

Analysis of end-of-odd-semester summative assessment questions based on r studio

Minarti ^{a,1,*}, Abdul Wahab ^{b,2},

^{a,b} Megister Pendidikan Agama Islam Institut Agama Islam Negeri Pontianak

¹ minarti.ja.far@gmail.com; ² abdulwahabassambasi@gmail.com

*Correspondent Author

Received: 02 April 2024

Revised: 10 Mei 2024

Accepted: 20 Juni 2024

KEYWORDS

Analysis;
Assessment;
PAI and BP.

ABSTRACT

This study aims to measure the quality of summative assessment items at the end of the odd semester of Islamic Religious Education and Budi Pekerti subjects for class XI A students in the 2023-2024 academic year at Al Mumtaz Pontianak Integrated Islamic Senior High School. Islamic Religious Education and Ethics learning carried out at the secondary level must be able to transform into quality learning. Quality learning will be achieved if the learning process starting from planning, implementing and evaluating learning is well designed and implemented by educators. It is very important for educators to evaluate learning. This is done so that the achievement of learning outcomes can be measured. This research method uses a descriptive quantitative approach. The subjects in this study were students of class XI A with a population of 19 students of class XI A. The variables of this study were the level of difficulty, differentiating power, effectiveness of triggers and the level of validity of odd ASAS questions. Data collection techniques are carried out with documentation in the form of questions, grids for writing questions, and answer sheets of odd ASAS students in the 2023-2024 academic year. Data analysis techniques using R software. 4.3.1 and R Studio. The results of the analysis show that the quality of the odd semester summative assessment questions shows 88% or 22 questions out of 25 ASAS questions with 17 questions at a good level as much as 68% and 5 questions with a fairly good level as much as 20%, and the results of the ASAS question reliability test with the results of 0.830 are at a good level, so that the question is feasible to be used as an evaluation instrument or feasible to use in the End of Odd Semester Summative Assessment of Islamic Religious Education and Budi Pekerti subjects for class XI A students at SMAIT AL MUMTAZ Pontianak.

This is an open-access article under the [CC-BY-SA](#) license.



Introduction

The paradigm shift in the world of education is marked by changing understandings of education over time (S. Kurniawan, 2019). This shift is indicated by changes in curriculum, methods, technology, and learning media (Rahayu et al., 2022). As educators, the changing times require teachers to innovate in designing learning experiences (S. Kurniawan, 2016), including in the context of evaluation or assessment. Islamic Religious Education and Character Education teachers must be able to transform (Mudarrisuna et al., 2022) in designing learning experiences in these subjects. This transformation ensures that Islamic Religious Education and Character Education can address contemporary challenges in their implementation. Therefore, designing quality assessments and evaluations is crucial (Azizah & Zainudin, 2020).

Islamic Religious Education and Character Education, along with assessments, are an inseparable part of the learning process. Educators and students should understand the competencies to be achieved (Fajri et al., 2021). In the Merdeka Curriculum, these competencies are the learning outcomes to be achieved. As educators, the learning process begins with planning, implementing, and assessing learning (McTighe et al., 2017). Assessment is an activity of proving the achievement of learning outcomes during the learning process (Miftachul Ulum, Abdul Mun'im, Erly Juliyan, 2021).

According to the learning and assessment guidelines from the Ministry of Education and Culture, assessment should be conducted to measure aspects accurately and holistically (McTighe et al., 2017). Assessments can be formative or summative (Direktorat Pembinaan SMA, 2010). Summative assessments are used to achieve all planned learning objectives (T. Kurniawan, 2015). The evaluation instruments for summative assessments must be able to measure the quality of questions according to the students' capabilities or learning outcomes. Theologically, Islamic Religious Education can draw spirit from the Quran, which specifically states that Allah does not burden a soul beyond its capacity, as mentioned in Surah Al-Baqarah/2 verse 286:

Allah does not burden a soul beyond that it can bear. It will have [the consequence of] what [good] it has gained, and it will bear [the consequence of] what [evil] it has earned. "Our Lord, do not impose blame upon us if we forget or make a mistake. Our Lord, and lay not upon us a burden like that which You laid upon those before us. Our Lord, and burden us not with that which we have no ability to bear. And pardon us; and forgive us; and have mercy upon us. You are our protector, so give us victory over the disbelieving people." (Al-Baqarah/2:286)

The Prophet Muhammad (peace be upon him) also demonstrated evaluating according to one's abilities. According to a hadith narrated by Imam Bukhari number 3940:

Muhammad ibn 'Abdullah ibn Numair told us, that his father informed him, who informed them from 'Abdullah, who heard from Nafi', who narrated from ibn 'Umar, who said that the Prophet Muhammad (peace be upon him) tested my combat skills during the Battle of Uhud when I was fourteen years old, but he did not allow me to participate. However, he tested me again during the Battle of the Trench when I was fifteen years old, and this time he permitted me to join. (HR. Muslim)

Therefore, a good evaluation instrument involves a question item analysis. To ensure the evaluation instrument is of quality and standard, educators must prepare test questions well (Castillo, 2023). Item analysis is necessary to ensure quality questions. The primary goal of item analysis is to evaluate the difficulty level, discrimination power, and the effectiveness of distractors or answer choices in each question (Jayali & Sriwahyuni, 2022), (Amalia & Widayati, 2012), (McTighe et al., 2017). This analysis can be conducted using one of two approaches, depending on the chosen test theory, which can be classical test theory (Sumin, Hasan Djidu, Raoda Ismail, Rahmaningtias, 2022), (Trimanansyah, 2014) or modern test theory (Mania et al., 2020).

One of the conditions for quality questions is paying attention to the difficulty level of the questions. Very difficult questions should not be used as they can demotivate and discourage students. Similarly, very easy questions do not stimulate students to think. Good questions match the students' capacity, being neither too difficult nor too easy. The difficulty level is expressed as an index. The following table, based on Iskandar & Rizal (2018), shows the difficulty index:

Table 1. Interpretation Index of Difficulty Level Analysis / Biserial Point (PBis)

PBis Value Interval	Interpretation
$P < 0.000$	Very Difficult
0.000-0.300	Difficult
0.310-0.700	Moderate
0.710-1.000	Easy

One of the main requirements for an effective test instrument is its ability to measure the discrimination power of each test item. This discrimination is necessary for the test to

distinguish between students with high capacity and those with low capacity. The discrimination index is used to assess how effectively a test item differentiates between students with different levels of capacity (Iskandar & Rizal, 2018). The purpose of this discrimination power is to detect the smallest possible individual achievement differences among the students. The discrimination index for test items is presented in the following table:

Table 2. Interpretation Index of Discrimination Power Analysis

Kpbis Value Interval	Interpretation
0.400-1.000	Very Good
0.300-0.390	Good
0.200-0.290	Fairly Good and needs slight revision
-0.000-0.190	Poor

The next requirement to assess the quality of test items is to examine the distractors. Distractors are answer options presented in a test that play a role in identifying students with high capacity. The effectiveness of a distractor can be evaluated based on how often it is chosen by students from the lower-performing group; if it is more frequently chosen by this group, the distractor is considered effective. Conversely, if the distractor is more frequently chosen by high-performing students, it is considered not optimally functional. According to Supranata, a distractor is considered good if it is chosen evenly, but it is considered misleading if it is more often chosen by high-performing students compared to low-performing students (Iskandar & Rizal, 2018). The effectiveness index of distractors in the test is presented in the following table:

Table 3. Interpretation Index of Distractor Analysis / Response Probability (rspP)

rspP Value Interval	Interpretation
76% - 125%	Very Good
126% - 150% or 51% - 75%	Good
151% - 175% or 26% - 50%	Fairly Good
rspP > 200%	Poor

Distractors are considered to function effectively if chosen by at least 5% of all students taking the test (Hamzah, B. U., & Koni, 2012). Furthermore, to assess whether a test item is of good, fairly good, or poor quality, one can examine its validity. The test validity index (Iskandar & Rizal, 2018) is as follows:

Table 4. Interpretation Index of Test Item Validity Analysis

Value Interval	Interpretation
0.60 - 1.00	Good
0.20 - 0.59	Fairly Good
< 0.19	Poor

Next, a question that is considered reliable is one that is consistent and dependable, according to Brown (2007). Here, reliable questions are those that are consistent and dependable. The following table presents the interpretation index of reliable questions:

Table 5. Interpretation Index of Test Item Reliability

Interval Nilai	Interpretasi
0,70 - 1,00	Baik
0,31 - 0,69	Cukup baik
r > 0,30	Tidak baik

Therefore, there are several important aspects that need to be explored regarding the evaluation of test items created by teachers. This study will describe the analysis of test items for the Odd Semester Summative Assessment in the academic year 2023-2024, with the aim of

assessing the quality and suitability of the test item instrument that has been created, based on identified issues and previous research findings. In relation to this, there are several studies, including research conducted by Sitti Mania, Fitriani, Ahmad Farham Majid, Nidya Nina Ichiana, Andi Ika Prasasti Abrar on the analysis of school final exam questions, concluding that 15% of the questions fall into the good category, 67% are categorized as fairly good, and 18% are categorized as less good. Thus, it can be described that these questions meet the criteria to be used as test instruments in school final exams (Mania et al., 2020). Similarly, research by Ata Nayla Amalia and Ani Widayati concluded the following: (1) Based on question validity, 87.5% of Series A questions were considered valid; Series B achieved 95% validity; Series C achieved 75% validity; Series D had a validity of 82.5%; and Series E had a validity of 75%. (2) In terms of reliability, these questions showed high reliability coefficients, with values of 0.833 for Series A, 0.843 for B, 0.80 for Series C, 0.785 for Series D, and 0.768 for Series E. (3) Regarding difficulty level, questions with moderate difficulty levels were 62.5% in Series A, 70% in Series B, 65% in Series C, 52.5% in Series D, and 47.5% in Series E. (4) Good discrimination power was found in questions with percentages of 55% in Series A, 60% in Series B, 57.5% in Series C, 55% in Series D, and 57.5% in Series E. (5) The effectiveness of distractor usage indicated very good quality with percentages of 62.5% in Series A, 37.5% in Series B, 40% in Series C, 50% in Series D, and 35% in Series E (Amalia & Widayati, 2012).

Therefore, it is necessary to conduct an analysis of the test items for the Odd Semester Summative Assessment to measure the quality of test items for the Odd Semester Summative Assessment in the subject of Islamic Education and Character Education for Grade XI students at Al Mumtaz Integrated Islamic High School, Pontianak, for the academic year 2023-2024.

Method

The quantitative approach used in this research is descriptive in nature. It is considered quantitative because the research data consists of numerical values (Sugiono, 2019) and is analyzed using statistical methods (John Creswell, 2015). The subjects of this study are Grade XI B students with a population of 19 students. The variables in this study include difficulty level, discrimination index, effectiveness of distractors, validity, and reliability of test items for the Odd Semester Summative Assessment (ASAS) in the academic year 2023-2024. Data collection techniques involve documentation such as the test item writing guidelines, answer keys, and student answer sheets for the Odd Semester Summative Assessment in the academic year 2023-2024.

Quantitative analysis is conducted to evaluate empirical data derived from 25 multiple-choice questions that were administered in the ASAS. Classical approach methods are used in this research analysis, focusing on aspects such as discrimination power, difficulty level, distractor effectiveness, validity, and reliability. The quantitative analysis is performed using R 4.3.1 and R Studio. R programming language was first introduced by Robert Gentleman and Ross Ihaka in 1992 (Sumin, Hasan Djidu, Raoda Ismail, Rahmaningtias, 2022). R is a powerful programming language widely used in the fields of data science and big data. Initially designed for statistical computation and graphics, its popularity has grown, enabling the implementation of machine learning and deep learning algorithms efficiently (Sumin, Hasan Djidu, Raoda Ismail, Rahmaningtias, 2022). Therefore, descriptive data analysis in this study utilizes R 4.3.1 and R Studio, which are developments of the R software environment.

Results and Discussion

The quantitative approach used in this study is descriptive in nature. It is considered quantitative because the research data consists of numerical data (Sugiono, 2019) and is analyzed using statistics (John Creswell, 2015). The subjects of this study are Grade XI B students with a population of 19 students. The variables in this study are the Difficulty Level, Discrimination Index, Effectiveness of Distractors, Validity, and Reliability of the Odd Semester Summative Assessment items. Data collection techniques include documentation such as the item writing guidelines, answer keys, and students' answer sheets for the Odd Semester

Summative Assessment in the academic year 2023-2024. Quantitative analysis was conducted to evaluate empirical data derived from 25 multiple-choice questions that were administered. A classical approach was used in the analysis, focusing on aspects such as discrimination index, difficulty level, effectiveness of distractors, validity, and reliability. The quantitative analysis in this study was performed using R 4.3.1 and R Studio.

The R programming language was initially introduced by Robert Gentleman and Ross Ihaka in 1992 (Sumin, Hasan Djidu, Raoda Ismail, Rahmaningtias, 2022). R is a powerful programming language widely used in data science and big data. Originally designed for statistical computing and graphics, its extensive popularity has led to its adaptation for machine learning and deep learning algorithms in a fast and efficient manner (Sumin, Hasan Djidu, Raoda Ismail, Rahmaningtias, 2022). Therefore, descriptive data analysis in this study is conducted using the R 4.3.1 and R Studio software, which are advancements of the R software environment.

Results of the Analysis of Difficulty Levels of Odd Semester Summative Assessment Items

The difficulty level of the Odd Semester Summative Assessment items represents the opportunity for students to answer correctly on each item, whether they are categorized as Very Difficult, Difficult, Moderate, or Easy.

Table 6. Difficulty Levels and Interpretation of Odd Semester Summative Assessment

Items		
Item Number	Point Biserial (pBis) / Difficulty Level	Interpretation
Soal_1	0.545	Moderate
Soal_2	0.295	Difficult
Soal_3	0.268	Difficult
Soal_4	-0.167	Very Difficult
Soal_5	0.431	Moderate
Soal_6	-0.244	Very Difficult
Soal_7	0.403	Moderate
Soal_8	0.533	Moderate
Soal_9	0.517	Moderate
Soal_10	0.228	Difficult
Soal_11	0.427	Moderate
Soal_12	0.479	Moderate
Soal_13	0.361	Moderate
Soal_14	0.643	Moderate
Soal_15	0.427	Moderate
Soal_16	-0.355	Very Difficult
Soal_17	0.112	Difficult
Soal_18	0.513	Moderate
Soal_19	0.633	Moderate
Soal_20	0.583	Moderate
Soal_21	0.596	Moderate
Soal_22	0.503	Moderate
Soal_23	0.715	Easy
Soal_24	0.358	Moderate
Soal_25	0.179	Difficult

Table 7. Analysis Results of Difficulty Level of ASAS Test Items

No.	Category	Test Items	Count
1.	Very Difficult	4, 6, 16	3

2.	Difficult	2, 3, 10, 17, 25	5
3.	Moderate	1, 5, 7, 8, 9, 11, 12, 13, 14, 15, 18, 19, 20, 21, 22, 24	16
4.	Easy	23	1

Figure A.1. Percentage of Difficulty Levels of ASAS Test Items



The description from Tables 6 and 7, and further clarification from Figure 1, reveal that 64% of ASAS test items are categorized as moderate difficulty, 20% as difficult, 12% as very difficult, and 4% as easy. These findings are consistent with a study by Sitti Mania, Fitriani, Ahmad Farham Majid, Nidya Nina Ichiana, and Andi Ika Prasasti Abrar, which found that out of 40 test items, 28% were of moderate difficulty, 22% were difficult, 20% were very difficult, 20% were easy, and 10% were very easy (Mania et al., 2020). Based on the difficulty index table, the ASAS test includes: 16 items (64%) categorized as moderate difficulty, specifically items 1, 5, 7, 8, 9, 11, 12, 13, 14, 15, 18, 19, 20, 21, 22, and 24. 5 items categorized as difficult, specifically items 2, 3, 10, 17, and 25. 3 items categorized as very difficult, specifically items 4, 6, and 16. 1 item categorized as easy, specifically item 23.

From the above presentation, items categorized as very difficult should ideally be avoided as they can diminish motivation and enthusiasm for tackling the questions. Similarly, easy items may not sufficiently stimulate students' thinking. A good question is one that aligns with students' abilities, not too difficult yet not too easy (Trimanansyah, 2014). Therefore, questions categorized as moderate difficulty are considered effective as they can accurately measure students' understanding of the learning material and their ability to achieve the intended learning outcomes. Hence, it can be concluded that the above questions are good questions because there are more moderate difficulty questions compared to very difficult, difficult, and easy questions.

Hasil Analisis Daya Pembeda Soal Asesmen Sumatif Akhir Semester Ganjil

The discrimination index of an item aims to describe its capacity or identify the difference between students with high and low abilities (Ramdani, 2012). Here are the results of the discrimination analysis using R 4.3.1 and R Studio.

Table 8. Discrimination Index and Interpretation of ASAS Items

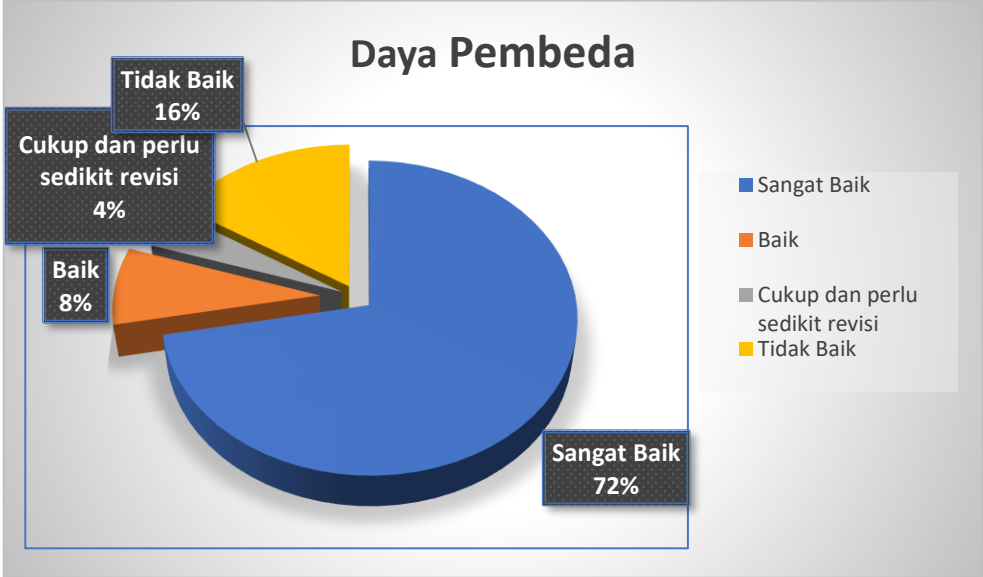
Item Number	Discrimination/Point Biserial Correlation/Discrimination Power	Interpretation
Soal_1	0.698	Very Good
Soal_2	0.417	Very Good
Soal_3	0.350	Good
Soal_4	-0.236	Not Good

Soal_5	0.551	Very Good
Soal_6	-0.508	Not Good
Soal_7	0.570	Very Good
Soal_8	0.697	Very Good
Soal_9	0.649	Very Good
Soal_10	0.308	Good
Soal_11	0.539	Very Good
Soal_12	0.647	Very Good
Soal_13	0.462	Very Good
Soal_14	0.807	Very Good
Soal_15	0.546	Very Good
Soal_16	-0.454	Not Good
Soal_17	0.159	Not Good
Soal_18	0.776	Very Good
Soal_19	0.827	Very Good
Soal_20	0.762	Very Good
Soal_21	0.747	Very Good
Soal_22	0.631	Very Good
Soal_23	0.896	Very Good
Soal_24	0.449	Very Good
Soal_25	0.229	Adequate, needs slight revision

Table 9. Analysis of Discrimination Index of ASAS Questions by Category, Item, and Number of Questions

No	Katagori	Butir Soal	Jumlah
1.	Sangat Baik	1,2,5,7,8,9,11,12,13,14,15,18,19,20,21,22,23,24	18
2.	Baik	3,10,	2
3.	Cukup dan perlu sedikit revisi	25	1
4.	Tidak Baik	4, 6, 16,17	4

Gambar A.2. Persentase Daya Beda Soal ASAS



Description of Tables 8 and 9 and Clarification Based on Figure 1. The tables and Figure 1 provide a detailed analysis of the discriminative power (daya beda) of ASAS (Odd Semester Summative Assessment) items. Here’s a breakdown of the findings: Table 8: Discriminative

Power and Interpretation of ASAS Items. This table categorizes ASAS items based on their discriminative power: Very Good: 18 items (72%) demonstrate strong discriminative power, including items numbered 1, 2, 5, 7, 8, 9, 11, 12, 13, 14, 15, 18, 19, 20, 21, 22, 23, and 24. Good: 2 items (8%) show moderate discriminative power, specifically items numbered 3 and 10.

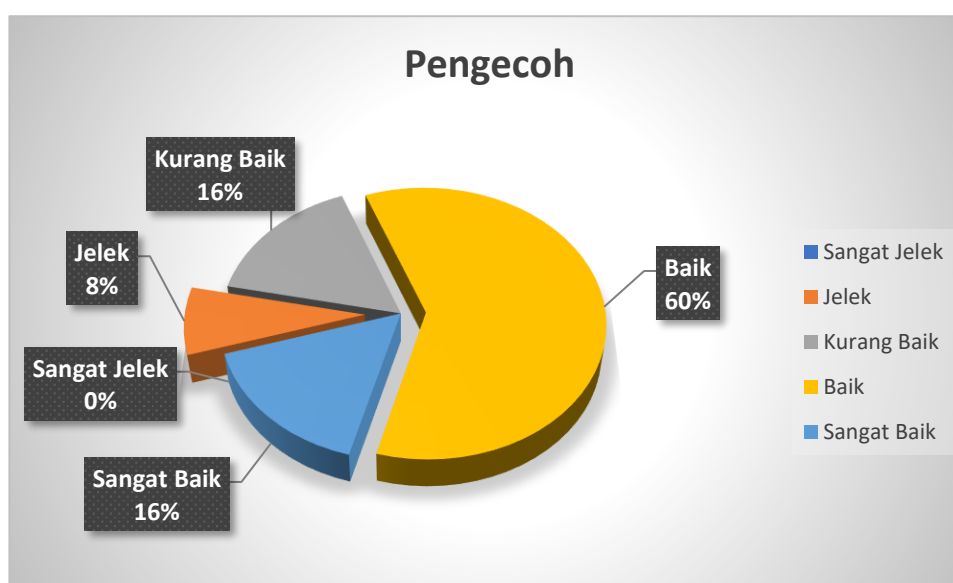
Fair and Needs Some Revision: 1 item (4%) requires slight revision for improved effectiveness, identified as item number 25. Not Good: 4 items (16%) have low discriminative power and are not recommended for reuse without substantial revision, listed as items numbered 4, 6, 16, and 17.

Table 9: Analysis of Discriminative Power of ASAS Items by Category, Item, and Total This table further classifies ASAS items according to their discriminative power categories: Very Good: 18 items (No. 1, 2, 5, 7, 8, 9, 11, 12, 13, 14, 15, 18, 19, 20, 21, 22, 23, 24) are highly effective in distinguishing between students who understand the material and those who do not. Good: 2 items (No. 3 and 10) exhibit acceptable discriminative power but may benefit from minor adjustments.

Fair and Needs Some Revision: 1 item (No. 25) shows moderate discriminative power and requires revision before reuse. Not Good: 4 items (No. 4, 6, 16, 17) demonstrate poor discriminative power and should be redeveloped before use in assessments. Clarification Based on Figure 1. Figure 1 highlights that: 72% of ASAS items are categorized as "Very Good" in terms of their discriminative power. 16% of items are categorized as "Not Good," indicating they require substantial improvement. 8% of items fall into the "Good" category, suggesting they have moderate effectiveness. 4% of items need slight revision to enhance their effectiveness.

These findings align with previous research conducted by Sitti Mania et al. (2020) in mathematics at SMAN 1 Pitumpanua, which also emphasized the importance of item quality based on discriminative power. Items categorized as "Very Good" are deemed suitable for inclusion in the question bank for Islamic Education and Character Education teachers at SMAIT AL MUMTAZ Pontianak. Items categorized as "Fair and Needs Some Revision" are identified for potential improvement, while those categorized as "Not Good" require redevelopment to ensure their effectiveness in assessing student comprehension. This comprehensive analysis ensures that ASAS items effectively measure student understanding and contribute positively to achieving educational goals in Islamic Education and Character Education at the school.

The analysis that can be evaluated from this research includes characteristics used to assess the quality of ASAS (Summative Assessment at the End of Odd Semester) items.

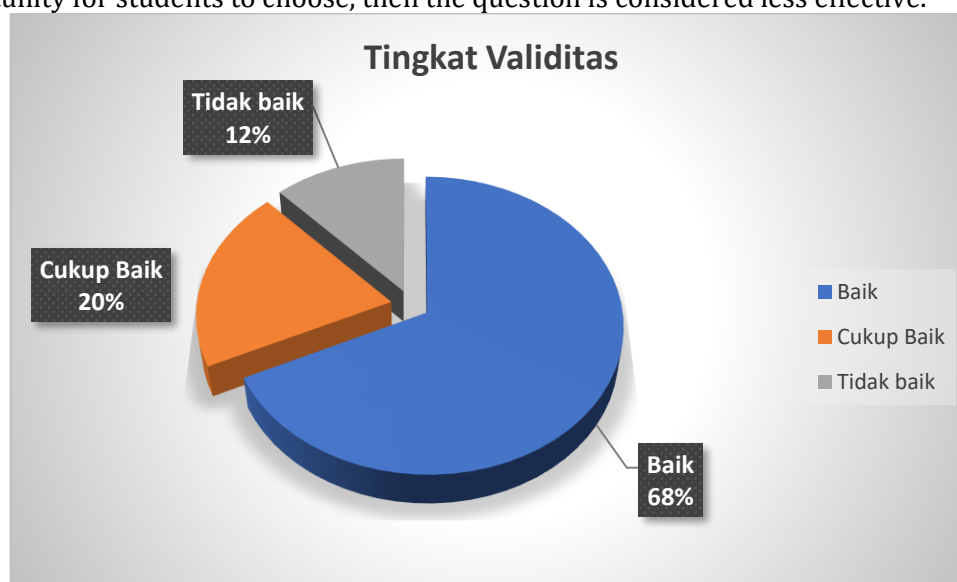


Based on tables 10 and 11, and figure 3 depicting the effectiveness percentage of distractors in ASAS questions, it can be described that 15 questions, accounting for 60% with a

good category, 4 questions, accounting for 16% with a very good category, 4 questions, accounting for 16% with a less good category, 2 questions, accounting for 8% with a poor category, and the very poor category has 0% percentage.

Based on the assessment of the effectiveness of the ASAS questions above, there are 4 questions categorized as very good with question numbers 4, 6, 17, and 18. Then, there are 15 questions categorized as good with question numbers 1, 3, 8, 9, 10, 11, 12, 14, 15, 16, 19, 20, 21, 22, and 23. Therefore, questions categorized as very good and good are considered effective distractor options for students to choose. Hence, these questions are included in the question bank inventory by PAI and BP teachers at SMAIT AL MUMTAZ Pontianak and can be used for subsequent tests. Questions categorized as less good include 4 questions with question numbers 5, 13, 24, and 25, which should be revised in terms of distractor options. For questions categorized as poor, there are 2 questions with question numbers 2 and 7, which should not be reused in subsequent tests.

Based on the above presentation, it can be concluded that questions are considered good if distractors effectively play their role, meaning distractor options provide equal opportunity for students to choose. Conversely, if the distractor options do not provide equal opportunity for students to choose, then the question is considered less effective.



Based on Table 12 and Figure 4, it can be described that the validity levels of ASAS questions categorized as good consist of 17 questions, accounting for 68%, while questions categorized as fairly good consist of 5 questions, totaling 20%. Additionally, questions categorized as not good amount to 3 questions, comprising 12%. The 17 questions categorized as good are identified with item numbers 1, 5, 7, 8, 9, 11, 12, 13, 14, 15, 18, 19, 20, 21, 22, 23, and 24. Meanwhile, the 5 questions categorized as fairly good include item numbers 2, 3, 10, 17, and 25. The questions categorized as not good consist of 3 questions with item numbers 4, 6, and 17.

From the gathered data, it is evident that the quality of item questions demonstrates a significant aspect in the process of constructing test instruments. Therefore, an analysis of the quality of item questions is necessary to assess the suitability of item questions to be used in tests, particularly in the context of this research, which is the Odd Semester Summative Assessment in Islamic Education and Character Education for Grade IX A students at SMAIT AL MUMTAZ Pontianak.

Results of ASAS Item Reliability Analysis. The analysis results of item reliability will be presented in the following table: [Please provide details if you need assistance in formulating this table.]

Tabel 13. Hasil Analisis Reliabel Soal ASAS

Coefficient Alpha
0,830

Based on Table 13 above, it can be interpreted that the ASAS questions that were tested and analyzed obtained a result of 0.830, indicating that these ASAS questions are at a good level.

Conclusion

Based on the results and discussion of this study, it can be concluded that the analysis of the Odd Semester Summative Assessment (ASAS) items in Islamic Education and Character Education for Grade XI students at SMAIT AL MUMTAZ Pontianak Academic Year 2023-2024 shows good quality and suitability as evaluation instruments. The findings from the analysis using R 4.3.1 and R Studio are summarized as follows:

Firstly, regarding the difficulty level of ASAS items, the results indicate that 64% of ASAS items fall into the moderate category, 20% are categorized as difficult, 12% as very difficult, and 4% as easy. Thus, it can be concluded that these items are of good quality because the moderate-level items are more prevalent compared to those categorized as very difficult, difficult, and easy.

Secondly, the discriminative power analysis reveals that 18 items (72%) are categorized as "Very Good," 4 items (16%) as "Not Good," 2 items (8%) as "Good," and 1 item (4%) as "Fair and Needs Some Revision." Items categorized as "Very Good" and "Good" are considered suitable for inclusion in the question bank used by Islamic Education and Character Education teachers at SMAIT AL MUMTAZ Pontianak, while those categorized as "Not Good" or needing revision are not recommended for reuse.

Thirdly, the effectiveness of distractors analysis shows that 15 items (60%) are categorized as "Good," 4 items (16%) as "Very Good," 4 items (16%) as "Fair," 2 items (8%) as "Poor," and no items are categorized as "Very Poor." Therefore, items categorized as "Very Good" and "Good" provide viable options for distractors and are included in the question bank inventory used by teachers, whereas items categorized as "Poor" or lower should not be reused in subsequent tests.

Fourthly, the validation level analysis indicates that 88% (22 out of 25) of ASAS items are categorized as "Good," comprising 68% as "Good" and 20% as "Fair." Hence, these items are deemed suitable for use as evaluation instruments in the Odd Semester Summative Assessment for Islamic Education and Character Education for Grade XI students at SMAIT AL MUMTAZ Pontianak. Fifthly, the reliability analysis of ASAS items yields a result of 0.830, which falls within the "Good" level. These findings collectively support the conclusion that the ASAS items analyzed in this study are of good quality, demonstrating suitability for evaluation purposes in the educational context specified.

References

- Amalia, A. N., & Widayati, A. (2012). Analisis Butir Soal Tes Kendali Mutu Kelas Xii Sma Mata Pelajaran Ekonomi Akuntansi Di Kota Yogyakarta Tahun 2012. *Jurnal Pendidikan Akuntansi Indonesia*, 10(1). <https://doi.org/10.21831/jpai.v10i1.919>
- Azizah, N., & Zainudin, M. (2020). Evaluasi Pembelajaran Pendidikan Agama Islam (Studi di SMK Muhammadiyah 1 dan SMK Muhammadiyah 2 Kota Palembang). ... *Journal: Pendidikan Agama Islam*. <https://jurnal.staibslg.ac.id/index.php/ej/article/view/139>
- Castillo, R. (2023). *Jurnal Indonesia Pendidikan Sejarah*. 6641(September).
- Direktorat Pembinaan SMA. (2010). *Panduan Analisis Butir Soal*. 1, 48.
- Fajri, N., Akbar, A., & Zakir, S. (2021). Penerapan Aplikasi Kahoot Dalam Evaluasi Pembelajaran PAI Pada Materi Masuknya Islam Ke Nusantara Di Kelas IX Di SMP N 1 Banuhampu Kabupaten Agam. *Jurnal Kajian Dan Pengembangan* <https://www.jurnal.umsb.ac.id/index.php/ummatanwasathan/article/view/3040>
- Hamzah, B. U., & Koni, S. (2012). *Assessment pembelajaran*. Bumi Aksara.
- Hadisi, L., Hakim, M. R., Musthan, Z., Nashihin, H., & Kendari, I. (2023). Implementation of Learning Management In Building an Attitude of Religious Tolerance at State High Schools In The Muna

- District. Edukasi Islami: Jurnal Pendidikan Islam, c, 1879–1892. <https://doi.org/10.30868/ei.v12i03.4296>
- Husna Nashihin, Yenny Aulia Rachman, Betania Kartika, Nurmasinta Fadhilah, T. H. (2023). Pendidikan TPQ Kontra Radikalisme berhaluan Aswaja (M. D. Yahya (ed.)). Academia Publication. <https://academiapublication.com/index.php/product/pendidikan-tpq-kontra-radikalisme-berhaluan-aswaja>
- Iskandar, A., & Rizal, M. (2018). Analisis kualitas soal di perguruan tinggi berbasis aplikasi TAP. *Jurnal Penelitian dan Evaluasi Pendidikan*, 22(1), 12–23. <https://doi.org/10.21831/pep.v22i1.15609>
- Jayali, A. M., & Sriwahyuni, E. (2022). Analisis Butir Soal Instrumen Tes Keterampilan Proses Sains pada Topik Sistem Periodik Unsur. *Journal on Teacher Education*, 4(2), 1581–1591. <http://journal.universitaspahlawan.ac.id/index.php/jote/article/view/10800%0Ahttp://journal.universitaspahlawan.ac.id/index.php/jote/article/download/10800/8354>
- Jatmiko Wibisono, Hafidz, Iffat Abdul Ghalib, H. N. (2023). Konsep Pemikiran Pembaharuan Muhammadiyah Bidang Pendidikan (Studi Pemikiran Muhammad Abduh). *Attractive : Innovative Education Journal*, 5(2), 514–522. <http://dx.doi.org/10.51278/aj.v5i2.772>
- Julkifli. (2022). Kedudukan dan Tugas Manusia dalam Perspektif Tafsir al-Qur'an Zubdatu At- Tafsir Karya Muhammad Sulaiman Abdullah Al Asyqar. *Amorti: Jurnal Studi Islam Interdisipliner*, 1(2), 103–110. <https://doi.org/10.59944/amorti.v1i2.24>
- John Creswell. (2015). *Riset Pendidikan Perencanaan, Pelaksanaan, dan Evaluasi Riset Kualitatif dan Kuantitatif* (5 ed.). Pustaka Pelajar.
- Kurniawan, S. (2016). *ILMU PENDIDIKAN ISLAM sebuah Kajian Komprehensif*.
- Kurniawan, S. (2019). Tantangan Abad 21 bagi Madrasah di Indonesia. *Intizar*. <http://jurnal.radenfatah.ac.id/index.php/intizar/article/view/3242>
- Kurniawan, T. (2015). Analisis Butir Soal Ulangan Akhir Semester Gasal Mata Pelajaran IPS Sekolah Dasar (Analysis of Odd Semester Final Test Items in Elementary School of Social Studies Subjects). *Journal of Elementary Education*, 4(1), 1–6. <https://journal.unnes.ac.id/sju/index.php/jee/article/view/7488>
- Kholis, N., & Yunita, V. (2023). Interpretation of Hadith-Sunnah from Schacht ' s Perspective : Multidisciplinary Study. *Amorti: Jurnal Studi Islam Interdisipliner*, 2(3), 148–159. <https://doi.org/10.59944/amorti.v2i3.100>
- Kholish, A. (2022). Pendidikan Multikultural di Pondok Pesantren Tradisional dan Modern sebagai Upaya Menjaga Negara Kesatuan Republik Indonesia (Studi di Pondok Pesantren Asy Syamsuriyah Brebes). *JIPSI: Jurnal Ilmu Pendidikan Dan Sains Islam Interdisipliner*, 1(1), 1–12. <https://doi.org/10.59944/jipsi.v1i1.1>
- Kholish, A., Hidayatullah, S., & Nashihin, H. (2020). Character Education of Elderly Students Based on Pasan Tradition at Sepuh Islamic Boarding Shool Magelang. *IJECA (International Journal of Education and Curriculum Application)*, 3(1), 48. <https://doi.org/10.31764/ijeca.v3i1.2061>
- Mania, S., Fitriani, F., Majid, A. F., Ichiana, N. N., & Abrar, A. I. P. (2020). Analisis Butir Soal Ujian Akhir Sekolah. *Al asma : Journal of Islamic Education*, 2(2), 274. <https://doi.org/10.24252/asma.v2i2.16569>
- McTighe, J., Wiggins, G., Warso, A. W. D. D., Zahroh, S. H., Parno, Mufti, N., & Anggraena, Y. (2017). Pembelajaran dan Penilaian. *Seminar Pendidikan IPA Pascasarjana UM*, 123.
- Miftachul Ulumm, Abdul Mun'im, Erly Juliyan, P. S. (2021). Evaluasi Pembelajaran Ujian Akhir Semester Mata Pelajaran Bisnis Online Kelas XII SMK Sunan Drajat Lamongan. *Paper Knowledge . Toward a Media History of Documents*, 3(2), 6.
- Mudarrisuna, J., Kajian, M., Agama, P., Vol, I., Uin, U., Utara, S., & Email, M. (2022). *REKONSTRUKSI PENDIDIKAN ISLAM KONTEMPORER PENDAHULUAN Pendidikan berintikan interaksi antar manusia , terutama antara pendidik dan terdidik demi mencapai tujuan pendidikan secara nasional Nasional . Pendidikan Islam memiliki peran yang sangat penting dala*. 10(1), 43–56.
- Nashihin, H. (2017). Pendidikan Karakter Berbasis Budaya Pesantren. *Formaci*. <https://books.google.co.id/books?id=X27IDwAAQBAJ>
- Nashihin, H. (2019a). Analisis Wacana Kebijakan Pendidikan (Konsep dan Implementasi). CV. Pilar Nusantara. <https://books.google.co.id/books?id=SXcqEAAAQBAJ>
- Nashihin, H. (2019b). تطبيق طريقة مباشرة في تعليم اللغة العربية في معهد الشيخ ابن باز يوجياكرتا. *Proceedings of 2nd International Conference on ASIC*, 275–281. <https://doi.org/10.1109/icasic.1996.562734>
- Nashihin, H. (2022). Konstruksi Pendidikan Pesantren berbasis Tasawuf-Ecospiritualism. *Edukasi Islami : Jurnal Pendidikan Islam*, 11(01), 1163–1176. <https://doi.org/10.30868/ei.v11i01.2794>
- Nashihin, H., Nur Iftitah, M., & Mulyani, D. (2019). Menepis Tuduhan Islam Misoginis Melalui Pengkajian Pendidikan Islam Perspektif Gender. *At Turots: Jurnal Pendidikan Islam*, 1(1).

- <https://doi.org/10.51468/jpi.v1i1.5>
- Nashihin, H., Yahya, M. D., & Aziz, N. (2020). Morality and Expression of Religious Moderation in "Pecinan". *Dalimunthe* 2016, 24158–24168. <https://doi.org/10.33258/birci.v5i3>
- Rahayu, R., Iskandar, S., & Abidin, Y. (2022). Inovasi Pembelajaran Abad 21 dan Penerapannya di Indonesia. *Jurnal Basicedu*, 6(2), 2099–2104. <https://doi.org/10.31004/basicedu.v6i2.2082>
- Ramdani, Y. (2012). Pengembangan Instrumen dan Bahan Ajar untuk Meningkatkan Kemampuan Komunikasi, Penalaran, dan Koneksi Matematis dalam Konsep Integral. *Jurnal Penelitian Pendidikan*, 13(1), 44–52. http://jurnal.upi.edu/file/6-yani_ramdhana-edl.pdf
- Sugiono. (2019). *Metode Penelitian Kuantitatif, Kualitatif dan R&D* (2 ed.). Alfabeta.
- Sumin, Hasan Djidu, Raoda Ismail, Rahmaningtias, I. (2022). *Analisis Instrumen Penelitian dengan Teori Tes Klasik dan Moderen Menggunakan Program R* (H. R. Syamsul Hadi (ed.)). UNY Perss.
- Trimanansyah, S. (2014). *Evaluasi Pembelajaran*. Pustaka Setya.