

## Educational Evaluation: Analyzing The Quality Of Summative Tests At Modern Junior High School Al Rifa Ie Malang

**Susilo Hidayah**

Universitas Islam Negeri Maulana Malik Ibrahim Malang, Indonesia  
[susilohidayah3@gmail.com](mailto:susilohidayah3@gmail.com)

**Khalishah Dyah Capriatin**

Universitas Islam Negeri Maulana Malik Ibrahim Malang, Indonesia  
[khalishahcapriatin71@guru.smp.belajar.id](mailto:khalishahcapriatin71@guru.smp.belajar.id)

**Indah Aminatuz Zuhriyah**

Universitas Islam Negeri Maulana Malik Ibrahim Malang, Indonesia  
[zuhriyah@pgmi.uin-malang.ac.id](mailto:zuhriyah@pgmi.uin-malang.ac.id)

DOI: 10.32528/tarlim.v8i1.2658

---

**Track:**

Received:

23 November 2024

Final Revision:

23 Januari 2025

Available online:

10 Maret 2025

Corresponding Author:

[susilohidayah3@gmail.com](mailto:susilohidayah3@gmail.com)

**ABSTRACT**

*This article aims to analyze the quality of even semester final summative questions of PAI subjects in class VIII of Modern Al Rifa Ie Junior High School in Malang in the 2023/2024 academic year. This research is a quantitative descriptive research. The results showed that the quality of question items based on validity, on short description questions there are 13 valid items, true-false questions there are 6 questions, the type of long description questions there are 5 items. The quality of question items based on reliability, on the short description questions the questions used have a high reliability coefficient of 0.798, true-false questions have a high reliability coefficient of 0.609, while the type of long description questions have a low reliability coefficient of 0.560. The quality of questions based on the level of difficulty, in the short description question there is 1 difficult item, 2 medium items, and 17 easy items. In true-false questions and long descriptions, all questions are categorized as easy items. The quality of questions based on differentiating power, in the short description question there are 12 items that have poor differentiating power, 2 items have sufficient differentiating power, 3 items have good differentiating power, and 3 items have excellent differentiating power. For the results of the analysis on true-false questions, there are 6 items that have poor differentiating power, 1 item has sufficient differentiating power, and 3 items have good differentiating power. As for the long description items, there are 2 items that have poor differentiating power, 2 items have sufficient differentiating power, and 1 item has excellent differentiating power.*

*Keywords: Evaluation, Quality, Summative Test*

---

### INTRODUCTION

In the learning process activities, evaluation is needed by a teacher to determine the level of understanding of students in solving learning problems (Yusuf, 2021). Evaluation is a very important component in the implementation of education, because with a good evaluation system, the quality of learning is expected to improve (Rizema, 2013). To improve the quality of learning, the evaluation is expected to pay attention to the existing components. However, the evaluation of education implemented so far can be said to still not provide sufficient distribution to improve the quality of education. This is due to the evaluation system used that is not appropriate or the implementation is not what is expected (Rukajat, 2018) Huda, H., Nursyamsiyah, S., & Alfian, M. (2022). Therefore, it is necessary to pay attention to the evaluation system used, with a good evaluation system it will encourage educators to determine a good teaching strategy

so that it can motivate students.

Student competency achievement refers to the ability or results achieved by students in mastering various competencies or abilities that have been determined in the applicable curriculum or learning standards. These competencies can include various aspects, such as knowledge, skills, attitudes, and certain values that students are expected to master after completing a learning process (Anwar, 2019). The achievement of student competencies is the main focus in the educational process because it measures the extent to which learning objectives have been achieved. Evaluation of student competency achievement is carried out through various evaluation instruments, such as exams, assignments, projects, observations, and so on (Wildan, 2017).

The step that must be taken to determine the quality of a test is to analyze the quality of test items or items. To measure the quality of a good learning outcome test, one of which can be determined from the quality of the questions given. Likewise, good questions are also determined by the quality of the points (Amrullah, 2018). A learning outcome test that contains high-quality questions, even in small numbers, will be much better and more useful than a learning outcome test that contains dozens of low-quality questions that will reduce the test function and deviate measurement results. Processing of learning outcomes tests in order to improve the learning process can be done in several ways, one of which is by making item analysis. (Arikunto, 2018: 230).

Quality analysis is a process of in-depth evaluation of various aspects or characteristics that determine the level of excellence or conformity of a product, service, process, or system to established standards or desired expectations (Septiani et al., 2020). The quality analysis referred to in this study is the analysis of the quality of the question items, the analysis activity aims to identify which questions are good, less good, and questions that are not good or bad so that they need improvement. To find out the quality of these items can be done by testing the quality of several aspects, namely validity, reliability, difficulty level, differentiating power and answer distribution patterns. (Ismail & Dkk, 2020).

Islamic Religious Education (PAI) is one of the subjects that has an important role in shaping students' character and spirituality. (Yeni Hartanti, 2021), both from elementary school to high school. In the context of formal education, evaluation of a student's competency achievement is often carried out through the End of Semester Summative (SAS) exam. (Astuti et al., 2024). This exam not only measures students' understanding of the material that has been learned, but also becomes an indicator of the extent to which they are able to apply Islamic values in everyday life. Item analysis conducted to measure the quality of end-of-semester summative questions in PAI subjects has an important role in determining the validity, reliability, and success of student learning achievement. By considering various aspects such as the level of difficulty of the questions, the clarity of the questions, and the direct relationship with the expected competencies, item analysis can provide an accurate picture of the extent to which students have mastered the learning materials, especially PAI subjects.

The purpose of item analysis is not only to assess individual student success, but also to evaluate the effectiveness of the curriculum and teaching methods applied. Thus, the results of item analysis can be the

basis for teachers in designing more effective learning strategies. In this context, the researcher will discuss the importance of item analysis in the end-of-semester summative examination of Islamic Education subjects at the junior high school level, namely Modern Junior High School AL Rifa Ie Malang, and how the results of this analysis can make a positive contribution in improving the achievement of student competencies and the quality of Islamic education in schools.

Based on observations in the field, it was found that the final exam questions such as the End-of-Semester Summative in Modern Al Rifa Ie Junior High School were made by the subject teachers at the school. The questions are made based on the provisions imposed by Pusmendik (Education Assessment Center), so it is hoped that the questions made can be of quality and can be accepted and can be understood properly by students.

In a study conducted by Nurdin, it was stated that based on the results of the analysis of validity, reliability, difficulty level and differentiating power, it can be concluded that objective questions with multiple choice types of PAI class VI subjects used in the Odd Semester Final Examination (UAS) for the 2020/2021 academic year at SDIK Nurul Qur'an are not all of very good/good quality. (Nurdin, 2020). The analysis results show that 6 questions (40%) have unfavorable criteria, 4 questions (26.67%) have unfavorable criteria, and 5 questions (33.33%) have excellent criteria. Therefore, it is necessary to have training in the development of item preparation so that in the future it will be able to provide an increase in the items to be given and tested. In other research also shows the results that overall the results of the analysis of the quality of the odd semester final test items of PAI class XII at SMA Negeri 2 Kroya Cilacap district show the quality of the items that are not ideal. (Pangesti, 2019). Likewise, the results of research conducted at SMA Negeri 1 Rao Utara in PAI class X in the 2023/2024 academic year showed that there were questions that were not well weighted (Nurlaila et al., 2023). For this reason, further improvements must be made to the question items so that the valid weights can be included in the question bank.

From some of the research above, it shows that one of the reasons for this condition is because teachers have not understood and have not developed questions and analyzed items in accordance with the mechanisms, principles, and procedures of assessment as described above. Therefore, the author is interested in analyzing the question items in PAI subjects in achieving student competence at Modern Al Rifa Ie Junior High School in Malang. The reason for choosing the place of research is because the institution does not use questions as in general. The questions are made by the subject teacher at the institution, which of all the questions are in the form of descriptions. Therefore, in this article the author seeks to try to analyze the items used in the institution to find out whether the question can be said to be good and feasible to enter into the question bank and be able to improve student competence.

The difference between this research and previous research lies in the types of questions analyzed, namely researchers analyzing short description questions, true-false questions, and long descriptions. Where in previous studies only analyzed the type of multiple choice objective questions. The results of this study are expected to be able to provide academic benefits and add insight to both students and teachers and in the world of education globally, especially regarding how to analyze the quality of items and why the analysis is

important to do. In addition, it is hoped that the results of this study can be used as one of the references in developing knowledge about educational evaluation, especially in analyzing the quality of tests and items either in PAI subjects or other general subjects.

## RESEARCH METHOD

This research is included in the quantitative descriptive approach. This research only provides a description of existing phenomena, which intends to analyze or search for data that can be used to describe tests in the form of question items. (Nurian, 2019). The population in this study were the answer sheets of class VIII students at Modern Al Rifa Ie Junior High School in Malang totaling 135 children, and the sample used in this study was 40 answer sheets, which contained short description questions, true-false questions and long descriptions. The sampling technique uses a random sampling technique with the provisions of 30% of the total population. (Sugiyono, 2017).

The research instrument was in the form of a summative question at the end of the even semester of PAI subjects in the 2023/2024 academic year along with the answer key obtained from the teacher who made the question and the results of students' answers after the test. Data collection techniques through observation and tests. As for the implementation process, the researcher actually only intends to obtain data and information which can be used to describe the quality of the items used.

Data analysis carried out on question items uses descriptive statistical analysis techniques, namely calculating validity, reliability, difficulty level, and distinguishing power by analyzing numerical data in order to provide a clear description of an event or situation, so that later conclusions can be drawn. In analyzing the descriptive statistics of the items, it was carried out using the help of the SPSS version 25 application program.

## RESULTS & DISCUSSION

The summative questions for the end of the even semester of PAI subjects in class VIII of Modern Al Rifa Ie Junior High School in Malang in the 2023/2024 academic year consisted of 20 short description questions, 10 true-false questions, and 5 long description questions.

### ***Result of Item Validity Analysis***

A test can be said to be valid if the test can measure or reveal what should be measured or revealed through the test (Supriyadi, 2011). So a question item can be declared valid if the item is produced as a measuring tool for the success of students' learning is done correctly, correctly, validly, can measure or reveal the learning outcomes that have been achieved by students.

Testing the validity of items on the Even Semester End Summative question of PAI subjects in class VIII of Modern Al Rifa Ie Malang Junior High School in the 2023/2024 school year using the biserial point formula with the help of the SPSS version 25 program. The results of the calculation are then consulted with rtable at a significance level of 5% or 0.05. The number of research subjects was 40 student answer sheets, so at a significance level of 0.05 and  $n = 40$ , the rtable value was 0.312. If  $r_{table} < r_{count}$  then the item is valid,

otherwise if  $r_{table} > r_{count}$  then the question is invalid.

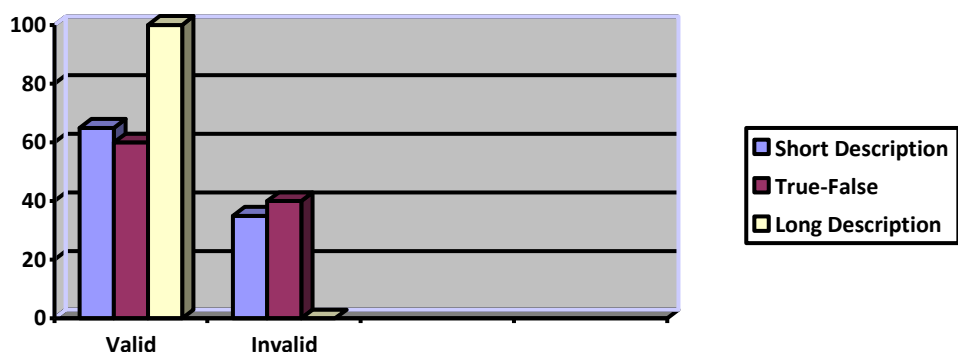
Based on the results of the analysis on the summative questions of the End of Even Semester of PAI subjects in class VIII of Modern Al Rifa Ie Junior High School in Malang in the 2023/2024 academic year, there were 13 questions or 65% declared valid and 7 questions or 35% declared invalid on short description type questions, 6 questions or 60% declared valid and 4 questions or 40% declared invalid on true-false type questions, and there were 5 questions on long description type questions, all of which were declared 100% valid. Based on this description, it can be concluded that the summative questions of the End of the Even Semester of PAI subjects in class VIII of Modern Al Rifa Ie Junior High School in Malang in the 2023/2024 academic year include good questions in terms of validity.

**Table 1.**

Distribution of Even Semester Final Summative Questions of PAI subjects in class VIII of Modern Junior High School Al Rifa Ie Malang in the 2023/2024 academic year Based on Validity aspects

Questions Type	Item Number	Total	Percentage
<b>Short Description</b>			
Valid	1,2,3,4,8,9,10,11,12,14,15,18,20	13	65%
Invalid	5,6,7,16,17,19	7	35%
<b>True-False</b>			
Valid	4,5,6,8,9,10	6	60%
Invalid	1,2,3,7	4	40%
<b>Long Description</b>			
Valid	1,2,3,4,5	5	100%
Invalid	-	-	-

The following is a chart of the percentage result of item validity analysis on the summative questions of the End of Even Semester of PAI subjects in class VIII of Modern Al Rifa Ie Junior High School in Malang in the 2023/2024 academic year.



**Picture 1. Percentage Result of Item Validity Analysis**

Questions that have been declared valid must be maintained by documenting the question in the

question bank, while invalid items should be improved by increasing the teacher's mastery of techniques in compiling items. The question items become valid because the construction is good and covers material that truly represents the target of the measure, namely test participants or testees from class VIII students of Modern Al Rifa Ie Junior High School in Malang in the 2023/2024 academic year.

Based on the results of the analysis on the Final Even Semester Summative question items of PAI subjects in class VIII of Modern Al Rifa Ie Junior High School Malang in the 2023/2024 school year, it can be seen that the validity test on the short description type questions there are 13 questions or 65% declared valid and 7 questions or 35% declared invalid. As for the results of the validity test of true-false type items, there are 6 questions or 60% declared valid and 4 questions or 40% declared invalid. And for the results of the validity test of long description type items, all of these items were declared 100% valid.

The results of the analysis above are almost similar to the results of research conducted by Ardilah Muluki et al. which states that when viewed from the validity aspect of the form of multiple choice questions, the valid questions are 70% and the invalid questions are 30% (Muluki, 2020). In accordance with the theory that says that the validity of an item or item of a test is the accuracy of measuring what is owned in measuring what should be measured through the test item (Sudijono, 2003). From the results of the analysis of the three types of items above, it can be said that the summative items of the End of Even Semester of PAI subjects in class VIII of Modern Al Rifa Ie Junior High School in Malang in the 2023/2024 academic year have been able to measure what should be measured.

Likewise, invalid question items, namely a total of 7 questions or 35% can be caused by several factors. As for several factors that affect the validity test results of the test results, namely the instrument factors used for the test, administration and scoring factors, and factors from the answers of test takers (Arifin, 2014). In the Even Semester Final Summative question of PAI subjects in class VIII of Modern Junior High School Al Rifa Ie Malang in the 2023/2024 academic year which is declared invalid can be influenced by several of the above factors. Both the error factor from making the instrument and the tendency of students to answer questions quickly so that they get answers that are less precise.

Based on the description above, it can be concluded that the summative questions of the End of Even Semester of PAI subjects in class VIII of Modern Al Rifa Ie Junior High School in Malang in the 2023/2024 academic year can be said to have good quality from the aspect of validity. Valid question items can be stored in the question bank, and conversely invalid question items need to be improved by adjusting indicators and increasing mastery of techniques regarding the preparation of question items. It is hoped that in the future teachers who make questions or subject teachers at Modern Al Rifa Ie Junior High School in Malang can make questions of very good quality so that they can improve student competency achievement.

The indicators of competency achievement that are achieved in these items are: 1) Students can understand the provisions and procedures for performing voluntary prayers in congregation and munfarid, 2) Students can practice voluntary prayers in congregation and munfarid, 3) Students can understand the provisions and procedures for prostration of gratitude, prostration of sahwī and prostration of recitation, 4) Students can practice prostration of gratitude, prostration of sahwī and prostration of recitation, 5) Students

can understand the procedures for voluntary and mandatory fasting, 6) Students can identify the wisdom of carrying out voluntary and mandatory fasting, 7) Students can understand the provisions of halal and haram food and drinks based on the Qur'an and Hadith, 8) Students can identify the wisdom of consuming halal and haram food and drinks based on the Qur'an and Hadith, 9) Students can understand the history of the growth of science during the Bani Umayyah period, and 10) Students can understand the history of the growth of science during the Bani Abbasiyah period.

### ***Results of Item Reliability Analysis***

Reliability is the result of an evaluation that has been carried out to show the same results when given to the same students at different times (Supriyadi, 2011). Reliability analysis is used with the aim of knowing how much consistency the test scores achieved by students or test takers on the test when done repeatedly.

Reliability testing on even semester final summative questions of PAI subjects in class VIII of Modern Al Rifa Ie Malang Junior High School in the 2023/2024 academic year using the Alpha Cronbach internal consistency technique using the help of SPSS version 25. The basis for making decisions on testing the reliability of the items is if the Alpha coefficient  $\geq 0.6$  then all items or items are said to be reliable (Sugiyono, 2017). So that if the item is used repeatedly to measure the achievement of student competence on the same object, it will provide the same data results.

The results of the analysis showed that 13 short description items produced Cronbach's Alpha of 0.798 and 6 true-false items produced Cronbach's Alpha of 0.609, which means that the items used can be said to be reliable because  $\geq 0.6$ . So that if the question items are used repeatedly, they will give the same results. Meanwhile, the long description type question item produces Cronbach's Alpha of 0.560, which means that the question items used are not reliable because they are  $< 0.6$ . So that the item although it can be said to be good but cannot be used repeatedly and with the same results.

**Table 2.** Distribution of Even Semester Final Summative Questions of PAI subjects in class VIII of Modern Junior High School Al Rifa Ie Malang in the 2023/2024 academic year Based on Reliability aspects

Questions type	<i>Alpha Cronbach</i>	Total Items	Description
<b>Short Description</b>	0,798	13	<i>Reliable</i>
<b>True-False</b>	0,609	6	<i>Reliable</i>
<b>Long Description</b>	0,560	5	<i>Un-Reliable</i>

A test instrument that has good validity on each item or item will have a high level of reliability as well. In line with the theory that states that a test consisting of many items will be more valid than a test consisting of only a few items (Arikunto, 2011). The high and low level of validity can indicate the high and low reliability coefficient, so the longer the test, the higher the reliability. Based on the description above, it can be concluded that the question of the Summative Question of the End of the Even Semester of PAI subjects in class VIII of Modern Al Rifa Ie Malang Junior High School in the 2023/2024 academic year in the type of short description questions and bear-false questions is a good question in terms of reliability.

Unlike the long description type questions, it shows that the items used are not reliable. If the item is

valid but not reliable, it means that the item cannot be used repeatedly to different test takers with the same test results. In accordance with the theory which states that there are several factors that affect the reliability coefficient, one of which is the factor of scoring objectivity (Sax, 1980). In addition to these factors due to the possibility of ambiguous words that cause multiple perceptions, either in the same language or in different languages and cultures of respondents, or also questions that may change their answers according to time even though the respondents remain the same.

Objective tests can be analyzed more accurately than subjective or description tests. It must be recognized that the quality of the description questions depends on the student's ability to answer questions, as well as the teacher's honesty in evaluating the test results. Description items can only be used once, while objective items can be used many times. (Kurniawan et al., 2022). Because basically the long description test is included in the type of subjective test where the scoring system depends on the subjectivity of the teacher or grader, it can be concluded that the summative item at the end of the even semester of PAI subjects in class VIII of Modern Junior High School Al Rifa Ie Malang in the 2023/2024 academic year on the type of long description question cannot be used multiple times.

**Results of Item Difficulty Analysis**

The quality or not of a learning outcome test can be known from the level of difficulty or level of difficulty possessed by each item or test item. The learning outcome test items can be declared as good items if the items are not too difficult and not too easy, in other words the degree of difficulty of the test items is moderate or sufficient. (Supriyadi, 2011). The calculation of the level of difficulty of the items can be interpreted in three categories, namely: items with a value of 0.00-0.30 belong to the difficult category, items with a value of 0.31-0.70 belong to the medium category, and items with a value of 0.71-1.00 belong to the easy category.

Based on the results of the analysis, it was found that in the type of short description questions, there were 1 item or 5% in the difficult category, 2 items or 10% in the moderate category, and 17 items or 85% in the easy category. In the type of true-false questions and long descriptions, all the items used are included in the easy category. The following is a description of the question items based on the level of difficulty index aspect.

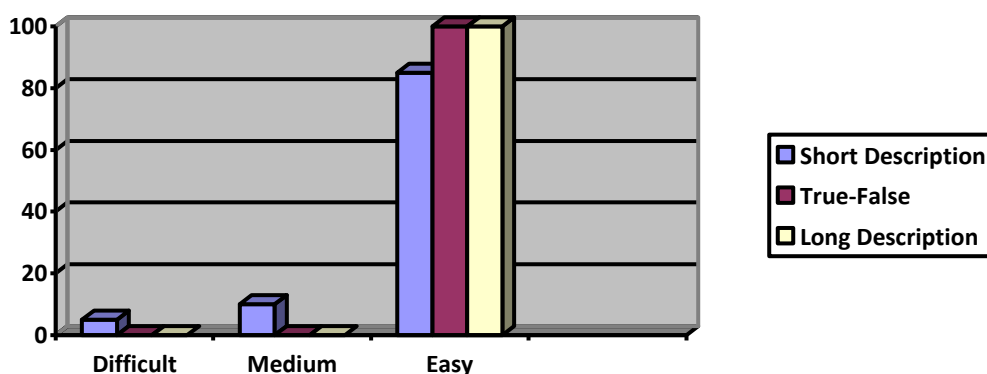
**Table 3.** Distribution of Summative Questions of the End of Even Semester of PAI subjects in class VIII of Modern Junior High School Al Rifa Ie Malang in the 2023/2024 academic year based on the aspect of Level of Difficulty.

Questions type	Item Number	Total	Percentage
<b>Short Description</b>			
0,00-0,30 (Difficult)	19	1	5%
0,31-0,70 (Medium)	15, 16	2	10%
0,71-1,00 (Easy)	1,2,3,4,5,6,7,8,9,10,11,12,13,14,17,18,20	17	85%
<b>True-False</b>			



0,00-0,30 (Difficult)	-	0	0%
0,31-0,70 (Medium)	-	0	0%
0,71-1,00 (Easy)	1, 2, 3, 4, 5, 6, 7, 8, 9, 10	10	100%
<b>Long Description</b>			
0,00-0,30 (Difficult)	-	0	0%
0,31-0,70 (Medium)	-	0	0%
0,71-1,00 (Easy)	1, 2, 3, 4, 5	5	100%

The following is a chart of the percentage result of item difficulty analysis on the summative questions of the End of Even Semester of PAI subjects in class VIII of Modern Al Rifa Ie Junior High School in Malang in the 2023/2024 academic year.



**Picture 2. Percentage Result of Item Difficulty Analysis**

From the results of the analysis above, it can be seen that of the three types of questions used, almost all items are included in the easy category. This shows that there are two possibilities that can be concluded, namely that the questions used really include easy questions and another possibility is due to the high ability of students in understanding and mastering the material.

A question item can be considered good if the item has a difficulty level between 0.30 - 0.70 (Arikunto, 2011). So it can be concluded that the summative questions of the End of Even Semester of PAI subjects in class VIII of Modern Junior High School Al Rifa Ie Malang in the 2023/2024 academic year both of the types of short description, true-false and long description include questions that have a low level of difficulty because each item used almost entirely has an easy level of difficulty.

Follow-up that can be done after analyzing the level of difficulty of the items at a low or easy level of difficulty there are 3 possible follow-ups. (Sudijono, 2003), namely: 1) The item is discarded and will not be issued again on the upcoming learning outcomes test, 2) Re-examined whether there are factors that cause the item to be answered correctly by almost all participants.

Can be answered correctly by almost all test takers so that improvements can be made by fixing options and making question sentences more complex, then the questions can be issued again on future learning outcomes tests, 3) Question items are retained for use in tests that are loose or formalities only.

### ***Results of Item Distinguishing Power Analysis***

Item discriminating power is the ability of a learning outcome test to be able to distinguish between test takers who have high abilities (understand the material) and test takers who have low abilities (do not understand the material). The number indicating the distinguishing power ranges from -1 to +1, for a negative sign indicates that test takers who have low ability can answer correctly while test takers who have high ability answer incorrectly. Thus, questions with a negative index of differentiating power indicate the inverse quality of test takers. (Supriyadi, 2011). The sample used in this study was 40 students, so to calculate the distinguishing power of the upper and lower groups, 27% of the total sample was taken.

The categories used to interpret the results of the differentiator analysis are items that have a differentiator index with a value of 0.71 - 1.00 including the excellent category, a value of 0.41 - 0.70 including the good category, a value of 0.21 - 0.40 including the fair category, and a value of 0.00 - 0.20 including the poor category. Thus a question with a differentiating index of less than 0.20 means that the item is not feasible because it is unable to distinguish the ability of test takers from high and low groups.

Based on the results of the analysis on the type of short description questions, there are 12 items or 60% have poor differentiating power, 2 items or 10% have sufficient differentiating power, 3 items or 15% have good differentiating power, and 3 items or 15% have excellent differentiating power. For the results of the analysis on the type of true-false questions, there were 6 items or 60% had poor differentiating power, 1 item or 10% had sufficient differentiating power, and 3 items or 30% had good differentiating power.

From the results of the analysis of the short description and true-false questions above, it can be concluded that these items are dominated by poor differentiating power, causing the questions to be less able to objectively differentiate students' abilities, both students from high and low groups.

As for the long description type items, there are 2 items or 40% that have poor differentiating power, 2 items or 40% have sufficient differentiating power, and 1 item or 2% has excellent differentiating power. This shows that the question can be said to be able to distinguish the abilities of students from high and low groups.

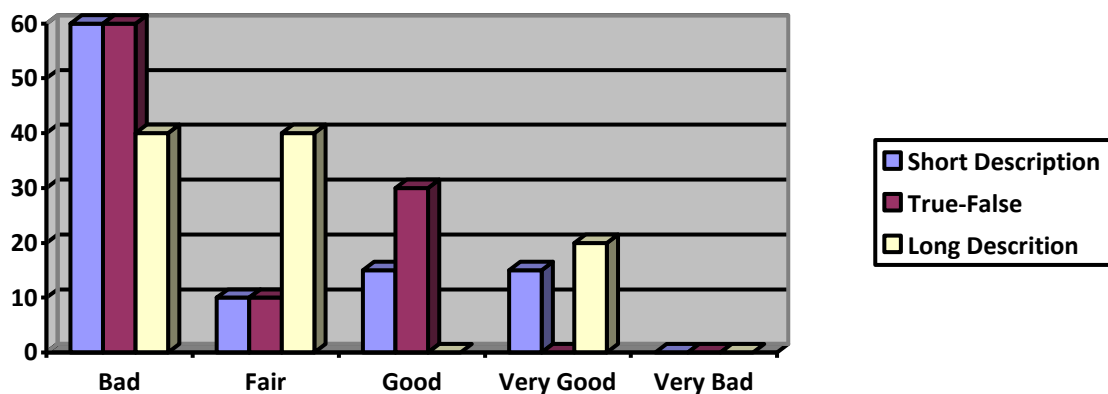
**Table 4.**

Distribution of Summative Questions of the End of Even Semester of PAI subjects in class VIII of Modern Junior High School Al Rifa Ie Malang in the 2023/2024 academic year based on the aspect of Distinguishing Power.

Questions type	Item Number	Total	Percentage
<b>Short Description</b>			
0,00-0,20 (Bad)	1, 2, 3, 4, 5, 6, 7, 8, 10, 16, 17, 19	12	60%
0,21-0,40 (Fair)	9, 13	2	10%
0,41-0,70 (Good)	12, 14, 15	3	15%
0,71-1,00 (Very Good)	11, 18, 20	3	15%
Negatif (Very Bad)	-	0	0%
<b>True-False</b>			

0,00-0,20 (Bad)	1, 2, 3, 4, 7, 8	6	60%
0,21-0,40 (Fair)	5	1	10%
0,41-0,70 (Good)	6, 9, 10	3	30%
0,71-1,00 (Very Good)	-	0	0%
Negatif (Very Bad)	-	0	0%
<b>Long Description</b>			
0,00-0,20 (Bad)	2, 3	2	40%
0,21-0,40 (Fair)	1, 4	2	40%
0,41-0,70 (Good)	-	0	0%
0,71-1,00 (Very Good)	5	1	20%
Negatif (Very Bad)	-	0	0%

The following is a chart of the percentage result of distinguishing power analysis on the summative questions of the End of Even Semester of PAI subjects in class VIII of Modern Al Rifa Ie Junior High School in Malang in the 2023/2024 academic year.



**Picture 3. Percentage Result of Item Distinguishing Power Analysis**

Items with excellent, good and sufficient differentiating power must be maintained by including these items in the question bank, while items with very poor and poor differentiating power must be completely repaired by tracing the cause of the failure. One of the efforts that can be made is to fix questions that are less clear in the meaning and intent of the items, so that they are ambiguous and confuse high-ability students in answering. The question items must be able to reflect the differences in ability between students who understand the material and students who do not understand the material. (Muluki, 2020).

Based on the discussion above, it can be concluded that the summative questions at the end of the even semester of PAI subjects in class VIII of Modern Al Rifa Ie Junior High School in Malang in the 2023/2024 academic year, the types of short description and true-false have poor differentiating power because most of the items are dominated by items that have poor differentiating power, so they have low differentiating power. As for the type of long description question, it has sufficient differentiating power.

Then the items that already have good differentiating power (good, and very good) should be included in the learning outcomes test question bank, and the items with poor differentiating power can be improved and developed.

**Quality of Question Items**

The results of the analysis of the aspects of validity, reliability, difficulty level and distinguishing power, the items are analyzed together to determine the quality of the questions as a whole. The quality of the items is adapted from the Likert scale which is divided into 5 categories, namely the categories of very good, good, medium, not good, and very bad.

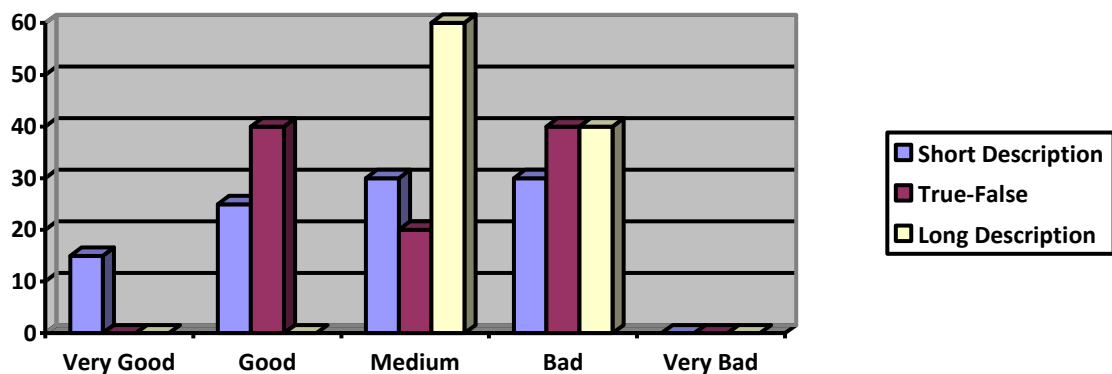
Based on the results of the analysis of the summative questions of the End of Even Semester of PAI subjects in class VIII of Modern Al Rifa Ie Junior High School Malang in the 2023/2024 academic year on the type of short description questions, there are 3 items of very good quality, 5 items of good quality, 6 items are included in the category of questions that have moderate quality, 6 items of poor quality, and there are no items that have very poor quality. In the type of true-false questions, there are 4 items of good quality, 2 items of moderate quality, 4 items are included in the category of questions that have poor quality, and there are no items that have very good and very poor quality. For the type of long description questions, there are 3 items of moderate quality, 2 items of poor quality, and there are no items of excellent, good and very poor quality.

**Table 5.**  
 Question Distribution Based on Question Quality

Questions type	Item Number	Total	Percentage
<b>Short Description</b>			
Very Good	15,18,20	3	15%
Good	9,1,11,12,14	5	25%
Medium	1,2,3,4,8,16	6	30%
Bad	5,6,7,13,17,19	6	30%
Very Bad	-	0	0%
<b>True-False</b>			
Very Good	-	0	0%
Good	5,6,9,10	4	40%
Medium	4,8	2	20%
Bad	1,2,3,7	4	40%
Very Bad	-	0	0%
<b>Long Description</b>			
Very Good	-	0	0%
Good	-	0	0%
Medium	1,4,5	3	60%
Tidak Good	2,3	2	40%
Very Bad	-	0	0%

The following is a chart of the percentage result of question distribution based on question quality on

the summative questions of the End of Even Semester of PAI subjects in class VIII of Modern Al Rifa Ie Junior High School in Malang in the 2023/2024 academic year.



Picture 4. Percentage Result of Question Distributin Based on Question Quality

Based on this, the summative question at the end of the even semester of PAI subjects in class VIII of Modern Al Rifa Ie Junior High School in Malang in the 2023/2024 academic year has not been able to carry out its function properly. The failure of the question item to carry out its function is due to the non-fulfillment of one of the question parameters.

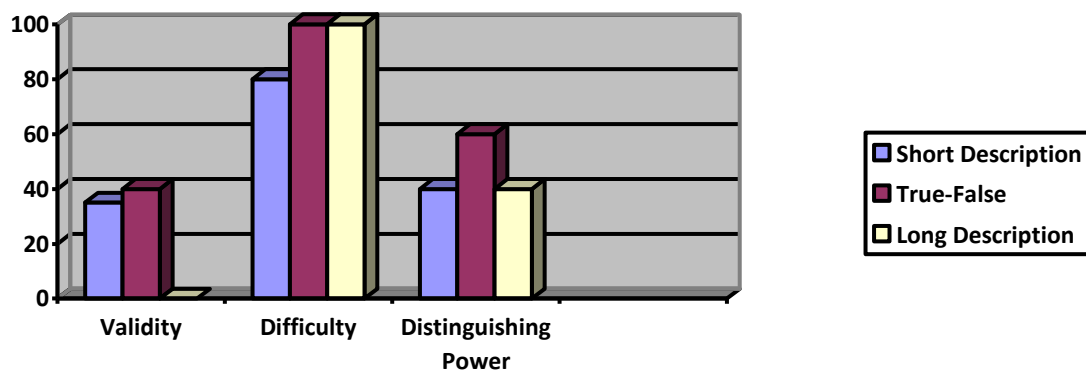
Table 6.

Causes of item failure

Questions type	Item Number	Total	Percentage
<b>Short Description</b>			
Validity (Invalid)	5,6,7,13,16,17,19	7	35%
Difficulty (Difficult & Easy)	1,2,3,4,5,6,7,8,9,10,11,12,13,14,17,19	16	80%
Distinguishing Power (Bad & negatif)	1,2,3,4,5,6,7,8,13,16,17,19	8	40%
<b>True-False</b>			
Validity (Invalid)	1,2,3,7	4	40%
Difficulty (Difficult & Easy)	1,2,3,4,5,6,7,8,9,10	10	100%
Distinguishing Power (Bad & negatif)	1,2,3,4,7,8	6	60%
<b>Long Description</b>			
Validity (Invalid)	-	0	0%
Difficulty (Difficult & Easy)	1,2,3,4,5	5	100%
Distinguishing Power (Bad & negatif)	2,3	2	40%

The following is a chart of the percentage result of causes of item failure on the summative questions

of the End of Even Semester of PAI subjects in class VIII of Modern Al Rifa Ie Junior High School in Malang in the 2023/2024 academic year.



Picture 5. Percentage Result of Causes of Item Failure

Based on the table above, it can be seen that in addition to some of the invalid items, the most dominating cause of failure is in the aspect of the difficulty level and distinguishing power of the items. This shows that the items have not been able to distinguish students who have high abilities and low abilities. Other failures are also caused because the items used are still too difficult and too easy.

The follow-up that can be done includes: 1) items with very good quality will be included in the question bank so that they can be used again as future tests, 2) items with good and moderate quality need to be revised first according to the failure indicator. 3) items with poor and very poor quality are better discarded and replaced with new questions.

As described above that all research results show from each type of question each item has moderate quality, it cannot be said unilaterally that the end-of-semester summative questions used in Modern Al Rifa Ie Junior High School all the items are not suitable for use to measure student competence. The failure of each item is due to several different factors in each type of question.

With this analysis, it is hoped that it will be able to provide an overview to readers or further researchers that in making questions, they must ensure that the items used are valid, reliable, not too easy and not too difficult, and able to differentiate the abilities of students from upper and lower groups, so as to provide an accurate evaluation of student abilities and support efforts to improve education. Hopefully, this research can encourage motivation for teachers or question makers to be more careful and pay attention and consider the test questions that will be used.

## CONCLUSION

Based on the results and discussion, analysis of the summative questions of the End of the Even Semester of PAI subjects in class VIII of Modern Al Rifa Ie Junior High School Malang in the 2023/2024 academic year on the type of short description questions, there are 3 items of very good quality, 5 items of

good quality, 6 items included in the category of questions that have moderate quality, 6 items of poor quality, and there are no items that have very poor quality. In the type of true-false questions, there are 4 items of good quality, 2 items of moderate quality, 4 items are included in the category of questions that have poor quality, and there are no items that have very good and very poor quality. For the type of long description questions, there are 3 items of moderate quality, 2 items of poor quality, and there are no items of excellent, good and very poor quality.

As described above that all research results show from each type of question each item has moderate quality, it cannot be said unilaterally that the end-of-semester summative questions used in Modern Al Rifa Ie Junior High School all the items are not suitable for use to measure student competence. The failure of each item is due to several different factors in each type of question. With this analysis, it is hoped that it will be able to provide an overview to readers or further researchers that in making questions, they must ensure that the items used are valid, reliable, not too easy and not too difficult, and able to differentiate the abilities of students from upper and lower groups, so as to provide an accurate evaluation of student abilities and support efforts to improve education.

## REFERENCES

- Amrullah. (2018). Analisis Instrumen Tes Soal Pilihan Ganda Ujian Sekolah (US) Mata Pelajaran Al Islam Kelas XII SMA Muhammadiyah se-kota Palembang. *Ad-Man-Pend: Jurnal Administrasi Manajemen Pendidikan*, 1(1), 11–20. <https://doi.org/https://doi.org/10.32502/amp.v1i1.956>
- Anwar, B. (2019). Kompetensi Pedagogik sebagai Agen Pembelajaran. *Shaut Al Arabiyyah*, 6(2), 114–125. <https://doi.org/10.24252/saa.v6i2.7129>
- Arifin, Z. (2014). *Evaluasi Pembelajaran*. Bandung: PT. Remaja Rosdakarya.
- Arikunto, S. (2011). *Prosedur Penelitian Suatu Pendekatan Praktek*. Jakarta: Rineka Cipta.
- Arikunto, S. (2018). *Dasar-Dasar Evaluasi Pendidikan*. Jakarta: PT. Bumi Aksara.
- Astuti, N. D., Hapsan, A., Herianto, & Dkk. (2024). *Prinsip-Prinsip Pengukuran dan Evaluasi Pendidikan*. Sulawesi Selatan: CV. Ruang Tentor.
- Ismail, I., & Dkk. (2020). *Asesmen dan Evaluasi Pembelajaran*. Makasar: Cendekia Publisher.
- Huda, H., Nursyamsiyah, S., & Alfian, M. (2022). The Community-based Character Education: Study of the 'Imaji Academy' Program in Madrasa. *Indonesian Journal of Islamic Education Studies (IJIES)*, 5(1), 113-127.
- Kurniawan, A., Febrianti, A. N., Hardianti, T., & Dkk. (2022). *Evaluasi Pembelajaran*. Padang: PT. Global Eksklusif Teknologi.
- Muluki, A. (2020). Analisis Kualitas Butir Tes Semester Ganjil Mata Pelajaran IPA Kelas IV MI Radhiatul Adawiyah. *Jurnal Ilmiah Sekolah Dasar*, 4(1), 86–96. <https://doi.org/10.23887/jisd.v4i1.23335>
- Nurdin, R. (2020). Analisis Kualitas Soal Ujian Akhir Semester Mata Pelajaran PAI. *Jurnal Pendidikan Aktual*, 6(2), 133–144. <https://repository.ar-raniry.ac.id/id/eprint/16456>
- Nurian, F. (2019). *Metodologi Penelitian Kuantitatif*. Semarang: CV Pilar Nusantara.
- Nurlaila, Kustati, M., & Amelia, R. (2023). Analisis Kualitas Butir-Butir Soal Ulangan Harian Pendidikan Agama Islam Dengan Menggunakan Iteaman. *SOKO GURU: Jurnal Ilmu Pendidikan*, 3(3), 142–155. <https://doi.org/10.55606/sokoguru.v3i3.3069>
- Pangesti, J. A. (2019). Analisis kualitas butir soal penilaian akhir semester 1 mata pelajaran pendidikan agama islam kelas XII di SMA Negeri 2 Kroya Kabupaten Cilacap. In *Institut Agama Islam Negeri*

*Purwokerto.*

- Rizema, P. S. (2013). *Desain Evaluasi Belajar Berbasis Kinerja*. Yogyakarta: DIVA Press.
- Rukajat, A. (2018). *Teknik Evaluasi Pembelajaran*. Yogyakarta: Deepublish.
- Sax, G. (1980). *Principles of Educational and Psychological Measurement and Evaluation* (2nd ed.). Phoenix Publishing Services.
- Septiani, Y., Aribbe, E., & Diansyah, R. (2020). Analisis Kualitas Layanan Sistem Informasi Akademik Universitas Abdurrahman Terhadap Kepuasan Pengguna Menggunakan Metode Sevqual (Studi Kasus : Mahasiswa Universitas Abdurrahman Pekanbaru). *JURNAL TEKNOLOGI DAN OPEN SOURCE*, 3(1), 131–143. <https://doi.org/10.36378/jtos.v3i1.560>
- Sudijono, A. (2003). *Pengantar Evaluasi Pendidikan*. Jakarta: Raja Grafindo.
- Sugiyono. (2017). *Metode Penelitian Kuantitatif, Kualitatif dan Mixed Methods*. Bandung: Alfabeta.
- Supriyadi, G. (2011). *Pengantar Teknik Evaluasi Pembelajaran*. Malang: Intimedia.
- Wildan, W. (2017). Pelaksanaan Penilaian Autentik Aspek Pengetahuan, Sikap dan Keterampilan di Sekolah atau Madrasah. *Jurnal Tatsqif*, 15(2), 131–153. <https://doi.org/10.20414/jtq.v15i2.3>
- Yeni Hartanti. (2021). Pembentukan Karakter melalui Pendidikan Agama Islam. *GUAU Jurnal Pendidikan Profesi Guru Agama Islam*, 1, 335–342. <https://studentjournal.iaincurup.ac.id/index.php/guau/article/view/69>
- Yusuf, R. (2021). Analisis Kualitas Butir Soal Ujian Nasional Mata Pelajaran Matematika Sekolah Menengah Pertama. *Journal of Didactic Mathematics*, 1(3), 158–164. <https://doi.org/10.34007/jdm.v1i3.417>