



An Analysis of English Mid-Term Test Items at the 7th Grade Students of SMPN 7 Bukittinggi

Salsa Billa Raihana¹, Merry Prima Dewi², Syahrul³, Irwandi⁴

^{1,2,3,4} UIN Sjech M. Djamil Djambek Bukittinggi, Indonesia

Corresponding Author: ✉ salsabillaraihana29@gmail.com

ABSTRACT

ARTICLE INFO

Article history:

Received

05 August 2024

Revised

20 August 2024

Accepted

24 October 2024

Key Word

How to cite

This study was motivated by several problems, such as the mismatch between the test questions and the material learned, unclear question instructions, and the lack of statistical analysis by English teachers. The main objective of this study was to identify the quality of English midterm test items for grade 7th students at SMPN 7 Bukittinggi, focusing on Validity, Reliability, difficulty level, Discrimination Power, and Distractor Effectiveness. The research method used was Quantitative Descriptive, with data collection through student answer sheets and analysis using various statistical formulas, including Point Biserial Correlation for validity, Spearman-Brown for reliability, and Narsya's formula for difficulty level and differentiating power, as well as Djumingin's formula for effectiveness. The results showed that most of the items were valid, and reliable, had a moderate difficulty level, and had good discriminating power, with only a few items showing minor problems. Overall, the English midterm exam items were considered good quality.

Test, Test Item, Item Analysis.

<https://pusdikra-publishing.com/index.php/jsr>



This work is licensed under a
[Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/)

INTRODUCTION

In the field of education, assessment is a crucial part of the learning process. It provides teachers with crucial information about students' learning achievements, identifying their knowledge, understanding, abilities, and feelings at various stages. According Syahrul (2024) Assessment is a process of activities in the implementation of learning with the aim of knowing the extent to which students have the ability to learn. Reflinda (2018) supports this by stating that Assessment is no longer simply the achievement of learning objectives, but it is an attempt to obtain a variety of information on a regular basis, continuous and thorough process and students' learning outcomes. Therefore, proper assessment is crucial for improving students' abilities, particularly in language skills, through evaluations or exams.

In Indonesia, teachers conduct various assessments throughout a semester to evaluate students' understanding and progress, including daily tests, practical tests,

midterm tests, final tests, and national tests. Each type of test varies in content, difficulty level, timing, and assessment process. Ensuring the validity and reliability of these tests is essential to maintain their quality. Tests, as supported by Putri (2022) and Rahman (2017), serve as instruments to measure competency achievements, aid in important decisions regarding students' progress, and allow for performance comparisons. Thus, tests are necessary tools in the assessment process.

Creating effective test items requires careful consideration of the material being tested. (Jenny Indrastoeti, 2017) emphasizes the importance of aligning test questions with the taught material to avoid gaps. Teachers must adjust exam materials to the existing syllabus and teaching modules, ensuring students are not confused during exams. A thorough item analysis, as suggested by (Elviana, 2020) and (Wibawa, 2019) is necessary to test the quality of exam items. Characteristics such as validity, reliability, difficulty level, and discrimination must be assessed to create high-quality tests.

Preliminary research conducted by the researcher at SMPN 7 Bukittinggi revealed several issues with test items for 7th-grade students. There was a gap between the learned material and the test questions, confusing instructions, and a preference for multiple-choice questions among students. Interviews with students indicated that many test questions did not match the material learned, and ambiguous instructions hindered their performance. Additionally, students preferred multiple-choice questions for their straightforward nature.

Interviews with an English teacher at SMPN 7 Bukittinggi highlighted that most students scored low on tests, particularly during midterm and semester exams. The teacher revealed the absence of statistical analysis on test questions, indicating a gap in ensuring quality assessments. As a result, the researcher decided to conduct a statistical analysis of the second mid-semester test items for 7th-grade students. Thus, this study aims to evaluate the validity, reliability, and difficulty of the test questions to uncover patterns in students' difficulties and provide solutions. Consequently, the article titled "An Analysis of English Mid-Term Test Items for 7th Grade Students at SMPN 7 Bukittinggi" aims to improve test item quality for the benefit of both students and teachers.

RESEARCH METHODE

According to Sugiono (2016) research methods are a scientific approach to obtaining data with specific purposes and applications. therefore, every research must have a method used in finding data. This study adopted a quantitative descriptive research design to present an in-depth understanding of the quality of midterm exam items at SMP N 7 Bukittinggi. Within the framework of this design, the main focus of the research was to conduct item analysis using quantitative methods to assess several aspects, including difficulty level, discriminating power, and distractor effectiveness.

There are several formulas used to measure various aspects, including Validity item, reliability, item difficulty, discriminating power, and distracter effectiveness. Several formulas can be used to measure various aspects of item analysis :

1. Validity Item

To find the validity of objective items, the coefficient point biserial correlation is used.

$$gpb_i = \frac{M_p - M_t}{S_t} \sqrt{\frac{p}{q}}$$

(Rahman & Nasryah, 2019)

The point-biserial correlation coefficient obtained from the calculation is compared with the r table at significance levels of 5% and 1%, based on the number of students (with degrees of freedom: db = N - Nr). If the point-biserial correlation coefficient (rpbis) is greater than the value in the r table, the test item is considered valid. (Djumingin, 2022)

2. Reliability

There is the Formula Spearman-Brown : (Djumingin, 2022)

$$r_{\frac{11}{2}} = \frac{N \sum xy - (\sum x)(\sum y)}{\sqrt{\{N \sum X^2 - (\sum X)^2\} \{N \sum y^2 - (\sum y)^2\}}}$$

$$r_{11} = \frac{2r_{\frac{11}{2}}}{1+r_{\frac{11}{2}}}$$

Additionally, interpreting the reliability coefficient (r11) serves as a standard benchmark. If r11 is less than 0.70, it indicates that the tested material has a low reliability index or is considered unreliable.. ((Djumingin, 2022)

3. Item Difficulty

$P = \frac{B}{JS}$	In which : P: Index Difficulty B: The number of students who answered the question correctly JS: The total number of students taking the test (Rahman & Nasryah, 2019)
--------------------	--

The following is the classification of the item difficulty into three categories (Rahman & Nasryah, 2019) :

Index Difficulty	Interpretation
Less than 0,30	Difficult
0,30 - 0,70	Moderate
More than 0,70	Easy

4. Discriminating Power

There is the following formula is employed to determine the item discrimination index. (Rahman & Nasryah, 2019):

$$D = \frac{BA}{JA} - \frac{BB}{JA} = PA - PB$$

The criteria : (Rahman & Nasryah, 2019)

Discirrimination Index
0.71 - 1.00 (Excellent)
0.41 - 0.70 (Good)
0.21 - 0.40 (Satisfactory)
≤0.20 (Poor)

5. Distractor Effectiveness

The distractor index is calculated employing the following formula (Djumingin, 2022) :

$$IP = \frac{P}{N} \times 100\%$$

The distractor has functioned well if its percentage exceeds 5%

RESULT AND DISCUSSION

Findings

In this part, the researcher showed the result from the mid-term test analysis of 7th-grade students of SMPN 7 Bukittinggi.

1. Validity

In this study, the researcher utilized the Point Biserial Correlation Coefficient (GPBI formula) to assess the validity of the test items. Each test item underwent a rigorous analysis to ensure its accuracy.

Table 1.
Calculation of Validity

No. Item	gpb _i	r table		Interpretation
		5%	1%	
1	0,2373	0,1308	0,1714	Valid
2	0,3792	0,1308	0,1714	Valid
3	0,5504	0,1308	0,1714	Valid
4	0,2622	0,1308	0,1714	Valid
5	0,0436	0,1308	0,1714	Invalid

6	0,0864	0,1308	0,1714	Invalid
7	0,5076	0,1308	0,1714	Valid
8	0,3234	0,1308	0,1714	Valid
9	0,492	0,1308	0,1714	Valid
10	0,4375	0,1308	0,1714	Valid
11	0,1022	0,1308	0,1714	Invalid
12	0,4536	0,1308	0,1714	Valid
13	0,53	0,1308	0,1714	Valid
14	0,255	0,1308	0,1714	Valid
15	0,464	0,1308	0,1714	Valid
16	0,3724	0,1308	0,1714	Valid
17	0,5347	0,1308	0,1714	Valid
18	0,5056	0,1308	0,1714	Valid
19	0,495	0,1308	0,1714	Valid
20	0,279	0,1308	0,1714	Valid

Based on the analysis results, it was found that out of the 20 questions tested, 3 questions did not meet the predefined validity criteria. This suggests that these 3 questions should be revised or replaced to enhance the overall quality of the test.

2. Reliability

To evaluate the test's reliability in this study, the Spearman-Brown Formula was employed. The test is deemed reliable if $r_{11} \geq 0.70$ and unreliable if $r_{11} < 0.70$. Upon conducting a reliability analysis on the first midterm test questions for 7th-grade students at SMPN 7 Bukittinggi, a reliability index of 1.7241 was obtained. This indicates that the r_{11} value obtained is 1.7241, which exceeds 0.70. Therefore, it is concluded that the midterm exam questions for students are reliable.

3. Item Difficulty

In this study, researchers employed Narsya's Formula to assess the test's difficulty level, dividing the number of students who answered each question correctly by the total number of students.

Table 2.
Calculation of P

No. Item	B	JS	P	Classification
1	207	225	0,92	Easy
2	161	225	0,72	Easy
3	140	225	0,62	Moderate
4	158	225	0,70	Moderate

5	135	225	0,60	Moderate
6	132	225	0,59	Moderate
7	107	225	0,48	Moderate
8	112	225	0,50	Moderate
9	134	225	0,60	Moderate
10	138	225	0,61	Moderate
11	79	225	0,35	Moderate
12	113	225	0,50	Moderate
13	114	225	0,51	Moderate
14	82	225	0,36	Moderate
15	162	225	0,72	Easy
16	179	225	0,80	Easy
17	114	225	0,51	Moderate
18	161	225	0,72	Easy
19	163	225	0,72	Easy
20	101	225	0,45	Moderate

The analysis of item difficulty in this study revealed that out of the 20 questions, 6 questions were classified as overly simple, while 14 questions were categorized as having a moderate level of difficulty.

4. Discrimination Power

In the item discrimination testing section, The researchers computed the discrepancy between the number of correct answers from the higher-scoring students and the lower-scoring students.

Table 3.
Calculation of D

No Item	PA	PB	D	Interpretation
1	1,00	0,73	0,27	Satisfactory
2	1,00	0,40	0,60	Good
3	1,00	0,17	0,83	Excellent
4	0,93	0,40	0,53	Good
5	1,00	0,30	0,70	Good
6	1,00	0,10	0,90	Excellent
7	0,83	0,23	0,60	Good
8	0,80	0,17	0,63	Good
9	0,90	0,23	0,67	Good
10	0,90	0,17	0,73	Excellent
11	0,30	0,30	0,00	Poor

12	0,93	0,17	0,77	Excellent
13	0,93	0,10	0,83	Excellent
14	0,67	0,27	0,40	Satisfactory
15	1,00	0,47	0,53	Good
16	1,00	0,53	0,47	Good
17	1,00	0,20	0,80	Excellent
18	0,97	0,23	0,73	Excellent
1	1,00	0,27	0,73	Excellent
20	0,70	0,23	0,47	Good

The test results showed that there was 1 question that was categorized as "poor", 2 questions as "satisfactory", 9 questions as "good", and 8 questions as "excellent".

5. Distracter Effectiveness

For testing distracter effectiveness, researchers use Djumingin's formula. For the classification of the effectiveness of distractors, the percentage of distractor test results must be more than 5% so that it can be declared effective. In the results of this analysis, it was found that 7 distractors were declared ineffective.

Discussion

After analyzing the data, the researcher did not find any significant problems in the mid-term test of grade 7 students at SMPN 7 Bukittinggi. This shows that in general, the test is running well and there are no significant obstacles that affect the test results. Referring to Abdul theory (2017) which states that There are five essential qualities that a good test should possess, namely reliability and validity tests, difficulty level, discriminatio and distracter effectiveness.

The results of the validity test showed satisfactory results, where of the 20 questions tested, only three questions were declared invalid. This means that most of the questions can measure what should be measured accurately and consistently. Furthermore, the results of the reliability test indicate that the test is dependable, demonstrating consistent outcomes across various time frames and student groups..

This finding supports the theory proposed by Abdul Qadir, which states that a quality test should adhere to standards of validity and reliability. Thus, as a result, the mid-term test used at SMPN 7 Bukittinggi has met good quality standards by this theory, although there are still some questions that need to be improved to increase the overall validity of the test.

Additionally, the findings from the analysis of the attributes of effective test items in this study indicate that the midterm test items at SMPN 7 Bukittinggi demonstrate high quality. According to Djumingin's theory (2022), a good test item is characterized by three types: first, a test item must have a difficulty level that is "Moderate," that is,

not too difficult and not too easy. Second, the test item must have a high level of reliability. Third, the test item must have effective distractors.

In this study, researchers discovered that the difficulty level of items in the mid-term test for 7th-grade students at SMPN 7 Bukittinggi yielded positive outcomes. Of the 20 items, only six items were declared too easy, although these items need to be evaluated, overall the majority of items are categorized as "Moderate." For the item discrimination test, the analysis results also showed quite good results. Of the 20 test items, only one item was declared "Poor," two items were declared "Satisfactory," nine items were declared "Good," and eight items were classified as "Excellent." Furthermore, in testing the distractor effectiveness of the 60 distractors there are only seven distractors that are declared ineffective, in line with the theory which states that a test item must have an effective distractor with a classification exceeding 5%.

Thus, from the results of this analysis, it can be concluded that the overall quality of the English mid-term test items of grade 7 SMPN 7 Bukittinggi has been categorized as good. In the validity, reliability, item difficulty, item discrimination, and distractor effectiveness tests, no bad analysis results were found. Even though certain questions do not completely align with the criteria for good test items, the overall assessment of the test indicates satisfactory results.

The results of item analysis that showed good quality in the mid-term test items of SMPN 7 Bukittinggi were contrary to the background of the problem which previously indicated significant difficulties in test preparation and implementation. Although initially in the background of this problem, it was thought that there were obstacles in validity, reliability, and distractor effectiveness, the findings of this study show that most of the questions have met the criteria of a good test item, as described by Djumingin's theory (2022). This shows that initial concerns about the low quality of the items were not proven, and on the contrary, the analysis showed that the items had been prepared well and effectively according to the expected standards.

Interestingly, this good result was achieved even though the English teacher had not previously conducted a statistical item analysis testing process. The implication is that even if the current analysis deems the questions to be good, it is crucial for teachers to still perform statistical item analysis testing in the future. In this way, the questions given to students can be truly categorized as good and appropriate for students' abilities, ensuring that the evaluation is not only based on intuition but also supported by valid and reliable statistical data.

CONCLUSION

This research aimed to assess the quality of the mid-term exam questions for 7th-grade students at SMPN 7 Bukittinggi. The data analysis revealed that the exam showed satisfactory results in key areas of a well-constructed test, including accuracy,

consistency, question difficulty, differentiation ability, and distractor efficacy. Specifically, the validity test showed that most questions were able to measure accurately and consistently, with only three out of 20 questions being invalid. The reliability test indicated the exam consistently produced similar outcomes, demonstrating high reliability. The item difficulty analysis revealed that most questions were of moderate difficulty, while the item discrimination test showed that the majority of questions effectively distinguished between students of different abilities. Lastly, the distractor effectiveness test indicated that most distractors were effective. These results suggest that the mid-term exam questions were of good quality, contrary to initial concerns about their low quality. However, teachers need to continue conducting statistical item analysis to ensure future tests align with students' abilities.

DAFTAR PUSTAKA/ REFERENCES

- Ahmed, M. S. A. (2013). *Language Testing*. LAB Lambert Academic Publising.
- Dewi, M. (2019). *An Analysis of Item Difficulties in PBT for the Graduate Candidates of IAIN Bukittinggi*. <https://doi.org/10.4108/eai.17-10-2019.2289740>
- Djumingin, S. (2022). *Penilaian Pembelajaran Bahasa & Sastra Indonesia Teori dan Penerapannya* (Azis (ed.); Revisi Jul). CV. Cahaya BintangCemerlang.
- 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.
- Inanna, Rahmatullah, & Hasan, M. (2021). *EVALUASI PEMBELAJARAN: Teori dan Praktek*.
- Jenny Indrastoeti, S. I. (2017). *Asesmen dan Evaluasi Pembelajaran di Sekolah Dasar* (Issue August 2017).
- Noly Shofiyah, S. B. S. M. P. (2018). Buku Ajar Asesmen Pembelajaran. In *Paper Knowledge . Toward a Media History of Documents* (Vol. 5, Issue 2).
- Putri, R. (2022). *Analisis Kualitas Butir Soal Ujian Tengah Semester Mata Pelajaran Biologi Kelas X di MAN 1 Solok Selatan*. 4(2), 70–75.
- Qodir, A. (2017). *Evaluasi dan Penilaian Pembelajaran* (M. I. K. Ngalimun, M.Pd. (ed.)). K-Media.
- Rahman. (2017). The Advantages and Disadvantages of Using Qualitative and Quantitative Approaches and Methods in Language “Testing and Assessment” Research: A Literature Review. *Journal of Education and Learning*, 6(1), 102–112. <https://doi.org/10.5539/jel.v6n1p102>
- Rahman, A. A., & Nasryah, C. E. (2019). Evaluasi Pembelajaran. In *Uwais Inspirasi Indonesia*.
- Reflinda. (2018). Authentic Assessment: The Implementation of Authentic Assessment to Assess Speaking Skill at the Eighth Grade Of MTs.N 6 AGAM. *Journal of*

Educational Studies, 1(01).

Sugiono. (2016). *Metode Penelitian Kuantitatif, Kualitatif dan R&D* (Issue April).

Syahrul. (2024). Students ' Perception on the Use of Online Assessment in English Summative Tests at MAN Kota Sorong . *Journal of Linguistic and English Studies*, 1(1), 15-21.

Wibawa, E. A. (2019). Karakteristik Butir Soal Tes Ujian Akhir Semester Hukum Bisnis. *Jurnal Pendidikan Akuntansi Indonesia*, 17(1), 86-96.
<https://doi.org/10.21831/jpai.v17i1.26339>