QUALITY OF END YEAR ASSESSMENT QUESTION ITEMS FOR ISLAMIC RELIGIOUS EDUCATION & CHARACTER SUBJECTS

Rahayu Ningsih[1], Karnadi[2]
[12]*Universitas Islam Negeri Walisongo Indonesia;*
Email: rahayu_1903016032@student.walisongo.ac.id

Abstract: This research aims to determine the quality of the End of Year Assessment (PAT) questions in Islamic Religious Education and Character Education subjects in class IX in Junior High School, Semarang City, for the 2022/2023 Academic Year in terms of level of difficulty, distinguishing power, effectiveness of distractors and reliability. This research is quantitative descriptive research, and the research subjects are class IX students for the 2022/2023 school year at SMP N 16 and SMP Islam Al-Azhar 29 Semarang, with a total of 90 students from each school. The data collection techniques used were observation, documentation, and testing techniques, and then the data obtained was analyzed using Anates Software version 4. The results of this research showed that the results of the analysis of the quality of the End of Year Assessment (PAT) questions for PAI and BP class IX subjects obtained: At SMP Negeri 16 Semarang, 20 questions are valid questions and can be stored in the school's question bank to be used in other tests relevant to PAI and BP subjects, while 20 questions are questions that are not valid and needs to be improved through improvements so that it can be used in subsequent tests. At Al-Azhar 29 Islamic Middle School Semarang, 15 questions are good questions and can be stored in the school's question bank so they can be used in other tests that are relevant to PAI and BP subjects, while 15 questions are good questions are invalid and need to be improved through revisions so that the items can be used in subsequent tests.
Keywords: Quality of Question Items, Islamic Religious Education and Character, SMP Negeri 16 Semarang, SMP Al-Azhar 29 Semarang

INTRODUCTION

Haris & Asep (2013) assume that learning outcomes are abilities possessed by students that are produced when they have completed the evaluation process. Several learning outcomes can be measured through an evaluation process (Agustiana et al., 2018). Apart from that, the evaluation also functions as a benchmark for students, teachers, parents, and even the community in teaching and learning activities (Mahirah 2017). The function of evaluation for students is that they will know the extent of their learning progress by looking at the graph of their learning results, whether they are increasing or decreasing. For a parent, it will also be easy for them to see the transparent development of the extent to which their child has good qualities or not. Meanwhile, the school (the principal, teachers, and staff) will be able to see and analyze more easily how students in the school have made significant progress up or down. Then, the community will know whether the school has good quality or not.

Evaluation can also be said to be a component of learning. Sanjaya (2015) explains that learning components include objectives, subject matter, learning methods or strategies, and evaluation. As a learning component, evaluation has an important role in determining the success of learning. As stated by Sudijono (2012), evaluation functions to determine the extent to which the efforts in the learning process carried out by the teacher have brought results. Thus, evaluation becomes a very important activity in the learning system.

In evaluating student learning outcomes, there are two techniques that are usually used in this case, namely test and non-test techniques. However, in the academic world, the test technique is often used. The tests usually use measuring instruments in the form of grades to measure student learning outcomes, mainly cognitive learning outcomes in mastery of teaching materials, which are then commensurate with the goals of the world of teaching and learning (Sudjana 2016). Then, according to Arifin (2017), testing is a technique that can be used to determine a student's learning outcomes. It is expressed through questions, statements, or a series of tasks that must be carried out by students so that they can be measured and find out the extent of an individual's ability to solve a problem or problem. Several studies show that a test is categorized as a quality test when the aspects of validity, reliability, discriminating power, level of difficulty, and distractor effectiveness are included in it (Lubis, 2008; Sudijono, 2015; Queenta & Yuliasma, 2020; Solichin, 2017). To find out whether these aspects need to be included in the test in order to get information about whether the test is worthy of being considered a good evaluation or not in an evaluation process.

The results of the question analysis will determine which questions are appropriate to be placed in the question bank, revised, and discarded. The activity of analyzing question items is an activity that must be carried out to improve the quality of the questions that have been written. This activity is a process of collecting, summarizing, and using information from students' answers to make decisions about each assessment (Nikto 1996). Question item analysis is an activity that must be carried out by teachers, which functions to improve the quality of the questions that have been created (Lestari et al., 2019). It is better to carry out question item analysis activities before the questions are distributed so that the teacher can find out whether the question items that have been prepared can carry out their function well or vice versa. In reality, there are many teachers who still do not carry out question item analysis activities because they feel burdened in carrying out the question item analysis process before the questions are tested on their students (Susanto et al., 2015).

Research conducted by Sarea (2022) found that the difficulty level of the questions at SMPN 17 Makassar and SMPN 19 Makassar was categorized as good for use. The item discrimination index of SMPN 17 Makassar was good at 92.5% and unfavorable at 7.5%. Meanwhile, SMPN 19 Makassar was categorized as good at 100%. The lowest SEM size for SMPN 17 Makassar UAS questions is 0.000173681, and the highest is 26.44767997, while the lowest SEM size for SMPN 19 Makassar UAS questions is 0.65658735 and the highest is 7.903108528. Then, research conducted by Muzayyanah (2020) showed that at SMP N 18 Semarang, there were 55% valid questions, 45% invalid, low reliability, easy difficulty level, poor differentiating power, and poor distractor function, whereas in SMP N 44 Semarang has 50% valid items, 50% invalid items, low reliability, easy difficulty level, poor discriminating power, and poor distractor function. Then, research conducted by Muluki et al. (2020) showed that the odd semester test items for the Natural Sciences class IV MI Radhiatul Adawiyah Makassar academic year 2018/2019 had a high-reliability coefficient of 0.70, the quality of the questions was based on differentiating power, questions that have poor discriminating power of 30%, fair discriminating power of 35%, good discriminating power of 25%, and negative discriminating power of 10%.

The results of research from Muslim (2017) show that the level of difficulty of the multiple choice questions for the Even Semester Final Examination test in the class VII Islamic Religious Education (PAI) subject made by the Semarang City PAI Subject Teacher Deliberation (MGMP) team for the 2016/2017 school year has 90% of the items are valid, 10% invalid, reliability is less than 0.80 (not reliable), the level of difficulty tends to be easy, namely 58%, the discrimination power tends to be in a bad category at 74%, the effectiveness of distractors tends to be in the very good category at 30%. Furthermore, research from Rahmawati (2015) shows that the Final Semester Exam Questions for the subjects of Islamic Religious Education and Character Education for Class VII for the 2014-2015 academic year are superior to PIRI 2 Middle School compared to Muhammadiyah Middle School because of the comparative validity of the test, the level of difficulty of the questions, and the strength. Differentiating questions and distracting functions. Meanwhile, Muhammadiyah Middle School 4 is ahead in terms of test validity and test reliability.

Based on observations made by researchers and several other relevant studies, teacher analysis of the quality of questions is quite rarely carried out due to several factors that influence it. An example is according to research from Rohmah (2017). The results of this research show that a public elementary school in Mojoroto District, Kediri City, found that the assessment of student learning outcomes carried out by a teacher was not fully able to describe the level of competency achievement of the students. This is indicated because teachers do not really know the indicators that students have not achieved. Then, research conducted by Septiana (2016) related to findings in the form of UAS questions in Biology subjects that were made by teachers but had never been analyzed in terms of both quality and quantity, including validity, reliability, distinguishing power, level of difficulty, and effectiveness of distractors.

This also coincided with what happened when researchers conducted observations at SMP Negeri 16 Semarang and SMP Islam Al Azhar 29 Semarang, where teachers had never carried out quality analysis activities because they were assumed to be quite complicated activities and took up quite a lot of time, especially for End of Year Assessment questions. It was found in preliminary observations that the questions had been made based on standards set by the government and made by teachers who had been equipped with certification, training, participation in organizations such as the Islamic Education Subject Teachers' Conference (MGMP), and of course, contained achievements such as The intended learning is according to the subjects taught by the teacher, especially in the subjects of Islamic Religious Education and Character (Researcher Observations 2023).

However, some of the problems that are clearly visible are that there are students who take things too lightly because the subject is already familiar with their environment, they don't need to study it, and in the end, there is a feeling of already knowing the answer. After proofreading, it turned out that there were parts that made the students' answers inaccurate. This is, of course, an evaluation for teachers who teach Islamic Religious Education and Characteristic subjects to fix the difficulties found and especially to improve morals or manners, which are the main milestones in students' daily lives both individually, in groups, and in their environment (Researcher Observations 2023).

Regarding students' understanding of the questions in the End of Year Assessment at each school, it was found that there were some students who thought it was very easy, and there were also those who thought that the questions could not be solved optimally because of unfamiliar Islamic terms or rarely heard by students, especially when lessons contain the history of the prophet or about Islamic dates. Another thing that sometimes becomes a difficulty is because there are some students who are not yet fluent in reading the Koran, so if one day the teacher wants students to come to the front of the class one by one to recite one or several verses in a material by memorizing, there is who have not been able to finalize this. Of course, it has an impact when a question item has to include several verses, and there are students who are still confused, can't read it, and can't even understand what the verse means (Researcher Observation 2023).

In the opinion of Wati (2013) stated that the reality in the field so far is that analysis of question items is rarely carried out because, in terms of material, question construction, language, validity, reliability, and analysis of question items, it is often included in the low category and even the quality is not known for certain.

In fact, analyzing question items is a recommended activity so that teachers can improve the quality of the questions they create (Sumiati et al., 2018). This activity is a process carried out by collecting, summarizing, and using relevant references from students' answers to reach a decision about each assessment. This analysis aims to identify questions that fall into the categories of good, poor, fair, bad, or very bad. After stating the results of the analysis, the step that a teacher can take is to make improvements to the question so that it becomes a higher quality question (Hasibuan 2013). The aim of this research is to determine the quality of the End of Year Assessment (PAT) questions in Islamic Religious Education and Character Education subjects in class IX in Junior High School, Semarang City, for the 2022/2023 Academic Year in terms of level of difficulty, distinguishing power, effectiveness of distractors and reliability.

METHOD

This type of research is quantitative with a descriptive approach. In this case, the researcher explained "The Quality of End-of-Year Assessment Questions for Religious and Character Education Subjects in Semarang City Middle Schools. The research data is in the form of data regarding the analysis of End of Year Assessment (PAT) questions on Islamic Religious Education and Character Education subjects in class IX from two schools, namely SMP Negeri 16 Semarang and SMP Islam Al-Azhar 29 Semarang, with a research population consisting of 90 students from SMP Islam Al-Azhar Semarang.

Data collection techniques were carried out by means of observation, interviews, and documentation. In terms of collecting this data, the researcher conducted direct interviews with teachers of Islamic Religious Education and Character Education subjects, while the researcher's documentation was in the form of a set of end-of-year assessment questions for class IX for the 2022/2023 academic year in the form of End-of-Year Assessment question sheets, question grids, students' names, answer keys, and students' answers. Then, data analysis using quantitative analysis techniques was carried out using the Anates. V.4 program.

RESULTS AND DISCUSSION

Results

The end-of-year assessment for the Islamic Religious Education and Character Education (PAI & BP) class IX at SMP N 16 Semarang consists of 40 multiple-choice questions, while Islamic Middle School Al Azhar 29 Semarang has 30 multiple-choice questions. Each question item has alternative answers A, B, C, and D. The results of the analysis of the question items are reviewed from the following aspects:

Validity

Validity can be seen in the correlation of item scores with the total score, which is divided into two categories, namely significant (valid questions) and the sign (-), which means not significant (invalid questions). In the Anates Software, the r-table value is immediately known in the correlation section of the item score with the total score.

Table 1. Analysis of the Validity of the End-of-Year Assessment Multiple Choice Test

PAI & BP subjects Class IX SMP Negeri 16 Semarang

No	Interpretation of Validity	Number of Questions	Question Number
1.	Valid	20	3, 6, 8, 10, 11, 13, 18, 19, 23, 24, 25, 26, 27, 30, 34, 35, 36, 37, 39, 40
2.	Invalid	20	1, 2, 4, 5, 7, 9, 12, 14, 15, 16, 17, 20, 21, 22, 28, 29, 31, 32, 33, 38

Table 2. Validity Analysis of End-of-Year Multiple Choice Tests

PAI & BP Subjects Class IX Islamic Middle School Al Azhar 29 Semarang

No	Interpretation of Validity	Number of Questions	Question Number
1.	Valid	15	3, 5, 9, 12, 13, 15, 17, 18, 19, 21, 23, 24, 25, 27, 29
2.	Invalid	15	1, 2, 4, 6, 7, 8, 10, 11, 14, 16, 20, 22, 26, 28, 30

Based on Table 1 above, it can be concluded that in the PAI and BP class IX questions at SMP Negeri 16 Semarang, there are 20 (50%) valid questions and 20 (50%) invalid questions. Meanwhile, in Table 2, it can be concluded that in the questions on the PAI and BP subjects for class IX Islamic Middle School Al-Azhar 29 Semarang, there are 15 (50%) valid question items and 15 (50%) invalid question items.

Reliability

Test reliability refers to the consistency of measurement results shown by the test. The reliability of the multiple choice test is calculated using the formula $\llbracket KR \rrbracket\_20$. Reliability is calculated only for valid questions.

Table 3. Reliability Analysis of End-of-Year Assessment Multiple Choice Questions

PAI & BP Subjects Class IX SMP N 16 Semarang

Number of Items	KR-20	Category
20	0,88	Very high

Table 4. Reliability Analysis of End-of-Year Assessment Multiple Choice Questions
PAI & BP Subjects Class IX Islamic Middle School Al Azhar 29 Semarang

Number of Items	KR-20	Category
15	0,44	Medium

Based on Table 3 above, it can be concluded that the questions on the PAI and BP class IX subjects at SMP Negeri 16 Semarang have a KR-20 value of 0.88, which means they have reliability in the very high category. Meanwhile, in Table 4, it can be concluded that the questions on PAI and BP class IX Islamic Middle School Al-Azhar 29 Semarang have a KR-20 value of 0.44, which means they have medium category reliability.

Difficulty Level

The level of difficulty is the opportunity to answer a question correctly, accompanied by the level of ability of an individual in the form of an index. It can also be defined as the proportion of students who answer the test correctly (Yusiana and Zamsir 2018).

Table 5. Analysis of the Difficulty Level of End-of-Year Assessment Multiple-Choice
Questions PAI & BP subjects Class IX SMP N 16 Semarang

No	Difficulty Level	Item No	Total	Percentage
1.	0,71 – 1,00 (Easy)	1, 2, 4, 6, 7, 9, 13, 15, 16, 17, 21, 22, 24, 31, 32, 38	16	40%
2.	0,31 – 0,70 (Medium)	3, 5, 8, 10, 11, 14, 18, 20, 23, 25, 26, 27, 28, 29, 30, 33, 34, 35, 36, 37, 39, 40	22	55%
3.	0,00 – 0,30 (Difficult)	12, 19	2	5%

Table 6. Analysis of the Difficulty Level of Multiple Choice Questions at the End of Year
Assessment PAI & BP Subjects Class IX Islamic Middle School Al Azhar 29 Semarang

No	Difficulty Level	Item No	Total	Percentage
1.	0,71 – 1,00 (Easy)	1, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 23, 24, 26, 27, 29.	25	83,33%
2.	0,31 – 0,70 (Medium)	2, 22, 25, 28, 30.	5	17%
3.	0,00 – 0,30 (Difficult)	-	0	0%

Based on Table 5 above, it can be concluded that the questions for PAI and BP class IX at SMP Negeri 16 Semarang have 16 questions, or 40% of which are in the easy category, 22 questions, or 55% of which are in the medium category and two questions or 5% is in the difficult category. Meanwhile, in Table 6, it can be concluded that the questions on PAI and BP class IX Islamic Middle School Al-Azhar 29 Semarang have 25 questions, or 83.33%, which are included in the easy category, five questions, or 17%, which are included in the medium category, and 0 questions or 0% which are in the difficult category.

Differentiating Power

Differentiating power is the ability of the questions and their inherent scores to distinguish participants who take the test from the high group or the low group (Hanifah 2014).

Table 7. Analysis of the Differentiating Power of End-of-Year Assessment Multiple Choice Questions PAI & BP Subjects Class IX SMP N 16 Semarang

No	Differentiating Power	Item Number	Total	Percentage
1.	0,00 – 0,20 (Low)	1, 12, 21, 32	4	10%
2.	0,21 – 0,40 (Medium)	2, 3, 4, 5, 6, 7, 9, 13, 14, 15, 16, 17, 22, 28, 29, 30, 31, 33, 38	19	47,5%
3.	0,41 – 0,70 (Good)	8, 10, 11, 18, 19, 20, 23, 24, 25, 26, 27, 34, 35, 36, 37, 39, 40	17	42,5%
4.	0,71 – 1,00 (Very Good)	-	0	0%
5.	Negative (Very Low)	-	0	0%

Table 8. Analysis of the Differentiating Power of End-of-Year Assessment Multiple Choice Questions PAI & BP Subjects Class IX Islamic Middle School Al Azhar 29 Semarang

No	Differentiating Power	Items Number	Total	Percentage
1.	0,00 – 0,20 (Low)	2, 5, 11, 12, 13, 14, 16, 19, 20, 23	10	33,33%
2.	0,21 – 0,40 (Medium)	1, 3, 4, 6, 7, 8, 9, 10, 17, 18, 21, 24, 26, 28, 29	15	50%
3.	0,41 – 0,70 (Good)	15, 22, 25, 27, 30	5	17%
4.	0,71 – 1,00 (Very Good)	-	0	0%
5.	Negative (Very Low)	-	0	0%

Based on Table 7 above, it can be concluded that the End of Year Assessment (PAT) questions for class IX PAI and BP subjects at SMP Negeri 16 Semarang have sufficient differentiating power, namely 47.5%. Meanwhile, in Table 8, it can be concluded that the End of Year Assessment (PAT) questions for PAI and BP class IX Islamic Middle School Al-Azhar 29 Semarang have sufficient differentiating power, namely 50%.

Distractor Effectiveness

The effectiveness of distractors can be obtained by calculating several answers from students who choose a, b, c, d, e, or do not choose any answer at all. Based on this, the effectiveness of the distractor can be stated as whether it is functioning well or not. A distractor can be declared good if the alternative chosen has a score of at least 5% of all test takers (Muluki et al., 2020).

Table 9. Analysis of the Effectiveness of Distracters on Multiple Choice End of Year Assessment Questions PAI & BP Subjects Class IX SMP N 16 Semarang

No.	Distractor Effectiveness	Distractor Effectiveness	Total	Percentage
1.	Very Good	2, 3, 4, 5, 6, 8, 9, 10, 11, 12, 13, 14, 16, 17, 18, 19, 20, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 33, 34, 35, 36, 37, 39, 40	34	85%
2.	Good	15	1	2,5%
3.	Medium	7, 32	2	5%
4.	Low	1, 38	2	5%
5.	Very Low	21	1	2,5%

Table 10. Analysis of the Effectiveness of Distracters on Multiple Choice Questions End of Year Assessment for PAI & BP Class IX Islamic Middle School Al-Azhar Semarang

No	Distractor Effectiveness	Distractor Effectiveness	Total	Percentage
1.	Very Good	2, 4, 15, 20, 21, 22, 24, 25, 26, 27, 28, 30	12	40%
2.	Good	1, 3, 7, 8, 9, 17, 23, 29	8	27%
3.	Medium	7, 32	2	7%
4.	Low	1	1	3,3%
5.	Very Low	5, 6, 11, 12, 13, 18, 19	7	23,33%

Based on Table 9 above, it can be concluded that the End of Year Assessment (PAT) questions for class IX PAI and BP subjects at SMP Negeri 16 Semarang in terms of the effectiveness of the distractors function very well, namely 85%, while in table 10, it can be concluded that the end of year Assessment questions Year (PAT) PAI and BP subjects class IX Islamic Middle School Al-Azhar 29 Semarang in terms of the effectiveness of the distractors also functioned very well, namely 40%.

Discussion

The results of the item analysis of validity, the End of Year Assessment (PAT) questions for class IX PAI and BP subjects at SMP Negeri 16 Semarang had 40 questions, with validity of 20 valid questions and 20 invalid questions, while the End of Year Assessment questions (PAT) PAI and BP subjects for class IX Islamic Middle School Al-Azhar 29 Semarang have 30 questions with validity, 15 questions are valid and 15 questions are invalid.

From these results, the number of questions that are declared valid must be maintained in one way, namely documenting them in the question bank of each school, and questions that are declared invalid should be corrected so that they can become higher quality questions by maximizing the technical mastery of each subject teacher in each school in preparing the question items. Sudjana (2016) states that the validity of the items of a test is measured by the accuracy of what should be measured through the items. The items that are declared valid in each school are items that have successfully carried out their function, namely as a measure of what should be measured. Then, invalid questions from each school can be caused by several factors, according to Groundland (2014), which can be several factors: (1) Instrument factors, (2) Administration

and scoring factors, and (3) Factors in students' answers. In this case, a factor in the possibility of invalid question items in each school is the factor of answers from students who have a tendency to answer quickly but inaccurately.

The results of the item analysis of reliability, the End of Year Assessment (PAT) questions for class IX PAI and BP subjects at SMP Negeri 16 Semarang had a reliability level of 0.88 in the very high category, while the End of Year Assessment (PAT) questions for class IX PAI and BP subjects and Islamic Middle School Al-Azhar 29 Semarang has a reliability level of 0.44 in the medium category. This is in line with the opinion of Arikunto (2013), who stated that if an instrument has a high validity test, then the level of reliability will also be consistent or high too.

The results of the item analysis of the level of difficulty of the End of Year Assessment (PAT) questions for class IX PAI and BP subjects at SMP Negeri 16 Semarang have 16 questions in the easy category, 22 questions in the medium category, 2 questions in the difficult category. Meanwhile, the End of Year Assessment (PAT) questions for PAI and BP class IX Islamic Middle School Al-Azhar 29 Semarang have 25 questions in the easy category, 5 questions in the medium category, 0 questions in the difficult category. According to Muluki et al (2020), whether a question is said to be of good quality or not can be seen from the level or degree of difficulty of the question. Items can be said to be good if they are neither too difficult nor too easy. From this opinion, it can be seen that each school has a tendency for questions to be in the easy category and this can be material for improvement for the two schools so that they can be further improved or paid more attention to the level of difficulty in End of Year Assessment (PAT) questions for PAI and BP class IX at SMP Negeri 16 Semarang and SMP Islam Al-Azhar 29 Semarang.

Furthermore, according to Arifin (2014), the calculation obtained from discriminating power is a type of measurement to determine the extent to which a question item is able to differentiate students who have mastered the material from students who have not mastered the material optimally according to certain criteria. Sudjana (2016) also agrees that with this differentiating power, what is present in each question item can be very important to know so that the question preparers are aware of the diversity of abilities possessed by students.

The results of item analysis of the discriminating power of End of Year Assessment (PAT) questions for class IX PAI and BP subjects at SMP Negeri 16 Semarang have sufficient discriminating power, namely 47.5%. Meanwhile, in Table 8, it can be concluded that the End of Year Assessment (PAT) questions for PAI and BP class IX Islamic Middle School Al-Azhar 29 Semarang have sufficient differentiating power, namely 50%. In this case, question items that have reached the level of excellent, good, and sufficient should be maintained and can even be improved further so that the results are of higher quality so that the question items can indicate that the student has good, intermediate, or poor abilities.

The results of the analysis of the effectiveness of the distractors on the End of Year Assessment (PAT) questions for class IX PAI and BP subjects at SMP Negeri 16 Semarang in terms of the effectiveness of the distractors functioned very well, namely 85%, while in table 10, it can be concluded that the End of Year Assessment (PAT) questions PAI and BP subjects for class IX Islamic Middle School Al-Azhar 29 Semarang in terms of the effectiveness of the distractors also function very well, namely 40%. According to Arikunto (2013), questions that can be categorized

as good-quality questions can indicate students' diverse abilities. In short, the higher the student's ability to understand the material presented by the teacher, the higher the opportunity for students to give the right or correct answer.

CONCLUSION

At SMP Negeri 16 Semarang, 20 questions are valid questions and can be stored in the school's question bank to be used in other tests relevant to PAI and BP subjects, while 20 questions are invalid and need to be improved through improvements so that it can be used in subsequent tests. At Al-Azhar 29 Islamic Middle School Semarang, 15 questions are good questions and can be stored in the school's question bank so they can be used in other tests that are relevant to PAI and BP subjects, while 15 questions are good questions are invalid and need to be improved through revisions so that the items can be used in subsequent tests.

REFERENCES

Agustiana, M., *et al.* 2018. Analisis Butir Soal Ulangan Akhir Semester Mata Pelajaran Bahasa Indonesia Kelas XI. *Jurnal Ilmiah Bina Edukasi Universitas Bina Darma*. 11(1):26–35.

Arifin, A.H. 2017. Analisis Butir Soal Ujian Akhir Semester Genap Mata Pelajaran Komputer Akuntansi Kelas XI Akuntansi Smk Muhammadiyah 1 Borobudur. *Jurnal Kajian Pendidikan Akuntansi Indonesia*. 1(1):1–18.

Arifin, Z 2014. *Evaluasi Pembelajaran: Prinsip, Teknik, Prosedur*. Bandung: PT Remaja Rosdakarya.

Arikunto, S. 2013. *Prosedur Penelitian Suatu Pendekatan Praktik*. Jakarta: Rineka Cipta.

Groundland. 2014. *Penyusunan Test Tertulis*. Jakarta: Prenada Media Group.

Hanifah, N. 2014. Perbandingan Tingkat Kesukaran Daya Pembeda Butir Soal Dan Reliabilitas Tes Bentuk Pilihan Ganda Biasa Dan Pilihan Ganda Asosiasi Mata Pelajaran Ekonomi. *Jurnal Sosio E-Kons*. 6(1):1–10.

Haris, A., & Asep, J. 2013. *Evaluasi Pembelajaran*. Yogyakarta: Multi Pressindo.

Hasibuan, E.S.B. 2013. Analisis Soal Ulangan Tengah Semester Bahasa Indonesia Kelas VII Mas Raudhatul Ulum MerantI Pontianak. *Jurnal Pendidikan Dan Pembelajaran Khatulistiwa*. 2(4):1–15.

Lestari, S., *et.al*. 2019. Analisis Butir Soal Ulangan Akhir Semester Ganjil Mata Pelajaran Matematika Kelas III SD Kecamatan Peturukan Tahun 2018/2019. *Jurnal Kajian Pendidikan Dan Pendidikan Dasar*. 9(2):118–125.

Lubis, M. 2008. *Evaluasi Pendidikan Nilai*. Yogyakarta: Pustaka Pelajar.

Mahirah, B. 2017. Evaluasi Belajar Peserta Didik (Siswa). *Jurnal Idaarah*. 1(2):257–267.

Muluki, A, *et.al*. 2020. Analisis Kualitas Butir Tes Semester Ganjil Mata Pelajaran IPA Kelas IV MI Radhiatul Adawiyah. *Jurnal Ilmiah Sekolah Dasar*. 4(1):86–96.

Muslim, A. 2017. *Analisis Butir Soal Ulangan Akhir Semester Mata Pelajaran Pendidikan Agama Islam (PAI) Kelas VIII Semester Genap Sekolah Menengah Pertama (SMP) Kota Semarang Tahun Pelajaran 2016/2017*. Skripsi Prodi Pendidikan Agama Islam. UIN Walisongo Semarang.

Muzayyanah, I. 2020. *Analisis Butir Soal Penilaian Tengah Semester Multiple Choice Pai & Bp Buatan Guru (Perbandingan SMPN 18 Dan SMPN 44 Semarang*. Skripsi Prodi Pendidikan Agama Islam. UIN Walisongo Semarang.

Nikto, A. 1996. *Educational Assessment Of Student Second Edition*. Ohio: Merril is an imprint of Prentice Hall Englewood Cliffs.

Observasi Peneliti. 2023. *Hasil Observasi Di SMP Negeri 16 Semarang Dan SMP Islam Al Azhar 29 Semarang*. Semarang.

Queenta, A.S., & Yuliasma. 2020. Analisis Butir Soal Ujian Tengah Semester Ganjil Mata Pelajaran Seni Budaya Kelas VII Di SMPN 5 Padang Tahun Ajaran 2019/2020. *Jurnal Ilmiah Pendidikan Seni Pertunjukan*. 9(1):1–13.

Rahmawati, A.N. 2015. *Analisis Kualitas Soal Ujian Akhir Semester Mata Pelajaran Pendidikan Agama Islam Dan Budi Pekerti Kelas VIII SMP Muhammadiyah 4 Dan SMP Piri 2 Tahun Ajaran 2014/2015*. Skripsi Prodi Pendidikan Agama Islam. UIN Walisongo Semarang.

Rohmah, W.N. 2017. Analisis Butir Soal Evaluasi Ulang Akhir Semester Ganjil Kelas IV Ilmu Pengetahuan Sosial Se-Kecamatan Mojoroto Kota Kediri Tahun Ajaran 2016/2017. *Jurnal Simki-Pedagogia*. 1(10):1–12.

Sanjaya, W. 2015. *Kurikulum Dan Pembelajaran: Teori Praktik Pengembangan Kurikulum Tingkat Satuan Pendidikan (KTSP)*. Jakarta: Kencana.

Sarea, M.S. 2022. Analisis Kualitas Soal Ujian Akhir Semester Pendidikan Agama Islam Dan Budi Pekerja Pada SMPN Kota Makassar. *Jurnal Pendidikan Islam*. 11(2):179–189.

Septiana, N. 2016. Analisis Butir Soal Ulangan Akhir Semester (UAS) Biologi Tahun Pelajaran 2015/2016 Kelas X Dan XI Pada MAN SAMPIT. *Jurnal Pendidikan Sains Dan Matematika*. 4(2):115–121.

Solichin, M. 2017. Analisis Daya Beda Soal, Taraf Kesukaran, Validitas Butir Tes, Interpretasi Hasil Tes Dan Validitas Ramalan Dalam Evaluasi Pendidikan. *Jurnal Manajemen Dan Pendidikan Islam*. 2(2):192–213.

Sudijono, A. 2015. *Pengantar Evaluasi Pendidikan*. Jakarta: PT Raja Grafindo Persada.

Sudjana, N. 2016. *Penilaian Hasil Proses Belajar Mengajar*. Bandung: PT Remaja Rosdakarya.

Sumiati, A., *et.al*. 2018. Workshop Teknik Menganalisis Butir Soal Dalam Meningkatkan Kompetensi Guru Di SMK Cileungsi Bogor. *Jurnal Pemberdayaan Masyarakat Madani*. 2(1):136–153.

Susanto, H., *et.al*. 2015. Analisis Validitas Reliabilitas Tingkat Kesukaran Dan Daya Beda Pada Butir Soal Ujian Akhir Semester Ganjil Mata Pelajaran Matematika." *Jurnal Pendidikan Matematika*. 6(2):203–217.

Wati, R. 2013. Analisis Butir Soal Matematika Pada UKA PLPG LPTK Fakultas Tarbiyah IAIN Antasari Banjarmasin. *Jurnal UIN Antasari*. 4(2):1–15.

Yusiana., & Zamsir. 2018. *Perbandingan Kualitas Butir Soal Buatan Guru Pada Mata Pelajaran Ekonomi Antara Sekolah Yang Berakreditasi A Dan Berakreditasi B*. Skripsi Prodi Pendidikan Ekonomi. Universitas Lampung.