Journal of Mathematics Education
Website: http://usnsj.id/index.php/JME
Email: pengelolajme@gmail.com
p-ISSN 2528-2468; e–ISSN 2528-2026

DOI: https://doi.org/10.31327/jme.v9i1.2021

ANALYSIS OF MATHEMATICS SUMATIVE EXAMINATION PROBLEMS IN CLASS VII OF SMP NEGERI 2 NGANTANG USING ANATES

Mohamad Kholil[1], Surya Sari Faradiba*[2]
[1,2] Universitas Islam Malang

Article Info

Article history:

Received Jul 31, 2023
Revised Mei 18, 2024
Accepted Jun 01, 2024

Keywords:

Mathematic Sumative
Examination Problems
Anates

ABSTRACT

Researching the quality of educational assessments is crucial for ensuring effective learning outcomes. This study aims to: (1) describe the level of difficulty, discriminatory power, and the effectiveness of distracting items on Year-End Assessment of mathematics using Anates software in class VII SMP Negeri 2 Ngantang, (2) describe the percentage quality of difficulty level, discriminating power, and The effectiveness of the item distractor on Year-End Assessment of mathematics subject using Anates in class VII of SMP Negeri 2 Ngantang. This study uses a qualitative approach. The data source used is a secondary data source. The data collection technique used is documentation. The research data was then analyzed using Anates software. The results showed that overall the items were quite valid, the reliability was high, the average level of difficulty was moderate but not balanced, the proportion of easy to medium difficulty should be 4 – 4 - 4 or 4 - 4 -5 - 3, the discriminating power of all questions was accepted, the distractor was generally effective, only 2 out of 36 or 5.56% that were badly needed to be replaced, while the significance of the items was 7 out of 12 questions or 57.33% significant.

Corresponding Author:

Surya Sari Faradiba,
The *Master of Mathematics Education* Program
Universitas Islam Malang
Email: suryasarifaradiba@unisma.ac.id
Phone Number : 085649532920

How to Cite:

Faradiba, S.S., & Kholil, M. (2024). Analysis of Mathematics Sumative Examination Problems in Class VII of SMP Negeri 2 Ngantang Using Anates. *JME: Journal of Mathematics Education*, *9*(1), 119-126.

1. INTRODUCTION

Evaluation is critical because it is a benchmark for students in achieving competence (Angriani et al., 2021). A test is an assessment tool to measure student learning outcomes throughout the learning process. Evaluation in mathematics education refers to collecting and analyzing data to measure students' progress, achievement, and understanding of mathematical concepts and the teaching methods' effectiveness. The evaluation aims to

provide accurate information about the extent to which students have achieved the learning objectives set and how the learning process can be improved (Mardiana et al., 2021).

Evaluation can take many forms, such as written tests, projects, presentations, or group activities. It is essential to combine different types of evaluation to get a more comprehensive picture of students' progress and understanding of mathematics. Using evaluation effectively, educators can improve the quality of learning mathematics, provide appropriate student support, and ensure that learning objectives are appropriately achieved. This study focuses on evaluation in the form of written tests. In written tests, item analysis is needed to assess the quality of good items so they can be reused in the next period or if the items are not good. Revisions can be made, while items that are not good do not need to be reused or replaced with a new question(Brown & Abdulnabi, 2017; Khairani & Shamsuddin, 2016; Kusumawati & Hadi, 2018).

The difficulty level of a question is a measure of how challenging it is for test takers, often determined by the percentage of students who answer it correctly. Analyzing this difficulty level involves categorizing questions into three groups: easy, medium, and difficult. This classification helps in understanding the overall test composition and ensuring a balanced assessment. Additionally, the discriminating power of a test item is crucial; it refers to the item's ability to distinguish between students with high and low levels of understanding or ability. A well-designed question will have high discriminating power, meaning it effectively identifies the differences in performance among test takers. Moreover, the role of distractors incorrect answer choices in multiple-choice questions—is vital in evaluating the quality of a test item. A distractor is considered effective if at least 5% of test takers choose it, indicating that it is plausible enough to mislead those who are unsure of the correct answer (Brown & Abdulnabi, 2017). The effectiveness of a distractor is thus measured by its ability to attract incorrect responses from those who do not know the right answer, contributing to the overall reliability and validity of the test (Rangan et al., 2021).

Anates is an application program specifically used to analyze multiple-choice tests and essays. Anates can analyze test items such as: calculating scores, calculating test reliability, grouping subjects into upper or lower groups, calculating discriminatory power, calculating the level of difficulty, calculating the correlation of item scores with the total score, determines the quality of the distractor, and is used on the Windows operating system. Anates functions are the same as other data processing items, such as ITEMAN (Item and Test Analysis) and AnBuso (Analisis Butir Soal), but it is easier to operate (Mawardi et al., 2023a, 2023b; Sanova et al., 2017a, 2017b; Stevani et al., 2022a, 2022b; Sudarto et al., 2023a, 2023b). In addition, the results have been directly analyzed by the program. So, analyzing it again with the existing criteria is optional. The functions and benefits of this are certainly for analyzing the data of the multiple-choice questions being tested. The advantages of the Anates program are: a) It can be used to analyze the items in the form of essays and multiple- choice questions, b) Analyze the item descriptions and multiple-choice questions quickly, c) Program instructions are easy to understand, d) use Indonesian, and e) the results of Anates can be printed immediately. The weaknesses of the Anates Program are: a) Data entry can only be done manually, and b) incorrect data entry will reduce the value of the final result. Based on the description above, the researcher conducted a study titled Analysis of Mathematics Sumative Examination Problems in Class VII of SMP Negeri 2 Ngantang using Anates.

2. METHOD

This study uses a qualitative approach. Researchers use the documentation method to find data about things or variables through notes, transcripts, books, newspapers, magazines, inscriptions, meeting minutes, and so on (Creswell & Guetterman, 2019). The data source used is

a secondary data source. Secondary data is already available and can be obtained by researchers with how to read, see or hear.

This research was conducted in December 2022 at SMP Negeri 2 Ngantang. Class 7A students are divided into two rooms, numbers 1 to 16 occupy room 1 in class 8A, while numbers 17 to 31 occupy room 2 in class 7D. The research procedure begins with the preparation stage, namely compiling the question grid, compiling questions, creating Google Forms, arranging schedules, and installing Anates Ver 4.0.9. At the data collection stage, all students are in the room, and each room is guarded or served by a room supervisor from the teacher element. The supervisor conveys a link students must access to work on exam questions. After the implementation phase, the analysis phase is continued, preceded by entering student answers into the Anates Ver 4.0.9 application. From the results of the analysis of the Anates Ver 4.0.9 application, it is continued with discussion, reporting, and publication.

The research begins with planning to make instruments through questions tested on students. Questions are presented in Google Forms. At the time of data collection, all students were in the room, and each room was guarded or served by a room supervisor from the teacher element. The supervisor conveys a link, and the students must access it to work on exam questions. After the students did the questions, the student's answers on Google Forms were exported to a spreadsheet. Then the data was entered into Anates Ver 4.0.9. The next processed Anates Ver 4.0.9 results became the subject of discussion in this research. Data analysis in this study used Anates Ver 4.0.9 so that the validity, reliability, level of difficulty, discriminating power, and distracting quality of the items were obtained based on Table 1-4 (Permansah et al., 2023).

Table 1. Criteria for determining the validity coefficient of the items

Value	Description
0,91 - 1,00	Very High
0,71 - 0,90	High
0,41 - 0,70	Moderate
0,21 - 0,40	Low
< 0,20	Very low

Table 2. Criteria for determining the reliability of the items

Value	Description
0,80 - 1,00	Very High
0,60 - 0,79	High
0,40 - 0,59	Moderate
0,20 - 0,39	Low
0 – 0,19	Very low

Table 3. Criteria for criteria of item difficulty level

Value	Description
0 – 15 %	Very Difficult
16 – 30 %	Difficult
31 – 70 %	Moderate
71 – 85 %	Easy
86 – 100 %	Very Easy

Table 4. Criteria of item discrimination

Value	Description
More than 25%	Approved
0 – 25 %	Revised
Less than 0%	Rejected

3. RESULTS AND DISCUSSION

3.1. Result

Table 5 shows that the correlation value is 0.60 (between 0.41 and 0.70), so the validity of this test item is in the sufficient category. Thus this test item is quite capable of measuring what should be measured.

Table 5. The Result of Reliability Analysis

	Value
Average	5,94
Deviation Standar	2,85
Correlation	0,60
Reliability	0,75

Table 6 shows that of the 12 multiple choice questions, there are two easy category questions (16.67%), namely numbers 1 and 7, and eight moderate category questions (66.66%), namely numbers 2, 3, 4, 5, 6, 9, 10 and 12. At the same time, there are two difficult category questions (16.67%), namely numbers 8 and 11. The assumption in preparing good quality test items is that there is a balance of the number of questions in their difficulty level and their validity and reliability. The balance of this difficulty level is the number of questions in the easy category, and there must be questions in the difficult category and the moderate category proportionally. So the proportions of two easy, eight moderate, and two difficult need to be balanced. Easy and difficult questions must be added, and moderate questions must be reduced.

Table 6. The Result of Analysis of Difficulty Level

No	Correct Amount	Difficulty Level (%)	Description
1	26	83,87	Easy
2	18	58,06	Moderate
3	13	41,94	Moderate
4	19	61,29	Moderate
5	17	54,84	Moderate
6	10	32,26	Moderate
7	23	74,19	Easy
8	8	25,81	Difficult
9	12	38,71	Moderate
10	13	41,94	Moderate
11	6	19,35	Difficult
12	19	61,29	Moderate

Table 7 shows that all questions have a discriminating power index value above 25%. According to the category, all questions are acceptable.

Table 7. The Result of Analysis of Discrimination Items

No	High Group	Low Group	Difference	Index (%)
1	8	3	5	62,50
2	8	2	6	75,00
3	8	1	7	87,50
4	8	2	6	75,00
5	7	2	5	62,50
6	5	0	5	62,50
7	7	3	4	50,00
8	5	1	4	50,00
9	6	1	5	62,50
10	6	2	4	50,00
11	3	0	3	37,50
12	8	2	6	75,00

Table 8 shows that there is a very good distractor in choice A questions numbers 4, 7, 8, 9, 10, and 11, choice B questions numbers 1, 6, and 7, choice C questions numbers 1, 2, 3, 4 and 10, choice D question number 10. There are distractors with good criteria in option A, numbers 5 and 6. Choice B questions numbers 3, 8, and 11. Choice C questions numbers 6, 7, and 8. Choice D questions numbers 1, 3, 5, 11, and 12. The distractor with fewer criteria is option A, numbers 2 and 12. Choice B question number 4, 5, and 9, choice C question number 12, and Choice D question number 9, this distractor needs to be considered for replacement. While the lousy criterion is in option D, question numbers 2 and 4, this distractor must be replaced. Criteria for terrible deception are not found in this problem.

Table 8. The Result of Distractor Efectiveness Analysis

No	A	B	C	D	*
1	26**	2++	2++	1+	0
2	7-	18**	5++	1--	0
3	13**	9+	5++	4+	0
4	4++	7-	19++	1--	0
5	3+	2-	17**	7+	0
6	5+	7++	9+	10**	0
7	3++	3++	2+	23**	0
8	8++	5+	10+	8**	0
9	6++	11-	12**	2-	0
10	5++	13**	7++	6++	0
11	7++	12+	6**	6+	0
12	7-	19**	2-	3+	0

Furthermore, table 9 shows the significance value of the correlation, where there are seven questions out of the twelve questions tested, namely numbers 1, 2, 3, 4, 6, 9, and 12, can be reused, while five questions, namely numbers 5, 7, 8, 10 and 11 need replaced.

Table 9. Item Analysis Recapitulation

No	Level of Discrimination (%)	Level of Difficulty	Correlation	Sig
1	62,5	Easy	0,533	Sig
2	75,0	Moderate	0,563	Sig
3	87,5	Moderate	0,579	Sig
4	75,0	Moderate	0,596	Sig
5	62,5	Moderate	0,441	-
6	62,5	Moderate	0,533	Sig
7	50,0	Easy	0,460	-
8	50,0	Difficult	0,434	-
9	62,5	Moderate	0,538	Sig
10	50,0	Moderate	0,462	-
11	37,5	Difficult	0,331	-
12	75,0	Moderate	0,548	Sig

3.2. Discussion

The findings from this study follow the analysis of the items carried out by Manalu et al. (2019), which shows that questions with poor quality need to be revised and even replaced with good quality. The questions that already have good quality can be put in the question bank and used again. In line with this opinion, Kusumawati & Hadi (2018); Manalu et al. (2019); Siregar & Panjaitan (2022) revealed that items that were not good enough could be revised.

Questions that have already demonstrated good quality can be placed in a question bank for future use. This practice not only ensures a consistent standard of assessment but also saves time and resources in the long run. By maintaining a repository of well-crafted questions, educators can quickly assemble tests and exams that meet the desired standards of quality. This approach aligns with the principles of efficient educational assessment and supports continuous improvement in teaching and learning practices (Stacey & Turner, 2015; Villarroel, 2018).

In line with this perspective, previous studies by Kusumawati & Hadi (2018) have also underscored the importance of revising items that do not meet quality criteria. These researchers have demonstrated through their analyses that revision is a crucial step in the development of effective assessment tools. By systematically reviewing and refining questions, educators can address issues such as ambiguity, bias, and inappropriate difficulty levels, thereby ensuring that assessments accurately reflect student learning and capabilities.

Moreover, the collaborative insights from these studies suggest that the process of item revision should be ongoing and iterative. Continuous feedback from both students and educators can provide valuable information about the effectiveness of assessment questions. This feedback loop enables the identification of problematic questions and the opportunity to improve them. In conclusion, maintaining a high-quality question bank and regularly revising assessment items are essential practices for creating reliable and valid assessments that truly measure student learning outcomes.

4. CONCLUSION

The findings of this study revealed that the items were generally valid, the reliability was high, the average level of difficulty was moderate but unbalanced, the proportion of easy

to medium difficulty should be 4 - 4 - 4 or 4 - 4 -5 - 3, the discriminating power of all questions was accepted, the distractor was generally effective, only 2 out of 36 or 5.56% of the items needed to be replaced, and the significance of the items was 7 out of 12 questions.

REFERENCES

Angriani, A. D., Mania, S., Alam, S., Rasyid, M. R., Kusumayanti, A., Tarbiyah, F., Uin, K., & Makassar, A. (2021). Analisis Soal Ujian Akhir Semester Matematika SMP. *Alauddin Journal of Mathematics Education*, *3*(1), 1–11. https://doi.org/10.24252/AJME.V3I1.20071

Brown, G. T. L., & Abdulnabi, H. H. A. (2017). Evaluating the Quality of Higher Education Instructor-Constructed Multiple-Choice Tests: Impact on Student Grades. *Frontiers in Education*, *2*. https://doi.org/10.3389/FEDUC.2017.00024

Creswell, J. W., & Guetterman, T. C. (2019). *Educational research : planning, conducting, and evaluating quantitative and qualitative research*. Pearson.

Khairani, A. Z., & Shamsuddin, H. (2016). Assessing Item Difficulty and Discrimination Indices of Teacher-Developed Multiple-Choice Tests. *Assessment for Learning Within and Beyond the Classroom*, 417–426. https://doi.org/10.1007/978-981-10-0908-2_35

Kusumawati, M., & Hadi, S. (2018). An analysis of multiple choice questions (MCQs): Item and test statistics from mathematics assessments in senior high school. *REID (Research and Evaluation in Education)*, *4*(1), 70–78. https://doi.org/10.21831/REID.V4I1.20202

Manalu, D., Sipayung, K. T., & Lestari, F. D. (2019). AN ANALYSIS OF STUDENTS READING FINAL EXAMINATION BY USING ITEM ANALYSIS PROGRAM ON ELEVENTH GRADE OF SMA NEGERI 8 MEDAN. *JETAL: Journal of English Teaching & Applied Linguistic*, *1*(1), 13–19. https://doi.org/10.36655/JETAL.V1I1.98

Mardiana, E., Haryati, F., & Wahyuni, S. (2021). Praktek Asessmen dalam Pembelajaran Matematika. *Jurnal Basicedu*, *5*(6), 5859–5876. https://doi.org/10.31004/BASICEDU.V5I6.1798

Mawardi, M. sholeh, Fuady, A., & Sunismi, S. (2023a). Analisis Butir Soal Pilihan Ganda Menggunakan Anates pada Penilaian Tengah Semester Kelas VII D SMP Negeri 1 Ngajum Kabupaten Malang. *Wahana*, *75*(1), 31–41. https://doi.org/10.36456/WAHANA.V75I1.6820

Mawardi, M. sholeh, Fuady, A., & Sunismi, S. (2023b). Analisis Butir Soal Pilihan Ganda Menggunakan Anates pada Penilaian Tengah Semester Kelas VII D SMP Negeri 1 Ngajum Kabupaten Malang. *Wahana*, *75*(1), 31–41. https://doi.org/10.36456/WAHANA.V75I1.6820

Permansah, S., Indrawati, C. D. S., Muhtar, M., & Rusmana, D. (2023). Effectiveness of simulation-based learning using “e-archive” technology in the archiving subject vocational school. *International Journal of Evaluation and Research in Education*, *12*(3), 1520–1527. https://doi.org/10.11591/IJERE.V12I3.25147

Rangan, A. Y., Qomariah, S., & Yusnita, A. (2021). Pendampingan Pemanfaaatan Media Sosial sebagai Sarana Membangun Citra Pondok Pesantren Nurul Mustafa Al-Husaini. *BANTENESE : JURNAL PENGABDIAN MASYARAKAT*, *3*(1), 1–9. https://doi.org/10.30656/PS2PM.V3I1.3344

Sanova, A., Bakar, A., & Afrida, A. (2017a). Standarisasi Instrumen Penilaian Hasil Belajar dengan Program Anates V4 Bagi Guru SMPN 17 Kota Jambi. *Jurnal Pengabdian Pada Masyarakat*, *2*(1), 1–10. https://doi.org/10.30653/002.201721.11

Sanova, A., Bakar, A., & Afrida, A. (2017b). Standarisasi Instrumen Penilaian Hasil Belajar dengan Program Anates V4 Bagi Guru SMPN 17 Kota Jambi. *Jurnal Pengabdian Pada Masyarakat*, *2*(1), 1–10. https://doi.org/10.30653/002.201721.11

Siregar, H. N. I., & Panjaitan, A. (2022). A Systematic Literature Review : the Role of Classical Test Theory and Item Response Theory in Item Analysis to Determine the Quality of Mathematics Tests. *Jurnal Fibonaci: Jurnal Pendidikan Matematika*, *2*(2), 29. https://doi.org/10.24114/JFI.V2I2.31056

Stacey, K., & Turner, R. (2015). Assessing mathematical literacy: The PISA experience. *Assessing Mathematical Literacy: The PISA Experience*, 1–321. https://doi.org/10.1007/978-3-319-10121-7

Stevani, S., Eprillison, V., & Gumanti, D. (2022a). PENERAPAN APLIKASI ANATES DALAM MENGANALISIS BUTIR SOAL PADA GURU MATA PELAJARAN EKONOMI SMA YAPI PADANG. *COMMUNITY SERVICE JOURNAL OF ECONOMICS EDUCATION*, *1*(2), 26. https://doi.org/10.24014/CSJEE.V1I2.20177

Stevani, S., Eprillison, V., & Gumanti, D. (2022b). PENERAPAN APLIKASI ANATES DALAM MENGANALISIS BUTIR SOAL PADA GURU MATA PELAJARAN EKONOMI SMA YAPI PADANG. *COMMUNITY SERVICE JOURNAL OF ECONOMICS EDUCATION*, *1*(2), 26. https://doi.org/10.24014/CSJEE.V1I2.20177

Sudarto, S., Abd. Hafid, & Abd. Kadir. (2023a). TINGKAT PENCAPAIAN KEMAMPUAN GURU SD DALAM MENGGUNAKAN APLIKASI ANATES MELALUI PELATIHAN. *J-ABDI: Jurnal Pengabdian Kepada Masyarakat*, *3*(5), 793–800. https://doi.org/10.53625/JABDI.V3I5.6595

Sudarto, S., Abd. Hafid, & Abd. Kadir. (2023b). TINGKAT PENCAPAIAN KEMAMPUAN GURU SD DALAM MENGGUNAKAN APLIKASI ANATES MELALUI PELATIHAN. *J-ABDI: Jurnal Pengabdian Kepada Masyarakat*, *3*(5), 793–800. https://doi.org/10.53625/JABDI.V3I5.6595

Villarroel, V. (2018). Authentic assessment: creating a blueprint for course design. *Assessment and Evaluation in Higher Education*, *43*(5), 840–854. https://doi.org/10.1080/02602938.2017.1412396