J-PAI : Jurnal Pendidikan Agama Islam Vol. 10 No. 1 Juli - Desember 2023
P-ISSN 2355-8237 | e-ISSN 2503-300X http://ejournal.uin-malang.ac.id/index.php/jpai
DOI : 10.18860/jpai.v10i1.25892

Analytical Review of Final Examination Questions for Fifth Grade at SDN 03 East Pontianak

[1]Sri Herlina; [2]Sahrani; [3]Moh. Ainin; [4]Muhamad Tisna Nugraha
[1, 2, 4] Institut Agama Islam Negeri (IAIN) Pontianak, Indonesia
[3]Universitas Negeri Malang, Indonesia
[1]sriherlina779@gmail.com; [2]sahrani176@gmail.com; [3]moh.ainin.fs@um.ac.id;
[4]tisnanugraha2014@yahoo.com

Abstract. *This study aims to evaluate the validity, reliability, level of difficulty, differentiating power, and effectiveness of checkers on the final test of the odd semester of PABP subject in class V SDN 03 East Pontianak in the 2023/2024 academic year. Using a quantitative approach, this study analysed data using IBM SPSS Statistics 23. This study involved 30 class V students as participants. The results of the analysis of 30 multiple choice items showed that 73% were valid and 27% were invalid. From the reliability aspect, it was found that the reliability level of the items was high, namely the r coefficient of 0.899. From the aspect of difficulty level, it was found that about 13% of the items were difficult, about 73% of the items were moderate and 13% of the items were easy. From the aspect of differential power, it was found that about 39% of items were good, about 42% of items were sufficient, about 19% of items were not good. And from the aspect of eliminators, it was found that about 89% were accepted, about 8% were revised and about 3% were rejected.*

Keywords*. Analytical Review; Final Examination Questions*

Abstrak. Penelitian ini bertujuan untuk mengevaluasi validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan efektivitas pengecoh soal ulangan akhir semester ganjil mata pelajaran PABP kelas V SDN 03 Pontianak Timur tahun pelajaran 2023/2024.Dengan menggunakan pendekatan kuantitatif, penelitian ini menganalisis data dengan menggunakan IBM SPSS Statistics 23. Penelitian ini melibatkan 30 siswa kelas V sebagai partisipan. Hasil analisis terhadap 30 butir soal pilihan ganda menunjukkan bahwa 73% valid dan 27% tidak valid. Dari aspek reliabilitas ditemukan bahwa tingkat reliablitas butir soal termasuk tinggi yaitu *r* koefisien sebesar 0,899. Dari aspek Tingkat kesukaran ditemukan bahwa sekitar 13% butir soal yang sukar, sekitar 73% butir soal yang sedang dan 13% butir soal yang mudah. Dari aspek daya beda ditemukan bahwa sekitar 39% butir soal yang baik, sekitar 42% butir soal yang cukup, sekitar 19% butir soal yang tidak baik. Dan dari aspek pengecoh ditemukan sekitar 89% diterima, sekitar 8% direvisi dan sekitar 3% ditolak.

Kata Kunci. *Telaah Analitik; Soal Ujian Akhir Sekolah*

A. INTRODUCTION

Education is an effort that is carried out consciously and well planned and full of awareness that must be followed by students so that they can be educated and guided effectively so that changes in knowledge, behavior and skills can be seen through experienced experiences, for the good of the world and the hereafter (Mastiah, 2019). So that education is the most important part needed in human life (Khoiriyah et al., 2023). Imam Al Ghazali said the purpose of education is to erode bad morals by instilling good morals. Progressive changes in human behavior become a systematic and well-organized element of (Purnomo & Solikhah, 2021). In the education process, tests and measurements are included as a way of evaluating. This is a factor that really needs to be considered because the results of the evaluation can determine various objectives in making decisions. Regarding the field of education, Gronlund (1985) explains that evaluation specifically aims to determine the extent to which students have mastered previously set learning objectives and diagnose student learning difficulties. In order for the information obtained to be a true picture of students' abilities, measurement instruments and procedures for implementing measurements are needed that can obtain results that are guided by high objectivity because we often find that measurements and decision making contain subjectivity because the evaluation process is an activity consisting of complex activities (Matondang, 2009).

Gronlund and Linn (1985) argue that assessment is interpreted as a systematic process of collecting, analyzing and interpreting information to determine the extent to which students achieve learning objectives (Arifianto et al., 2021). The same thing was also expressed by Rina Febriana (2019: 5) that assessment is a series of activities to obtain, analyze and interpret data about the process and results of student learning which is carried out systematically and continuously so that it becomes meaningful information in making decisions (Febriana, 2019). This assessment activity should be carried out based on the principle that the assessment given regarding student learning outcomes is intended as corrective feedback for both students and teachers. Because assessment is an activity that takes place systematically and continuously in order to obtain information about the process or results of student learning, in order to determine decisions based on previously determined criteria (Tausih & Marno, 2021). With the results obtained, it can increase students' interest and motivation to get even better learning results. Without evaluation, it is impossible to know the needs and abilities of students. That is why assessment and evaluation are needed to determine whether the learning carried out has achieved educational goals.

Assessment and evaluation in the classroom can assess the cognitive domain because the cognitive domain is very important to determine the ability of students in terms of knowledge. This cognitive domain assessment aims to determine the extent to which the learning process can be successful and the extent of students' understanding of the material that has been taught by the teacher over a period of time. In the assessment and to evaluate, the teacher needs a tool. Evaluation tools in education used to collect data can be tests or non-tests. Learning outcomes evaluation tools, test techniques are the techniques most often used by teachers to conduct cognitive domain assessments so that it can be seen to what extent students can master and understand the material. With this test, teachers usually conduct evaluations.

Evaluation is an important component and stage that must be taken by teachers to determine the effectiveness of learning. The results obtained can be used as feed-back for teachers in improving and perfecting learning programs and activities (Dayu & Suharmon, 2021). The evaluation carried out can run optimally and hit the target if the measurements

taken are in accordance with the competencies to be measured. Therefore, teachers must know the steps of preparing tests in the cognitive domain according to the material that has been given to students. The first step in preparing a cognitive test is to compile a grid of material to be tested. In this case, the teacher must make a summary outline of the material that has been taught to students. This is done so that the test instrument given can cover important material from the lessons that have been done before so that it can actually measure what should be measured. With a grid, students will be fully prepared for the test because they can narrow the learning area so that their abilities can be truly measured (Sridadi et al., 2020).

After giving the test, the teacher is required to conduct item analysis. Item analysis is very important in the world of education, item analysis is carried out to determine the level of quality of questions used as an assessment of the development of student learning outcomes. Because quality questions are questions that can provide precise information about students who master the material and who have not mastered the material (Masitoh, 2022). Arikunto (2013) also argues that analyzing the quality of items is a procedure that will systematically be able to provide very specific information about the test items we compile. The items commonly used in school exams are multiple choice questions. Multiple choice questions are questions that have several alternative answers. In general, multiplechoice questions have alternative answers ranging from 3 (three) to 4 (four) choices, if the alternative answers to multiple choice questions are more than that, it will make it difficult to prepare the items, and confuse the students.

The purpose of item analysis is also to help improve tests through revision or discarding ineffective questions, as well as to find out diagnostic information on students whether they have understood the material that has been taught or vice versa. The results of the analysis will show how good the quality of a question is. A quality question can be known from the validity, reliability, difficulty level of the question, differentiating power, and the effectiveness of the question examiner. Through the results of the analysis, the teacher can evaluate the question. Item analysis is carried out as a follow-up to improving the quality of a test. If the quality of the items is good, the evaluation will run optimally. Evaluation that runs optimally will make the teaching and learning process better and improve student abilities from year to year. Anastasi and Urbin in Kusaeri and Suprananto (2012) state that item analysis activities have many benefits, including: 1) can help test users in evaluating the quality of the tests used; 2) relevant for the preparation of informal tests such as tests prepared by teachers for students in class; 3) support effective item writing; 4) materially can improve classroom tests; 5) increase the validity and reliability of questions.

The results of research conducted by Rouf Arokhmah (2017), showed that the items of the Second Midterm Examination of Islamic Religious Education Class VI SDIT Mutiara Hati Purwareja Klampok Banjarnegara in the 2016/2017 academic year, the quality of the items can be said to be quite good because of the validity of 78% valid and 20% invalid. The reliability is high having a coefficient of 0.88. The level of difficulty of the question 50% is categorized as very easy, 16% is categorized as easy, 32% is categorized as medium, 0% is categorized as difficult, and 2% is categorized as very difficult. Distinguishing power 28% categorized as very good, 20% categorized as quite good, 12% categorized as poor, 40% categorized as not good. The effectiveness of the function of the exceptions; 15% of the exceptions function very well, 29% of the exceptions function well, 19% of the exceptions function poorly, 23% of the exceptions function poorly, and 14% of the exceptions function very poorly.

Given that item analysis is very important, but in reality there are still schools that analyze the items only up to content validity, and not up to the trial and item analysis in

terms of validity, reliability, level of difficulty, distinguishing power, and the effectiveness of examiners. This was also the case for teachers in East Pontianak. Teachers conduct item analysis only to see how many levels of achievement of the learning objectives they teach each item. For this reason, researchers are interested in analyzing the items of the odd semester final test prepared at the sub-district level. Researchers are eager to see the results of item analysis in terms of validity, reliability, difficulty level, differentiation and effectiveness of the checker. However, this research was specifically conducted at SDN 03 East Pontianak. The aim is that teachers at the school can do the same for other subjects. Then they can understand that making good questions is not only because teachers are good at making questions, but students can answer these questions well.

Based on this background, researchers are interested in conducting research on "Item Analysis of Odd Semester End Summative Questions (SAS) of PABP Class V SDN 03 East Pontianak in the 2023/2024 School Year". With the hope that it can provide an overview and further action so that in the future the teacher can improve the quality of the next question.

B. METHODS

This research uses a quantitative approach. The data source is the odd semester final summative question paper (SAS) of PABP class V at SDN 03 East Pontianak. Data collection was carried out through documentation in the form of question grids, answer keys, and answer sheets of 30 fifth grade students. The data collected was then analyzed using IBM SPPS statistics 23, including to describe the level of validity, reliability, level of difficulty, differentiating power and effectiveness of checkers The following are the criteria for identifying the level of difficulty of question items (Mania et al., 2020). Because in these schools the teacher has not implemented specific item analysis. This research aims to provide an overview and further action so that in the future teachers can improve the quality of questions.

The following are the identification criteria for each aspect. Validity or validity means the extent of the accuracy and accuracy of a measuring instrument in performing its measurement function. The following are the criteria for the value of item validity. The validity of this criterion is more related to comparing the test results of a test with the test results of a standardized test that is already considered valid.

Table B. 1 Validation Value Criteria

No.	Interpretation of validity value	
1.	0,800-1,00	Very High
2.	0,600-0,799	High
3.	0,400-0,599	Fair
4.	0,200-0,399	Low
5.	0,000-0,199	Very Low

A reliable test is a test that is consistent, not indecisive or doubtful *(adamu at-tadzabdzub)*. This definition is also reinforced by Brown (2007), that a reliable test is a consistent and reliable test. Specifically, consistency refers to the constancy of the score. that is, if the test is tested more than once (for example twice) on the same subject at different times.

Table B. 2 Criteria for Reliability Value

No.	Interpretation of Reliability Value	
1.	0,90 - 1,00	Very High
2.	0,70 - 0,89	High
3.	0,50 - 0,69	Fair
4.	0,30 - 0,49	Low
5.	less than 0,30	Very Low

The level of difficulty of school final exam questions is the chance for students to answer correctly on each question item that has a medium, difficult or very difficult level of difficulty.

Table B. 3 Criteria for Level of Difficulty

No.	Interpretation of Level of Difficulty	
1.	0,00 – 0,30	Difficult
2.	0,31 – 0,70	Medium
3.	0,71 – 1,00	Easy

One of the requirements for a good test instrument must look at the discrimination power. The differential power of a test item aims to distinguish between test takers with high abilities and low abilities. The following is a criterial table for the differentiating power of butirl questions.

Tabel B. 4 Criteria for Differential Power

No.	Interpretation of Distinguishing Power	
1.	0,00 - 0,20	0 -25,9%
2.	0,21 – 0,40	26 – 50,99%
3.	0,41 – 0,70	51 – 75,99%
4.	0,71 - 1,00	76 -100%

The effectiveness of the distractor is how well the wrong choice can fool test takers who do not know the answer key. The more test takers who choose these distractors, the better the distractor can perform it’s signaling function. The criteria for the item's butler index are presented in the following table.

Table B. 5 Criteria for Checker Effectiveness

No.	Interpretation of Exposure Effectiveness	
1.	>200%	not good
2.	0% - 25%	Low good
3.	26% - 50%	fair
4.	51% -75%	good
5.	76% -125%	very good

Furthermore, based on several considerations, the quality of the questions that have been analyzed can be determined, including questions with good quality, good enough, low good and not good.

C. RESULT AND DISCUSSION

The results of the data obtained by researchers in the form of answers from all students of Class V PABP subjects SDN 03 East Pontianak in the following 2023 / 2024 academic year.

Table C. 1 Results of Student Answer Data

No.	Name	Question Item Analysis																														Total Score	Score
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30		
1	Afkar	1	1	1	0	0	1	1	1	1	1	1	1	1	0	1	0	0	1	1	1	1	1	1	1	1	0	1	0	0	0	21	70
2	Aisyah	1	0	0	1	1	1	0	0	0	0	1	0	1	0	0	1	0	1	1	0	0	0	0	0	0	1	0	0	0	1	11	37
3	Allaman	1	1	0	0	1	1	0	1	1	1	0	1	1	1	1	0	1	1	1	1	1	0	1	1	1	1	1	0	1	0	22	73
4	Alfin	1	1	1	0	1	1	1	1	1	0	1	0	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1	0	1	0	23	77
5	Alit	1	1	0	0	0	1	0	0	1	1	0	0	1	1	1	0	0	1	0	0	0	0	1	1	1	0	1	0	0	1	14	47
6	Anggun	0	0	0	0	1	0	0	0	0	1	1	1	0	0	1	0	1	1	1	0	1	0	0	0	0	0	0	1	1	1	12	40
7	Aqilla	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	0	0	1	0	0	1	24	80
8	Fahmi	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	1	0	0	1	0	0	0	0	0	1	6	20
9	Farhan	0	0	0	0	1	0	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	0	0	0	1	0	1	1	20	67
10	Farid	0	0	0	0	0	1	0	1	0	0	1	1	0	0	1	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	9	30
11	Hermaiwan	1	0	0	0	0	1	1	1	1	1	0	1	1	1	1	0	0	0	1	0	1	1	1	1	1	0	0	0	0	1	17	57
12	Fikri	1	0	0	1	1	1	1	1	1	0	1	1	1	1	1	0	1	0	1	1	1	1	1	0	1	1	1	0	1	0	22	73
13	Jelsy	1	1	1	0	1	1	1	0	0	0	1	1	1	1	1	0	0	0	0	1	0	1	1	1	1	1	1	0	0	1	19	63
14	Keisha	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	0	1	0	1	1	26	87
15	Elma	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1	1	0	1	0	1	1	0	0	0	0	0	21	70
16	Khalisah	1	1	1	0	1	1	1	0	1	1	1	1	0	1	1	0	0	0	0	0	0	1	1	0	1	1	0	1	0	0	17	57
17	Khilqa	1	1	0	1	0	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	0	1	1	0	1	1	1	0	1	1	24	80
18	Meri	1	1	1	0	0	0	1	0	0	1	1	0	1	1	0	0	0	0	0	1	1	1	1	1	0	1	0	1	0	0	15	50
19	Hajidi	1	0	1	0	0	1	0	0	0	0	0	1	0	0	0	0	0	1	0	1	0	0	1	0	1	0	1	0	0	1	10	33
20	Fahri	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	1	0	1	0	0	0	0	0	0	0	4	13
21	Naufal	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	1	1	0	1	0	0	1	0	0	0	0	0	0	0	1	7	23
22	Naraya	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	1	1	0	0	0	0	0	0	1	1	21	70
23	Novita	0	0	0	0	1	1	1	0	1	1	0	0	1	0	0	0	0	1	1	1	1	0	1	0	1	1	0	0	0	1	14	47
24	Puan	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	0	1	0	1	1	27	90
25	Putra	1	1	1	0	0	0	1	1	1	1	0	1	0	1	1	0	1	1	1	0	1	0	1	0	0	0	1	0	1	1	18	60
26	Raisya	1	1	0	0	1	0	1	0	0	1	0	0	0	0	1	0	1	0	1	1	0	1	1	1	1	1	1	0	0	1	16	53
27	Rendi	1	0	0	0	1	1	1	1	1	0	1	0	1	1	0	0	0	1	0	1	0	1	0	0	0	0	1	0	1	0	14	47
28	Rizki	0	1	0	0	1	1	0	1	1	1	1	0	0	1	1	0	1	1	1	0	1	1	1	0	1	0	1	0	0	0	17	57
29	Rayhan	1	1	1	0	0	0	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	0	0	1	24	80
30	Syafa	1	0	0	0	1	0	0	1	1	1	1	1	1	1	1	0	1	1	1	0	1	1	1	1	1	0	0	0	1	0	19	63

The results of this study are then presented and presented in the form of the results of the calculation of validity, reliability, difficulty level, differential power and the effectiveness of the checkers on item questions End of Semester Summative odd class V at SDN 03 East Pontianak in the 2023/2024 academic year.

1. Validity

The validity test is a test that serves to see whether a measuring instrument is valid (valid) or invalid (Janna & Herianto, 2021). A test is said to be valid or legitimate, if the test really measures what it wants to measure, it must be able to measure the level of learning outcomes achieved in the implementation of a desired goal (Damayanti, 2012). According to Sukardi, a test is said if the coefficient of 0.5 can be accepted, if it is the only one, otherwise if it turns out that there is another prediction test that is similar and has a higher coefficient then the coefficient of 0.5l is not accepted. So a valid test has a minimum coefficient of 0.5. (Elviana, 2020).

Table C. 2 Validity of Problem Items

No	Person correlation	Sig value	Conclusion	Interpretation
1	.605**	0	valid	high
2	.649**	0	valid	high
3	0,359	0,052	invalid	
4	.392*	0,032	valid	low
5	0,312	0,094	invalid	
6	0,245	0,191	invalid	
7	.664**	0	valid	high
8	.689**	0	valid	high
9	.598**	0	valid	sufficient
10	.547**	0,002	valid	sufficient
11	.512**	0,004	valid	sufficient
12	.548**	0,002	valid	sufficient
13	.571**	0,001	valid	sufficient
14	.724**	0	valid	high
15	.484**	0,007	valid	sufficient
16	-.363*	0,049	valid	low
17	.581**	0,001	valid	sufficient
18	0,264	0,159	invalid	
19	0,355	0,055	invalid	
20	.484**	0,007	valid	sufficient
21	0,217	0,25	invalid	
22	.525**	0,003	valid	sufficient
23	.454*	0,012	valid	sufficient
24	.437*	0,016	valid	sufficient
25	.491**	0,006	valid	sufficient
26	0,244	0,194	invalid	
27	.541**	0,002	valid	sufficient
28	-0,118	0,536	invalid	
29	.482**	0,007	valid	sufficient
30	-0,006	0,974	invalid	

Based on the results of data analysis obtained using IBM SPSS statistics 23, the validity coefficient (*r* table) is 0,316 or a question item if the significance is < 0,05, it can be said that the question item is valid. After observing the table above, it is known that of the 30 items analyzed, there are 8 items that are insignificant or invalid, namely items 3, 5, 6, 18, 19, 21, 28 and 30, the rest are 22 significant items or valid items.

2. Reliability

Wahyudin (2014) argues that a measurement instrument is said to be reliable if the instrument is used repeatedly will show the same results (Ovan & Andika Saputra, 2020). Reliability testing is carried out on 22 items with a significance level of 5%. And the degree of freedom obtained the value of *r* table = 0,537 or the question is said to be reliable if the Cronbach alpha value is > 0,6. This test is carried out only for questions that are said to be valid, invalid questions do not need to be tested for reliability again because it will affect other question items.

Table C. 3 Reliability of Problem Items

Cronbach's Alpha	N of Items
0,872	22

From the results of the analysis, it can be seen that there are 22 items that can be tested for reliability and the correlation coefficient (*r*) is 0,872. To determine whether this *r* (0,872) indicates that the test has a high or low level of reliability, reliable or unreliable, it must be confirmed with the correlation table or with the correlation index, because the *r* table of 22 items is 0,537, it can be concluded that the correlation coefficient of 0,872 is higher than the score in the correlation table, then the test can be said to be reliable. However, to see the interpretation, we can measure it using the correlation index, that *r* count0,872, it can be concluded that the level of reliability is high because it is in the range of 0,70 – 0,89.

A reliable test is a test that is consistent, not indecisive or doubtful (adamu at-tadzabdzub). This definition is also reinforced by Brown (2007), that a reliable test is a consistent and reliable test. Specifically, consistency refers to the constancy of the score. that is, if the test is tested more than once (for example twice) on the same subject at different times, but has more or less the same or even the same results.

For this reason, researchers use the split-half technique or tajzi'atu al-ikhtibar (split-half method). This technique is done by separating the scores into two groups, namely odd groups and even groups. What is meant by odd and even groups are the answer in the form of scores from odd number items and even number items. The following are the results of data using the two-split technique using IBM SPSS statistics 23.

Table C. 4 Reliability of Question Item 2

Cronbach's Alpha	Part 1	Value	0,853
		N of Items	11[a]
	Part 2	Value	0,673
		N of Items	11[b]
	Total N of Items		22
Correlation Between Forms			0,720
Spearman-Brown Coefficient	Equal Length		0,837
	Unequal Length		0,837
Guttman Split-Half Coefficient			0,816

The correlation coefficient from this calculation (0,816) only indicates the reliability of the test for half of the questions. Meanwhile, what is expected is the level of reliability for all questions. To get the correlation coefficient of the reliability level of all questions, the Spearman-Brown formula is used as follows (Magdalena et al., 2021).

Rumus:

$$\text{Reliability of the whole test} = \frac{2 \times \text{reliability of half the test}}{1 + \text{half test reliability}}$$

$$r = \frac{2 \times 0{,}816}{1 + 0{,}816} = \frac{1{,}632}{1{,}816}$$

$$r = 0{,}899$$

From the results of the analyies it's can be seen, that *r* count of 0,899> *r* table (0,537). This means that the level of reliability of the test with the two-split technique is high / reliable. Similarly, when confirmed with the correlation index, it can be stated that the level of reliability is high, because it is between 0,70 and 1,89.

3. Level of Difficulty

Analysis of the level of difficulty means examining test questions in terms of their difficulty, so that questions can be obtained that are feasible to be given to students, to find out which questions are easy, which questions are moderate and which questions are difficult. What is meant by good items or questions is that the level of difficulty can be known not too difficult and not too easy (Fatimah & Alfath, 2019). The level of difficulty of the question can be seen from how many students can answer it, not from the point of view of the teacher who made the (Magdalena et al., 2021). The results of the analysis of the level of difficulty are:

Table C. 5 Level of Difficulty of Problem Items

	N		Mean	the level of difficulty
	Valid	Missing		
Item1	30	0	0,733	easy
Item 2	30	0	0,567	Medium
Item 3	30	0	0,467	Medium
Item 4	30	0	0,200	difficult
Item 5	30	0	0,567	Medium
Item 6	30	0	0,700	Medium
Item 7	30	0	0,633	Medium
Item 8	30	0	0,600	Medium
Item 9	30	0	0,700	Medium
Item 10	30	0	0,667	Medium
Item 11	30	0	0,667	Medium
Item 12	30	0	0,633	Medium
Item 13	30	0	0,633	Medium
Item 14	30	0	0,667	Medium
Item 15	30	0	0,800	easy
Item 16	30	0	0,067	difficult
Item 17	30	0	0,567	Medium
Item 18	30	0	0,667	Medium
Item 19	30	0	0,767	Medium
Item 20	30	0	0,600	Medium
Item 21	30	0	0,667	Medium
Item 22	30	0	0,700	Medium
Item 23	30	0	0,733	easy
Item 24	30	0	0,467	Medium

Item 25	30	0	0,633	Medium
Item 26	30	0	0,333	Medium
Item 27	30	0	0,600	Medium
Item 28	30	0	0,067	difficult
Item 29	30	0	0,400	Medium
Item 30	30	0	0,633	Medium

Based on the table, the results show that there are 23 items or about 77% in the moderate category, there are 4 items or about 13% in the easy category, and there are 3 items or about 10% in the easy category. It can be concluded that out of 30 questions, there are 23 questions or about 73% can be reused for future exams, there are 4 questions or about 13% that should be re-examined to find out why the questions can be easily done by students, and with the ease of paral students can find the answer key and the only thing that is bad is an exception. Then there are 3 items or about 10% butirl questions that are difficult, there may be several follow-up actions, namely: (1) the item is no longer used and is no longer issued in the next test results. (2) re-examined, researched and traced so that the factors causing children to be unable to answer the question can be known.

4. Differentiating power

According to Daryanto, question discriminating power is the ability of a question to distinguish between students who are smart (high ability) and students who are stupid (low ability) (Rahmaini & Taufiq, 2018). The higher the discriminating power of the question means that the better the question in question distinguishes students who understand and do not understand the material (Dewi et al., 2019). Good test items must be able to distinguish between capable test takers and incapable test takers, in other words, good test items should be answered correctly by capable test takers and answered incorrectly by incapable test takers (Rusmayani, 2020). The following is the differential power obtained from analysis using IBM SPSS statistic 23.

Table C. 6 Differentiating Power of Problem Items

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted	Differentiation Index
Item1	33,53	141,913	0,581	0,718	good
Item2	33,7	140,631	0,624	0,715	good
Item3	33,8	144,166	0,321	0,723	good enough
Item4	34,07	144,616	0,364	0,724	good enough
Item5	33,7	144,769	0,274	0,724	good enough
Item6	33,57	145,771	0,209	0,726	good enough
Item7	33,63	140,654	0,64	0,715	good
Item8	33,67	140,23	0,666	0,714	good
Item9	33,57	141,771	0,572	0,718	good
Item10	33,6	142,179	0,519	0,719	good
Item11	33,6	142,593	0,482	0,72	good
Item12	33,63	142,033	0,519	0,718	good
Item13	33,63	141,757	0,543	0,718	good

Item14	33,6	140,11	0,705	0,714	very good
Item15	33,47	143,706	0,458	0,722	good
Item16	34,2	150,648	-0,381	0,735	No good
Item17	33,7	141,459	0,553	0,717	good
Item18	33,6	145,49	0,227	0,726	good enough
Item19	33,5	144,81	0,323	0,724	good enough
Item20	33,67	142,713	0,452	0,72	good
Item21	33,6	146,041	0,179	0,727	No good
Item22	33,57	142,599	0,496	0,719	good
Item23	33,53	143,568	0,424	0,722	good
Item24	33,8	143,2	0,402	0,721	good
Item25	33,63	142,723	0,459	0,72	good
Item26	33,93	145,72	0,207	0,726	good enough
Item27	33,67	142,023	0,511	0,718	good
Item28	34,2	149,131	-0,138	0,732	No good
Item29	33,87	142,74	0,449	0,72	good
Item30	33,63	148,654	-0,046	0,732	No good

Based on this table, it is known that there are 18 (60%) items categorized as medium, there are 7 (23%) items categorized as sufficient, there are 4 (13%) items categorized as unfavorable and there are 1 (3%) item categorized as very good. That it can be concluded that the odd semester-end summative items of class V PABP lessons at SDN 03 East Pontianak in 2023/2024 have good quality because they are in the range of 51 – 75,99%.

5. Excerpt Effectiveness

Exterminators can also be called distractors, in multiple choice questions the possible answers are divided into two, namely the answer key and distractors. Of the many alternative answers only one is correct, namely the answer key and the possibility of an incorrect answer is called an exception. Distractors function to identify test takers who have high abilities. The purpose of analyzing distractors is to find out how many students answer correctly according to the answer key and how many choose distractors. If the distractor has a large attraction, it can be said to be effective. In other words, if a distractor is chosen by many test takers, then the distractor is said to be functional. Vice versa, a poor or bad distractor is one that is not chosen or chosen a little by the tested participants (Pradita et al., 2023).

The following is a test of distractors or tricky using IBM SPSS statistic 23 then manually clarified using excel and the results also show the same percentage, so for more clarity we can see from the formula and the following table.

Formula:

D= A/ N x 100%

D: Distractor/Electrifier Level

A: Number of students who chose the answer

N: Total number of students

Table C. 7 Percentage of Eligibility for Each Option

No	Answer	A	B	C	D
1	C	20,00%	3,33%	73,33%	3,33%
2	D	10,00%	23,33%	10,00%	56,67%
3	B	13,33%	46,67%	33,33%	6,67%
4	C	40,00%	20,00%	20,00%	20,00%
5	A	56,67%	6,67%	33,33%	3,33%
6	C	13,33%	6,67%	70,00%	10,00%
7	D	6,67%	20,00%	10,00%	63,33%
8	A	60,00%	20,00%	13,33%	6,67%
9	A	70,00%	10,00%	13,33%	6,67%
10	D	16,67%	10,00%	6,67%	66,67%
11	D	10,00%	10,00%	13,33%	66,67%
12	A	63,33%	23,33%	13,33%	0,00%
13	B	13,33%	63,33%	3,33%	20,00%
14	C	20,00%	10,00%	66,67%	3,33%
15	D	6,67%	10,00%	3,33%	80,00%
16	B	66,67%	6,67%	23,33%	3,33%
17	C	20,00%	13,33%	56,67%	10,00%
18	A	66,67%	10,00%	10,00%	13,33%
19	A	76,67%	20,00%	3,33%	0,00%
20	B	13,33%	60,00%	20,00%	6,67%
21	A	66,67%	13,33%	13,33%	6,67%
22	A	70,00%	13,33%	6,67%	10,00%
23	A	73,33%	13,33%	0,00%	13,33%
24	C	6,67%	6,67%	46,67%	40,00%
25	B	6,67%	63,33%	26,67%	3,33%
26	D	6,67%	50,00%	10,00%	33,33%
27	B	16,67%	60,00%	16,67%	6,67%
28	B	13,33%	6,67%	70,00%	10,00%
29	C	26,67%	20,00%	40,00%	13,33%
30	B	0,00%	63,33%	16,67%	20,00%

Based on the results of research using IBM SPSS statistic 23, there are 30 items of multiple choices questions with the effectiveness of triggers in the very good category as many as 3 items (10%), good category as many as 4 items (13%), good enough category as many as 21 items (70%), poor category as many as 2 items (7%).

According to Suharsimi Arikunto no further after analyzing the effectiveness of the distractor, it can be treated in the following 3 ways: a. Maximized because it is good. This means that all the distractors on the question have been selected by 5% of the test participants. b. Rejected because it is not good. This means that no distractors were chosen by the test takers (0%). c. Rewritten (revised) because it is not good. This means that the distractor has not performed its function well (distractors are selected less than 5%) (Rahmaini & Taufiq, 2018). The following is an explanation of whether the effectiveness of an exception can be accepted, revised or rejected.

Table C. 8 Effectiveness of Examiners

Item	A	B	C	D
1	Retrieved	revised	Retrieved	revised
2	Retrieved	Retrieved	Retrieved	Retrieved
3	Retrieved	Retrieved	Retrieved	Retrieved
4	Retrieved	Retrieved	Retrieved	Retrieved
5	Retrieved	Retrieved	Retrieved	revised
6	Retrieved	Retrieved	Retrieved	Retrieved
7	Retrieved	Retrieved	Retrieved	Retrieved
8	Retrieved	Retrieved	Retrieved	Retrieved
9	Retrieved	Retrieved	Retrieved	Retrieved
10	Retrieved	Retrieved	Retrieved	Retrieved
11	Retrieved	Retrieved	Retrieved	Retrieved
12	Retrieved	Retrieved	Retrieved	rejected
13	Retrieved	Retrieved	revised	Retrieved
14	Retrieved	Retrieved	Retrieved	revised
15	Retrieved	Retrieved	revised	Retrieved
16	Retrieved	Retrieved	Retrieved	revised
17	Retrieved	Retrieved	Retrieved	Retrieved
18	Retrieved	Retrieved	Retrieved	Retrieved
19	Retrieved	Retrieved	revised	rejected
20	Retrieved	Retrieved	Retrieved	Retrieved
21	Retrieved	Retrieved	Retrieved	Retrieved
22	Retrieved	Retrieved	Retrieved	Retrieved
23	Retrieved	Retrieved	rejected	Retrieved
24	Retrieved	Retrieved	Retrieved	Retrieved
25	Retrieved	Retrieved	Retrieved	revised
26	Retrieved	Retrieved	Retrieved	Retrieved
27	Retrieved	Retrieved	Retrieved	Retrieved
28	Retrieved	Retrieved	Retrieved	Retrieved
29	Retrieved	Retrieved	Retrieved	Retrieved
30	rejected	Retrieved	Retrieved	Retrieved

Based on this analysis, it can be concluded that of the 30 questions analyzed, there are several choice items that are rejected because they do not reach 5% of the provisions, namely question nos. And from the aspect of eliminators, it was found that about 89% were accepted, about 8% were revised and about 3% were rejected.

An exemplar is said to function well if it is selected by 5% of test takers. If an exemplar is selected evenly, then the exemplar is considered very good. Making good exams on multiple-choice tests is difficult, because poor exams will result in low discriminating power, and if one or two exams are not functioning low.

D. CONCLUSION

Item analysis is a very important thing in the world of education, item analysis is carried out to determine the level of quality of questions used as an assessment of the development of student learning outcomes. Because a quality question is a question that can provide precise information about students who master the material and who have not

mastered the material. The results of the analysis will show how good the quality of a question is. Quality questions can be known from the validity, reliability, level of difficulty of the questions, differentiating power, and the effectiveness of the questions. Through the results of the analysis, the teacher can evaluate the question.

Based on the results of the research and discussion regarding the analysis of the Final Examination Question Item of Gasal Semester Class V SDN 03 East Pontianak, it can be concluded that the quality of the question items is good. So that in the future, all teachers are no exception to all teachers in Indonesia so that they can always analyze after providing an assessment or evaluation. The existence of this evaluation can improve teacher competence in preparing end-of-semester exam items, because teachers have the opportunity to discuss together to review and solve problems based on the situation in the field, then can improve or follow up in the next cycle continuously if the problem has not been resolved.

REFERENCE

Arifianto, M. L., Ainin, M., Irhamni, Ahsanuddin, M., Hikmah, K., Anwar, M. S., & Fitria, N. (2021). *Evaluasi Dan Pengembangan Tes Interaktif Bahasa Arab*.

Damayanti, E. T. (2012). *Analisis Butir Soal Evaluasi Pembelajaran Pendidikan Agama Islam Kelas XI IPA & IPS DI SMAN 3 Probolinggo* (Vol. 01).

Dayu, R., & Suharmon. (2021). Analisis Butir Soal Tes Pilihan Ganda Mata Pelajaran PAI BP Kelas IPS SMAN 1 Sungai Lasi Tahun Ajaran 2021 / 2022. *International Conference on Islam, Law, and Society (INCOILS)*.

Dewi, S. S., Hariastuti, Rachmaniah, & Utami, A. U. (2019). Analisis Tingkat Kesukaran Dan Daya Pembeda Soal Olimpiade Matematika (Omi) Tingkat Smp Tahun 2018. *Transformasi : Jurnal Pendidikan Matematika Dan Matematika*, *3*(1), 15–26. https://doi.org/10.36526/tr.v3i1.388

Elviana. (2020). Analisis Butir Soal Evaluasi Pembelajaran Pendidikan Agama Islam Menggunakan Program Anates. *Analisis Butir Soal Evaluasi Pembelajaran Pendidikan Agama Islam Menggunakan Program Anates*, *10*(2), 58–74.

Fatimah, L. U., & Alfath, K. (2019). analisis Kesukaran Soal, Daya Pembeda dan Fungsi Distraktor. *Jurnal Komunikasi Dan Pendidikan Islam*, *8*, 37–64.

Febriana, R. (2019). *Evaluasi Pembelajaran*.

Janna, N. M., & Herianto. (2021). Artikel Statistik yang Benar. *Jurnal Darul Dakwah Wal-Irsyad (DDI)*, *18210047*, 1–12.

Khoiriyah, N., Maslikhatun Nisak, N., & Niam, K. (2023). Students' Perceptions of Self-Assessment in Islamic Education Learning. *Jurnal Pendidikan Agama Islam*, *10*(1), 2023. http://ejournal.uin-malang.ac.id/index.php/jpai

Magdalena, I., Fauziah, S. N., Faziah, S. N., & Nupus, F. S. (2021). Analisis Validitas, Reliabilita, Tingkat Kesulitan Dan Daya Beda Butir Soal Ujian Akhir Semester Tema 7 Kelas III SDN Karet 1 Sepatan. *BINTANG : Jurnal Pendidikan Dan Sains*, *3*(2), 198–214. https://ejournal.stitpn.ac.id/index.php/bintang

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

Mastiah, M. (2019). Upaya Meningkatkan Hasil Belajar PAI melalui Model Pembelajaran Kooperatif Tipe Numbered Heads Together (PTK di Kelas VIII SMPI Nurul Ikhsan Kibin-Serang)". *Diss. UIN SMH BANTEN*, 1–37. http://repo.iain-tulungagung.ac.id/8073/4/BAB I .pdf

Matondang, Z. (2009). *Evaluasi Pembelajaran*.

Ovan, & Andika Saputra. (2020). *CAMI: Aplikasi Uji Validitas dan Reliabilitas Instrumen Penelitian Berbasis Web*.

Pradita, E., Megawanti, P., & Yulianingsih. (2023). Analisis Tingkat Kesukaran, Daya Pembeda, dan Fungsi Distraktor PTS Matematika SMPN Jakarta. *Jurnal Ilmiah Mahasiswa Pendidikan Matematika*, *3*(80), 109–118.

Purnomo, P., & Solikhah, P. I. (2021). Konsep Dasar Pendidikan Islam Inklusif : Studi Tentang Inklusivitas Islam Sebagai Pijakan Pengembangan Pendidikan Islam Inklusif. *J-PAI: Jurnal Pendidikan Agama Islam*, *7*(2), 114–127. https://doi.org/10.18860/jpai.v7i2.13286

Rahmaini, A., & Taufiq, A. N. (2018). Analisis Butir Soal Pendidikan Agama Islam Di SMK N 1 Sedayu Tahun Ajaran 2017/2018 (Analisis Tingkat Kesukaran, Daya Pembeda dan Fungsi Distraktor pada Soal Pilihan Ganda Kelas XI). *Jurnal Mudarrisuna*, *8*(1), 1–24. https://www.jurnal.ar-raniry.ac.id/index.php/mudarrisuna/article/view/2787

Rusmayani. (2020). Analisis Butir Soal Penilaian Akhir Semester Genap Mata Pelajaran Pendidikan Agama Islam di SMP Bintang Persada Tabanan-Bali. *Widya Balina*, *5*(1), 41–49. https://doi.org/10.53958/wb.v5i1.50

Sridadi, Dwihandaka, R., & Bagiastomo, A. (2020). The analysis of test item on learning outcomes of physiical education subject of class VIII students. *Jurnal Pendidikan Jasmani Indonesia*, *16*(1), 28–40.

Tausih, T. U., & Marno, M. (2021). Pelaksanaan Penilaian Ranah Afektif Menggunakan Google Form di Era New Normal. *Jurnal Pendidikan Agama Islam*, *7*(2), 103–113. https://doi.org/10.18860/jpai.v7i2.12270