questions to determine which items need to be maintained, discarded or revised. This analysis provides information about the quality of the questions seen from the level of difficulty, distinguishing feature and effectiveness of distractor (Muhson et al., 2015: 200). The purpose of item analysis is to identify good, bad and bad questions (Daryanto, 2012: 179; Sudjana, 2011: 135).

The activity of analyzing the items can be said as one of the 'obligations for every teacher'. It is because every teacher must be able to convey information both for the institution or for students about the extent of mastery or students' ability to the material or certain skills in connection with the material that has been given. The reality in the field often shows that the scores obtained by students from the results of the final exams or final school grades are still low. The low results are not only caused by the low ability of students to answer the questions, but can also be caused by the low quality of the questions.

## **METHODS**

This research is an evaluation research. The evaluation was carried out on the items of Civic Education Subjects at the end of the odd semester assessment at SMPN Banjarmasin City Academic Year 2019/2020. The problem analyzed is the 8<sup>th</sup> grade question test, totaling 50 multiple choice forms with research subjects totaling 200 students from nine schools. The data collected are in the form of: (1) final exam question sheets, (2) question answer key sheets, and (3) student answer sheets obtained by documentation techniques.

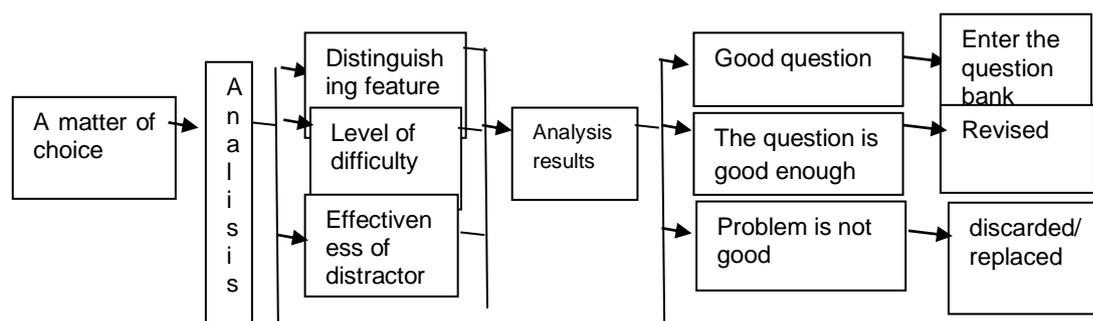
In analyzing the data, researcher used the AnBuso version 8.0 program. The problem is said to be valid if it has a minimum correlation of 0.2. While referring to the distinguishing feature criterion items, the item is good, if the coefficient of distinguishing feature exceed of 0.3. While the coefficient between 0.2-0.3 is considered good enough, and the coefficient below 0.2 is considered poor.

The level of difficulty criteria range from  $<0.3$  which included on the difficult category,  $0.3-0.7$  which included on the moderate category and  $>0.7$  which included on the easy category. The criteria for a good level of difficulty are between  $0.3 - 0.7$ .

The criteria for distractor alternative answers of good items are if the alternative is answered by at least 5% of test takers, so that the alternative is considered effective. The criteria used to interpret the effectiveness of the distractor items are as follows: (a) If all the distractor on the item functions, it means the question is said to be very good which can then be stored in the question bank, (b) if there is one non-functional distractor on the item, then the problem is said to be good and can be stored in the question bank with the condition that the non-functional option is revised, (c) if there are two distractors in the non-functional item, then the problem is said to be bad and cannot be stored in the question bank. The question must be revised until it meets the criteria of good questions, and (d) if there are three or more non-functional distractors in the items, then the problem is said to be very bad and cannot be stored in the question bank. The question must be revised until it meets the criteria of a good question or the problem is discarded and replaced with a new one.

While the criteria for determining the quality of items are: (a) questions are considered good if the distinguishing feature is good/good enough, the level of difficulty is moderate and all alternative answers are effective, (b) revision of alternative answers, if the distinguishing feature is good/good enough and the level of difficulty is intermediate, but there are ineffective alternative answers, (c) good enough, if the distinguishing feature is good/good enough, but the level of difficulty is easy/difficult, and (d) is not good, if the distinguishing feature is not good (Muhson, 2015).

Evaluation of the quality of the questions, designed as follows.



## RESULTS AND DISCUSSION

The results of the analysis of the distinguishing feature index of Civics Education issues at Junior High School in Banjarmasin can be seen in the following table:

**Table 1.** Percentage of Distinguishing Feature of Problems

Category	Item Number	Number of Items	%
<b>Good <math>&gt;0.3</math></b>	1, 2, 5, 7, 8, 9, 10, 13, 18, 19, 21, 28, 29, 34, 42, 47	16	32
<b>Good Enough</b>	6, 12, 14, 24, 32, 35, 40, 48, 49	9	18

<b>0.2 - 0.3</b>			
<b>Not Good &lt;0.2</b>	3, 4, 11, 15, 16, 17, 20, 22, 23, 25, 26, 27, 30, 31, 33, 36, 37, 38, 39, 41, 43, 44, 45, 46, 50	25	50

From the table above, it is known that the questions that have a distinguishing feature are categorized as good only 32%, and 50% of the distinguishing feature of questions categorized as not good. The follow-up of the items after analyzing the distinguishing feature is as follows: (1) items that have good differentiation are stored in the question bank. These items can be re-issued when the next learning achievement test, (2) items with low differentiation items have two possibilities of not continuing, namely: (a) traced for later revised and then re-issued in future learning achievement tests to determine the distinguishing features increased or not. (b) discarded/replaced, and (3) the items with distinguishing index number is negative, should be discarded because of their low quality (Sudijono, 2012).

Next analysis is about the difficulty level of the questions. The difficulty level of an item is one of the item parameters that is very useful in the analysis because it can provide information about the difficulty of the item. Questions are said to be difficult if the difficulty level is close to 0 and questions are considered easy if it is approaching 1 (Muhson et al., 2015). The results of the analysis of the difficulty level index at the end-semester test can be seen in the following table:

**Table 2. Percentage of Problem Difficulty Levels**

Category	Item Number	Number of Items	%
<b>Easy &gt;0.7</b>	8, 18, 24, 47, 49, 50	6	12
<b>Medium 0.3-0.7</b>	1, 2, 5, 6, 7, 9, 10, 12, 14, 16, 19, 21, 22, 26, 27, 28, 29, 30, 32, 34, 35, 36, 37, 39, 42, 44, 48	27	54
<b>Difficult &lt;0.3</b>	3, 4, 11, 13, 15, 17, 20, 23, 25, 31, 33, 38, 40, 41, 43, 45, 46	17	34

The table illustrates that only 12% of questions is easy, 54% is medium, and 34% is difficult category. For the effectiveness of distractor, the results of this study indicate that all questions have good distractor, meaning that the distractor in each item has been chosen by more than 5% of students. From the analysis of the distinguishing feature, difficulty level and effectiveness of distractor, the quality distribution of questions can be seen in the following table:

**Table 3. Percentage of the Quality of the Questions**

Category	Item Number	Number of Items	%
<b>Good</b>	1, 2, 5, 6, 7, 9, 10, 12, 14, 19, 21, 28, 29, 32, 34, 35, 42, 48	18	36
<b>Revised Alternative Answers</b>	-	0	0
<b>Good Enough</b>	8, 13, 18, 24, 40, 47, 49	7	14
<b>Not good</b>	3, 4, 11, 15, 16, 17, 20, 22, 23, 25, 26, 27, 30, 31, 33, 36, 37, 38, 39, 41, 43, 44, 45, 45, 50	25	50

From the table above, it can be concluded that the quality of the end-semester Civics Education test items at 8<sup>th</sup> grade of Junior High School in Banjarmasin, academic year 2019/2020 is 50% poor quality.

## DISCUSSION

The results of this study indicate that 50% of end-semester Civics Education test items used at State Junior High School in Banjarmasin, academic year 2019/2020 is poor. The results of the analysis of the distinguishing feature are only 32% were categorized as good, 18% were categorized as moderate, and 50% were categorized as poor. The item of distinguishing feature is the ability of items to distinguish between high and low ability test takers (Muhson et al., 2015: 200; Azwar, 2010: 137; Mansyur, Harun Rasyid and Suratno, 2015: 189; Sukiman, 2012: 215). Thus, it can be said that of the 50 items used, 50% of the questions cannot distinguish students who have high abilities and students who have low abilities. Knowing the distinguishing feature items is important because one of the basics in preparing test items for learning outcomes is the assumption that the ability of the test items is the assumption that the ability of one question is different from another and the test items must be able to provide the test results that illustrate the differences in ability among the test takers. (Sudijono, 2012: 386).

The analysis shows that item number 3, 4, 11, 15, 16, 17, 20, 22, 23, 25, 26, 27, 30, 31, 33, 36, 37, 38, 39, 41, 43, 44, 45, 46, and 50 have poor distinguishing feature with distinguishing feature coefficients  $<0.2$ . It means that the item cannot distinguish between students who can answer (high ability) and students who cannot answer (low ability). Therefore, the item should not be used anymore or must be discarded. The examples of problems that have distinguishing feature can be seen in the following figure

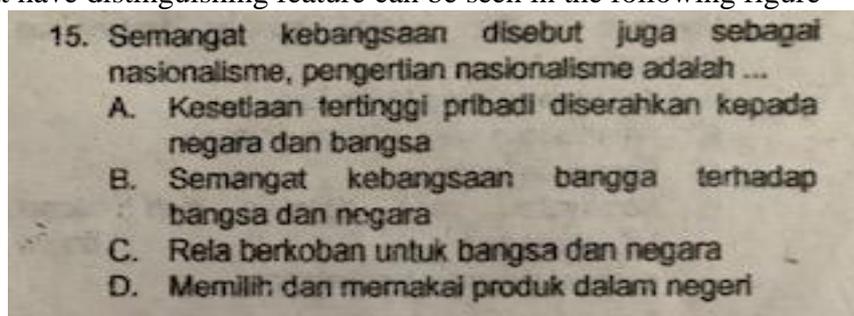


Fig. 1. Examples of problems with poor distinguishing feature.

While for item number 6,12,22,23,25,30,35,38, 48 and 49 is good enough/moderate differences with coefficients between 0.2 - 0.3, meaning that these items can slightly differentiate ability of each student and the item can still be maintained although it is not yet satisfying and needs to be improved. The examples of problem can be seen in the following picture.

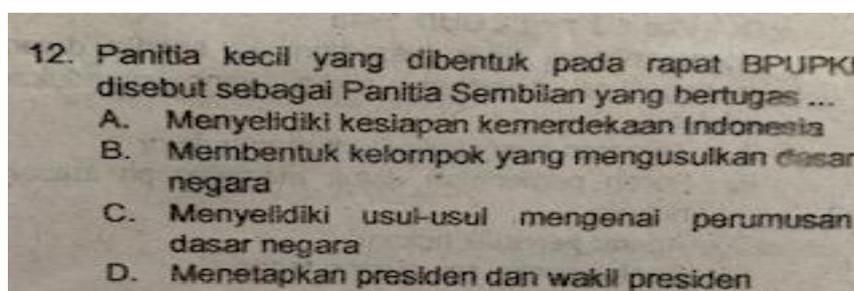
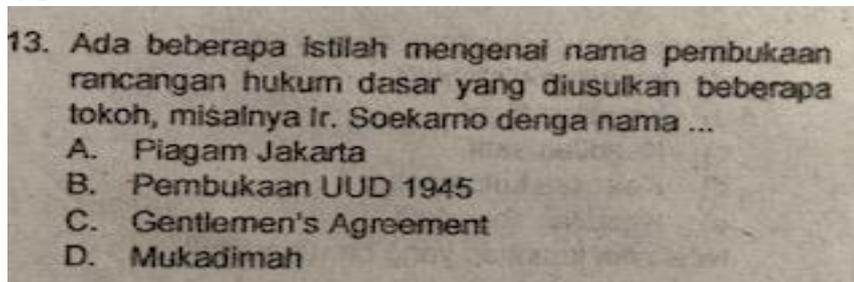


Fig. 2. Examples of problems with moderate distinguishing feature.

Whereas for item number 1, 2, 5, 7, 8, 9, 10, 13, 18, 19, 21, 28, 29, 34, 42, and 47 have good/high difference with coefficients  $>0.3$ . The examples of problem can be seen in the following picture.

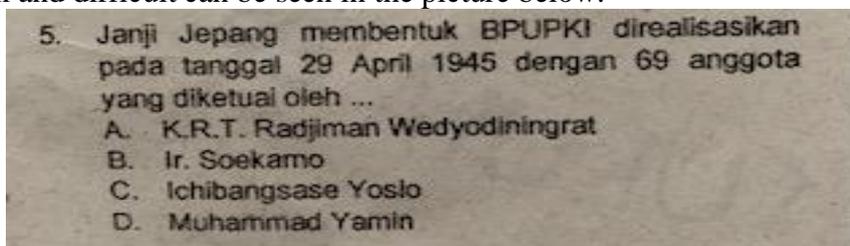


**Fig. 3. Examples of problems with good distinguishing feature.**

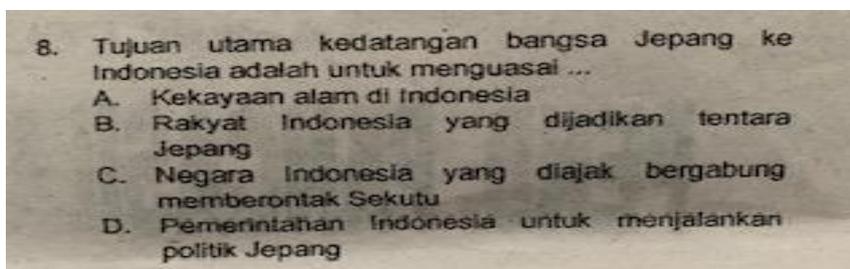
Based on these data, it can be concluded that 50 questions that Civics Education used in the end-semester test are only 50% were able to distinguish students who were classified as high and low. In other words, there are only 50% of the questions if it given to students who are able, the results will show high achievement, and if given to students who are weak the results are low.

To find out whether an item is good or bad, it can be seen from how much the ability of these questions in distinguishing the intelligent and less intelligent students (Shermis & Vesta, 2010: 282). Some questions, namely questions number 9,10,11, 17,19,34,40,46,47 and 49 show negative differences. Negative index of distinguishing feature is bad, because it indicates that the item cannot distinguish according to the stated purpose (Chatterji, 2003: 385-386).

While, the difficulty level of this study showed that 12% of the categories are easy, 54% are medium and 34% are difficult. The examples of problems that are categorized as easy, medium and difficult can be seen in the picture below.



**Fig. 4. Example of easy category question**



**Fig. 5. Example of medium category question**

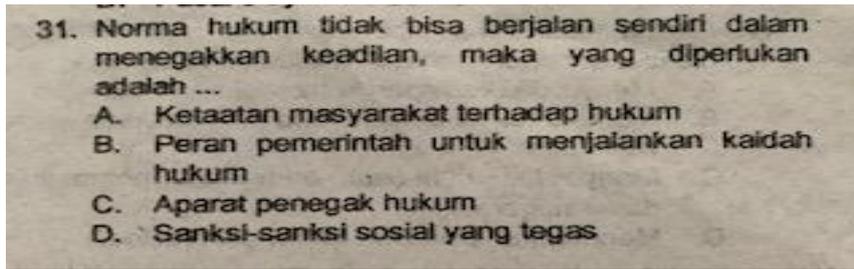


Fig. 6. Example of difficult category question

The items is categorized as good if the item is not too difficult and not too easy (Anas, 2011: 370). The questions that are too easy do not stimulate students to try to solve them, and the questions that are too difficult cause students to become discouraged and have no enthusiasm in trying because it is beyond of their range (Arikunto, 2013: 222). The assumption used to obtain good quality questions, especially for the difficulty level of the questions has existence of balance, in addition to meet the validity and reliability. The balance of questions is the existence of questions that include proportionally easy, medium and difficult (Mansyur, Harun Rasyid and Suratno, 2015: 180). Therefore it is expected that in a test there is a proportional level of difficulty between questions with easy, medium, and difficult categories (Rasyid & Mansyur, 2008: 239).

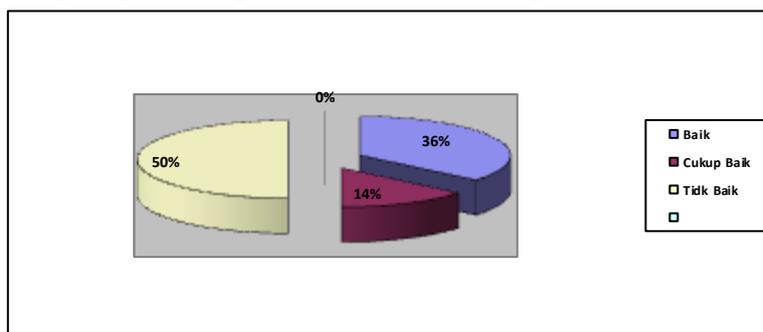
The results of the question analysis, for items number 8, 18, 24, 47, 49, and 50, are in easy category with a coefficient value  $>0.7$ . Item number 1, 2, 5, 6, 7, 9, 10, 12, 14, 16, 19, 21, 22, 26, 27, 28, 29, 30, 32, 34, 35, 36, 37, 39, 42, 44, and 48 are in intermediate category with coefficient values between 0.3 - 0.7, and item number 3, 4, 11, 13, 15, 17, 20, 23, 25, 31, 33, 38, 40, 41, 43, 45, and 46, are in difficult category with coefficient values  $<0.3$ . The data shows that the number of easy, intermediate, and difficult questions do not meet the proportional requirements because the number of easy questions is only 12% (6 questions), questions that have medium category is 54% (27 questions) and questions that have difficult category is 34% (17 questions). Therefore to meet the provisions of the proportional difficulty level, some considerations in determining the proportion of the number of easy, medium and difficult questions are: (1) there is a balance of the number of questions for the three categories, (2) the proportion of the number of questions for the three categories is based on the normal curve, in which most of the questions are at medium category, then the proportion of easy and difficult items is proportional (Mansyur, Harun Rasyid and Suratno, 2015: 181).

From the 50 items analyzed, item number 25 was the most difficult one with a difficulty index of 0.020. Difficulty of items can be low (low coefficient value) if the contents of the learning target presented are not good, for example, the writing format is unclear, the language of writing items is confusing/ambiguous, and students do not yet understand the competencies/contents conveyed in the items (Chatterji, 2003: 385). Items that are too difficult show failure in constructing items (Shermis & Vesta 2011: 299). For the effectiveness of distractor, the results of the analysis show that all items have good distractor. In other words, all distractor have been chosen by more than 5% of the test takers.

The effectiveness of distractor question items is how the ability of distractor questions to distract students who are less able to choose the alternative answers. Making questions in multiple choice forms must have the effectiveness of distractor means that the answer should not be too easy for students, but the answer can show the real ability related to who has knowledge, lacks knowledge or is confused with the material presented (Chatterji, 2003: 386). A distractor can be said to function well, if the distractor has a great traction for test

takers who do not understand the concept or do not master the material (Arikunto, 2015: 234). A distractor can be said to function well if at least selected by 5% of test takers or more are chosen by the lower classes (Daryanto, 2010: 193; Sudijono, 2005: 411). Another opinion said that good question items are the distractor which is chosen equally by the test takers. Otherwise, bad question items are the distractor which is chosen unequally (Arifin, 2013: 279).

Based on the results of the items analysis by AnBuso program version 8.0 on the distinguishing feature of questions, the difficulty level and the effectiveness of distractor, it can be seen that the quality of the questions are shown in picture below:



**Fig.7. The Percentage of Questions Quality**

From the picture, question items categorized as good are 36%. Item number 1, 2, 5, 6, 7, 9, 10, 12, 14, 19, 21, 28, 29, 32, 34, 35, 42, and 48, are good. Item number 8, 13, 18, 24, 40, 47, and 49 are good enough. The rest are item number 3, 4, 11, 15, 16, 17, 20, 22, 23, 25, 26, 27, 30, 31, 33, 36, 37, 38, 39, 41, 43, 44, 45, 45, 50 are poor. The questions that are poor quality, of course, they must be discarded or replaced with new questions.

## CONCLUSION

The results of this study indicate that the distinguishing feature item index is 32% in the good category, 18% in the moderate category, and 50% in the poor category. For the level of difficulty, 12% is categorized as easy, 54% is categorized as medium and 34% is categorized as difficult. For the effectiveness of distractor, the results of the analysis show that all items have good distractor. In other words, all distractors have been chosen by more than 5% of the test takers. From the results of the analysis of the distinguishing feature, the level of difficulty, and the effectiveness of the distractor, it can be seen that the quality of the items: 36% is good, 14% is good enough, and 50% is poor.

## REFERENCES

- Anderson, L.W. 2003. Classroom Assessment: Enhancing The Quality of Teacher Decision Making. New Jersey: Lawrence Erlbaum Associates, Inc.
- Arifin, Zainal, 2013. Evaluasi Pembelajaran. Bandung: Remaja Rosdakarya.
- Arikunto, Suharsimi. 2013. Dasar-Dasar Evaluasi Pendidikan (Edisi 2). Jakarta: Bumi Aksara.
- Azwar, Syaifuddin. 2003. Tes Prestasi. Yogyakarta: Pustaka Pelajar.
- Bers, T.H. 2008. "The Role of Institutional Assessment in Assessing Student Learning Outcomes". *New Directions for Higher Education*, 141, 31-39.
- Chatterji, Madhabi. 2003. Designing and Using Tools For Educational Assessment. USA: Pearson Education, Inc.
- Daryanto. 2012. Evaluasi Pendidikan. Jakarta: Rineka Cipta.

- Kubiszyn, T. & Borich, G.D. 2013. *Educational Testing and Measurement: Classroom Application and Practice* (10<sup>th</sup> ed.). Hoboken, NJ: John Wiley & Sons, Inc.
- Mardapi, Djemari. 2012. *Pengukuran, Penilaian dan Evaluasi Pendidikan*. Yogyakarta: Nuha Medika.
- \_\_\_\_\_. 2008. *Teknik Penyusunan Instrumen Tes dan Nontes*. Yogyakarta: Mitra Cendikia Press.
- Muhson, Ali, Berkah Lestari, Supriyanto & Kiromin Baroroh. 2015. "Kelayakan AnBuso sebagai Software Analisis Butir Soal Bagi Guru". *Jurnal Kependidikan*. Fakultas Ekonomi Universitas Negeri Yogyakarta. Vol. 45. No. 2, Hlm.198-210.
- Muhson, Ali. 2015. *Panduan Penggunaan AnBuso*. Yogyakarta: Universitas Negeri Yogyakarta.
- Nunnally, J. C & Bernstein, I. H, 1994. *Psychometric Theory* (Third Edition). New York: McGraw-Hill, Inc.
- Pophan, W. James. 2008 dan Eva L, Baker. 2008. *Teknik Mengajar Secara Sistematis*. (Terjemahan Amirul Hadi, dkk.) Jakarta: Rineka Cipta.
- Purnomo, A., 2007. *Kemampuan Guru dalam Merancang Tes Berbentuk Pilihan Ganda pada Mata Pelajaran IPS untuk Ujian Akhir Semester (UAS)*. *Lembaran Ilmu Kependidikan*, Jilid 36, No. 1, Juni 2007.
- Rasyid, Harun dan Mansyur. 2008. *Penilaian Hasil Belajar*. Bandung: CV. Wacana Prima.
- Raymond, J.E., Homer,C.S.E., Smith, R., & Gray, J.E. 2012. "Lerning through Authentic Assessment: An Evaluation of A New Development in The Undergraduate Midwifery Curriculum". *Nurse Education in Practice*, 30, 1-6.
- Russell, M.K., & Airasian, P.W. 2012. *Classroom Assessment: Concepts and Applications*. (7<sup>th</sup>ed.). New York: McGraw-Hill.
- Shermis, Mark D. &Di Vesta Francis J. 2011. *Classroom Assessment In Action*. USA: Rowman & Littlefield Publisher, Inc.
- Sudijono, Anas. 2012. *Pengantar Evaluasi Pendidikan*. Jakarta: Raja Grafindo: Persada.
- \_\_\_\_\_. 2011. *Pengantar Evaluasi Pendidikan*. Jakarta: Raja Grafindo: Persada.
- Sudjana, Nana. 2011. *Penilaian Hasil Proses Belajar Mengajar*. Bandung: PT. Remaja Rosdakarya.
- Supardi. 2015. *Penilaian Autentik: Pembelajaran Afektif, Kognitif, dan Psikomotor (Konsep dan Aplikasi)*. Jakarta: PT. RajaGrafindo Persada.
- Sukiman. 2012. *Pengembangan Sistem Evaluasi*. Yogyakarta: Insan Madani.
- Weeden, P., Winter, J., & Broadfoot, P. 2002. *Assessment: What's in it for school?*. London and new York: Routledge Falmar.
- Wening, Sri. 2012. *Materi Evaluasi Pembelajaran*. Yogyakarta: Fakultas Teknik Universitas Negeri Yogyakarta.
- Widoyoko, E.P. 2014. *Evaluasi Program Pembelajaran Panduan Praktis bagi Pendidik dan Calon Pendidik*. Yogyakarta: Pustaka Pelajar.