

**ANALYSIS OF ENGLISH TEST ITEMS AT MUHAMMADIYAH
WAIPARE JUNIOR HIGH SCHOOL, MAUMERE**

Yuliana Clarita Hayong¹, Skolastika Dau², Skolastika Dau³,

Theresia Claudia Evonia Bude⁴, Elisabeth Uba Lelaona⁵

cehayong5@gmail.com¹, astindau242@gmail.com², urnitalaurensia@gmail.com³,

claudiabude04@gmail.com⁴, ubaelisabeth2@gmail.com⁵

Universitas Nusa Nipa Maumere

ABSTRACT

This research deal with the analysis of test item constructed by English as a foreign language (EFL) teachers of Muhammadiyah Waipare Junior High School, Maumere. It aims to know discrimination power, difficulty level and distracter of test items constructed by English teacher. The design used in this research is descriptive qualitative. Data resource is semester test items of grade VII students in the academic year 2025/2026. The result of this study show that discrimination power has 35,71 % test items that could not be used to test the students, 21,43 % test items that must be revised, and 42,86 % test items that are really good to be used to test the students. Difficulty index none of the test items were classified as easy 0%, 50% test items as difficult level and 50% test items as moderate level. Distracted Level involve 39,02% very poor distracter, 21,95% poor distracter, 31,70% fair distracter, 4,87% good distracter, 2,43% very good distracter.

Keywords: *Analysis, Test Item, Constructed, EFL Teacher, Resource.*

INTRODUCTION

Assessment plays a fundamental role in the teaching and learning process because it provides essential information about students' learning progress, academic achievement, and instructional effectiveness. Through assessment, teachers are able to identify students' strengths and weaknesses, monitor learning development, and make appropriate decisions to improve classroom instruction. Effective assessment also enables teachers to determine whether the intended learning objectives have been achieved and to identify areas that require further improvement. According to Hasen Allehaiby & Al-Bahlani (2021:9), assessment is a continuous process of collecting and interpreting information about students' learning to support educational decision-making. Similarly, recent studies in English language assessment emphasize that assessment has become an indispensable component of language education because it not only measures learning outcomes but also supports instructional improvement and educational quality. Therefore, assessment should be viewed as an integral part of the teaching and learning process rather than merely a tool for assigning grades.

Assessment is closely associated with evaluation because the information obtained through assessment serves as the basis for evaluating the effectiveness of teaching and learning activities. While assessment focuses on collecting evidence of students' performance, evaluation involves interpreting the evidence to determine whether educational goals have been achieved and to make informed instructional decisions. According to Tractenberg (2021:5), assessment should provide meaningful evidence that supports decision-making and promotes student learning. Therefore, assessment and evaluation complement each other in improving educational quality by providing teachers with reliable information for planning, implementing, and revising instructional practices.

Assessment is an essential component of the teaching and learning process because it provides information about students' learning achievement and supports teachers in making appropriate instructional decisions. One of the most frequently used assessment instruments in educational is the test. According to Girolamo et al., (2022:981), state that the ultimate aim of an

assessment is to help examiners make valid conclusions about an individual's skill given their performance on a particular measure. This means that a test is not only used to measure students' achievement but also to provide valid information that helps teachers evaluate students' abilities the learning process. Among various forms of educational tests, multiple-choice tests are widely used because they are practical, objective, and efficient in assessing students' achievement. According to Gonsalves, (2023:3), explain that multiple-choice assessment remains valuable because it allows objective scoring and efficient measurement of students' understanding. This means that multiple-choice tests are suitable for classroom assessment, especially when teachers need to evaluate a large number of students consistently.

However, a test cannot accurately measure students' achievement unless its items are of good quality. Therefore, test items need to be analyzed before they are used as assessment instruments. According to Dewi et al., (2022:18), item analysis is conducted to determine the quality of test items based on aspects such as validity, reliability, difficulty level, and discrimination index. This means that items analysis enables teachers to identify the strengths and weaknesses of each test item before it is administered or reused in future assessments. As a result, teachers can revise ineffective items and retain high-quality items, thereby improving the validity and reliability of classroom assessments.

Item analysis can be conducted through quantitative analysis, which focuses on the statistical characteristics of test items. According to Rezigalla et al., (2024:2), item analysis is conducted by examining several psychometric indicators, including the difficulty index, discrimination index, and distractor efficiency, to evaluate the quality of multiple-choice test items. This means that item analysis provides statistical evidence of how well each test items functions based on students' responses. Through these indicators, teachers can identify items that should be retained, revised, or discarded to improve the quality of classroom assessments.

The difficulty index is one of the main indicators used to determine the quality of test items. According to Logayah et al., (2024:2), the difficulty index and discrimination index are essential indicators for evaluating the quality of multiple-choice test items. The difficulty index indicates the proportion of students who answer an item correctly, whereas the discrimination index measures how effectively an item distinguishes between high and low achieving students. Therefore, these indicators help teachers evaluate whether test items accurately measure students' learning achievement and identify items that require improvement. Another important indicator in item analysis is the discrimination index because it shows whether a test item can effectively distinguish students based on their level of achievement. According to Khare et al., (2023:1039), the discrimination index is one of the most important tools for assessing the quality and validity of multiple-choice items because it indicates an items ability to discriminate between students with high and low test scores. This means that test items with high discrimination indices function effectively in differentiating high-achieving students from low-achieving students. Therefore, items with good discrimination indices should be retained, whereas items with low discrimination indices should be revised or discarded to improve the quality of classroom assessments.

Conversely, items with low discrimination power may fail to measure students' actual abilities and should be revised or discarded. In addition to difficulty index and discrimination power, distractor effectiveness is also an essential aspect of multiple-choice test quality. According to Chauhan et al., (2023:2), effective distractors contribute to appropriate difficulty and discrimination indices in multiple-choice tests. This means that plausible distractors encourage students who do not know the correct answer to select alternative options, thereby improving the quality and accuracy of the assessment. Distractors that are rarely selected by students are considered ineffective because they fail to perform their intended function. Therefore, evaluating distractor effectiveness helps improve the reliability and validity of multiple-choice tests.

Several previous studies have investigated the quality of teacher made English test items using item analysis to determine whether the items function effectively as assessment instruments. According to Darmawan et al., (2022:499), although some English teacher made test items were found to be valid and reliable, several items still had poor discrimination indices and inappropriate difficulty levels. This indicates that not all teacher made test items are capable of accurately measuring students' learning achievement. Therefore, item analysis is necessary to identify ineffective items and revise them before they are used in classroom assessment. Similarly, Rame et al., (2024:61), found that several English multiple-choice test items did not meet the expected standards in terms of difficulty level, discrimination index, and distractor effectiveness. This means that some test items still require revision to improve their quality and ensure that they accurately measure students' learning achievement. Furthermore, Noviasmy & Risma, (2024:48), concluded that the teacher-made English summative test generally had good item validity, an acceptable level of difficulty, and effective distractors, although several items still needed revision before being reused. This suggests that item analysis provides valuable information for teachers to improve the quality of classroom assessments by identifying which items should be retained, revised, or discarded. Likewise, Logayah et al., (2024:2), emphasized that evaluating test items based on the difficulty index, discrimination index, and distractor efficiency provides important information about the quality of assessment instruments. This indicates that comprehensive item analysis helps teachers improve test quality and ensure that classroom assessments accurately measure students' learning achievement. These findings indicate that the quality of English teacher-made test items remains an important issue in educational assessment and requires continuous evaluation to ensure valid and reliable measurement of students' achievement.

Although previous studies have examined test item quality in different educational contexts, limited studies have focused on English multiple-choice test items at junior high schools in Maumere, particularly at Muhammadiyah Waipare Junior High School. In addition, many previous studies were conducted in different regions and educational settings, making it necessary to investigate whether similar issues also occur in this context. Therefore, this study provides a specific analysis of English test items used at Muhammadiyah Waipare Junior High school, Maumere.

Based on the research gap above, this study aims to analyse the quality of English multiple-choice test items administered to seventh-grade students at Muhammadiyah Waipare Junior High School, Maumere. The analysis focuses on three indicators, namely difficulty index, discrimination power, and distractor effectiveness. The findings of this study are expected to provide useful information for teachers in improving the quality of English assessment instruments and developing more effective test items in the future.

METHOD

This study employed a descriptive quantitative research design. The study was conducted at Muhammadiyah Waipare Junior High School. The data source of this study was an English multiple-choice test administered to seventh grade students. The test consisted of fourteen items and was taken by fourteen students. The students' answer sheets were used as the primary data for the item analysis.

This study followed several procedures. First, the researcher scored the students answers based on the answer key. Second, the student scores were arranged from the highest to the lowest and divided into upper and lower groups. Third, the researcher calculated the Difficulty Index of each item to determine whether the items were categorized as easy, moderate, or difficult. Fourth, the researcher calculated the Discrimination Power to determine whether the items should be categorized as good, revised, or dropped. Finally, the researcher analyzed the Distractor Effectiveness of each item to identify whether the distractors functioned effectively.

RESULT AND DISCUSSION

In this section, the researcher presents the result of the items analysis of the English multiple-choice test administered to seventh grade students at Muhammadiyah Waipare Junior High School. The test consisted of 14 items answered by 14 students. The analysis was conducted to determine the quality of the test items based on three indicators, namely Discrimination Power, Difficulty Index, and Distractor Effectiveness. The result of the analysis are presented in the following sections.

Tabel 1. Discrimination Power

NO	DP	DECISION
1	0,28	Good
2	0,28	Good
3	-0,14	Dropped
4	0,14	Corrected
5	0,00	Dropped
6	0,57	Good
7	0,28	Good
8	0,00	Dropped
9	0,14	Corrected
10	0,00	Dropped
11	0,28	Good
12	0,14	Corrected
13	-0,14	Dropped
14	0,57	Good

Tabel 2. Difficulty Index

NO	DI	DECISION
1	0,57	Moderate
2	0,28	Difficult
3	0,21	Difficult
4	0,21	Difficult
5	0,28	Difficult
6	0,28	Difficult
7	0,57	Moderate
8	0,50	Moderate
9	0,42	Moderate
10	0,50	Moderate

11	0,42	Moderate
12	0,28	Difficult
13	0,21	Difficult
14	0,42	Moderate

Table 2 presents the difficulty index of the English test items administered to seventh-grade students at Muhammadiyah Waipare Junior High School. The analysis shows that 7 out of 14 test items (50%) were categorized as difficult items. These items were found in item numbers 2, 3, 4, 5, 6, 12, and 13. Furthermore, 7 test items (50%) were categorized as moderate items, including item numbers 1, 7, 8, 9, 10, 11, and 14. No test items were categorized as easy. Therefore, the findings indicate that the test consisted mainly of moderate and difficult items, which may provide an appropriate level of challenge for seventh-grade students.

Distracter

The result of distracter analysis of 14 test items administered to seventh-grade students of Muhammadiyah Waipare Junior High School is in the following analysis:

Item number 1

$$DtI = \frac{nDt}{(N-nC)/(Alt-1)}$$

$$\text{Option A} = \frac{3}{\frac{(14-8)/(4-1)}{6/3}} = \frac{3}{\frac{2}{2}} = 1,5$$

$$\text{Option C} = \frac{2}{\frac{(14-8)/(4-1)}{6/3}} = \frac{2}{\frac{2}{2}} = 1$$

$$\text{Option D} = \frac{1}{\frac{(14-8)/(4-1)}{6/3}} = \frac{1}{\frac{2}{2}} = 0,5$$

Test number 2 until 14 are analysis as the example of the analysis above. It present on the following table:

Table 3. Distracter

NO	A	B	C	D
1	1,5	*	1	0,5
2	0,60	1,81	*	0,60
3	1,1	*	1,6	0,27
4	*	0,27	1,38	1,38
5	1,51	*	0,60	0,90
6	1,51	0,60	*	0,90
7	1	*	1,5	0,5
8	0,86	1,73	*	0,43
9	*	1,15	0,76	1,15
10	2,17	*	0,86	0
11	0,76	1,15	*	1,15
12	1,51	0,90	*	0,60
13	1,6	*	0,83	0,5
14	0,76	*	1,53	0,76

The results of the distracter analysis above were interpreted by using the standardized distracter levels shown in the following table:

Tabel 4. Distracter level

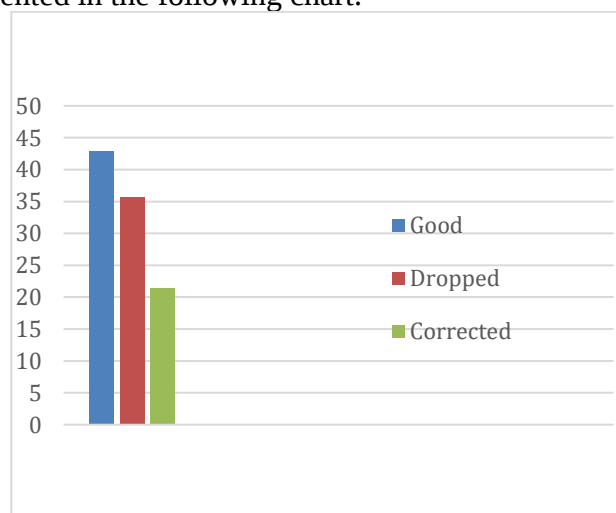
Level	A	B	C	D	Σ	%
Very Poor	1	0	0	0	16	39,02
Poor	0	1	0	1	9	21,95
Fair	4	2	2	5	13	31,70
Good	2	1	3	3	2	4,87
Very Good	5	2	4	5	1	2,43
	12	6	9	14	41	100

Tabel 4. Show that there are 16 of 41 or 39,02% very poor, 9 of 41 or 21,95% poor distracter, 13 of 41 or 31,70% fair distracter, 2 of 41 or 4,87% good distracter and 1 of 41 or 2,43% very good distracter.

In this section, the results of the data analysis are described and discussed. The findings obtained through the standardized formulas are explained and interpreted. The discussion focuses on three aspects of item analysis, namely Discrimination Power, Difficulty Index, and Distractor Effectiveness of the English multiple-choice test administered to seventh grade students at Muhammadiyah Waipare Junior High School. The result of the analysis are presented in the following:

Discrimination Power

Discrimination power is used to distinguish between high-achieving students and low-achieving students in answering the test items. It is calculated based on the responses of both upper and lower groups. The purpose of this analysis is to determine whether the test items are categorized as good, need revision, or should be dropped. The results of the Discrimination Power analysis are presented in the following chart.

**Chart 1. Discrimination Power**

The chart shows that there are 5 of 14 test items constructed by EFL teachers that must be dropped. It involves 35,71 % test items that could not be used to test the students. In addition, there are 3 test items or that must be corrected or revised. It involves 21,43 %. Thus, there are 6 of 14 test items that are really good to be used to test the students. It includes 42,86 %.

Difficulty index

The difficulty index represents the percentage of students in both the high and low groups who answered each test items correctly. This Index is used to classify test items into three

categories: easy, moderate, and difficult. The result of the difficulty index analysis for the test items constructed by EFL teacher are illustrated in the following chart.

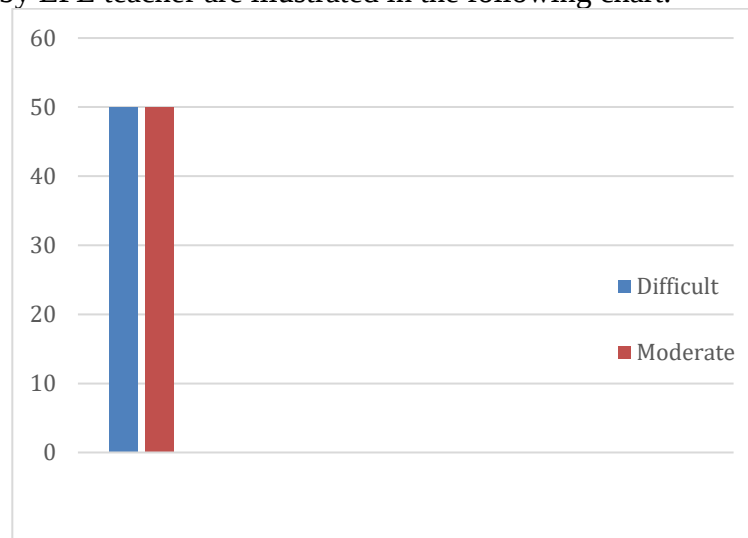


Chart 2. Difficulty Index

The difficulty index chart above indicates that none of the test items constructed by the EFL teacher were classified as easy 0%. There are 7 test items involves difficulty level or 50% of 14 test items, and there are 7 test items involves moderate level or 50% of 14 test items.

Distracter

Distracter level refers to incorrect answer options provided in multiple choice test items. The effectiveness of distracters be observed through the difficulty index and discrimination index. Ineffective distracter may reduce the quality of test items by resulting in poor discrimination and inappropriate levels of difficulty. Therefore, distracters are generally classified into five categories: very poor, poor, fair, good, and very good. The quality of the distracters in the test items constructed by the EFL teacher is presented in the following chart.

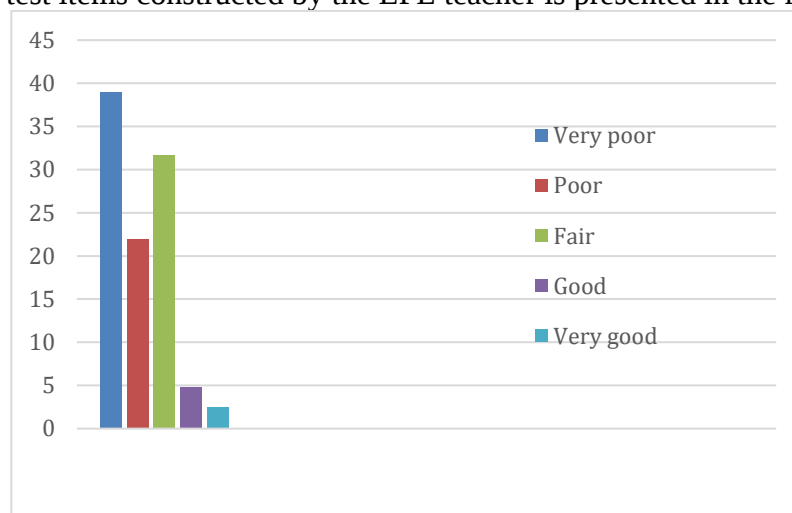


Chart 3. Distracter

The chart show that distracter of test items constructed by EFL teachers are 16 of 41 or 39,02% very poor, 9 of 41 or 21,95% poor distracter, 13 of 41 or 31,70% fair distracter, 2 of 41 or 4,87% good distracter and 1 of 41 or 2,43% very good distracter.

CONCLUSION

This study was limited to the analysis of the English test items developed by the EFL teacher at Muhammadiyah Waipare Junior High School, Maumere, in the 2025/2026 academic year. Based on the result of the test items analysis, it was found that only 6 out of the test items,

representing 42,86%, met the criteria of good quality test items and were considered appropriate for assessing students learning achievement. Therefore, teacher as test developers should pay careful attention to important aspects of test constructions, including the discrimination index, difficulty index, and distracter effectiveness. One strength of this study is that the analysis was conducted using established criteria for item analysis, making the findings useful as a reference for English teacher and other test developers in evaluating and improving the quality of their test items.

REFERENCES

- Chauhan, G. R., Chauhan, B. R., Vaza, J. V., & Chauhan, P. R. (2023). Relations of the Number of Functioning Distractors With the Item Difficulty Index and the Item Discrimination Power in the Multiple Choice Questions. *Cureus*, 15(7). <https://doi.org/10.7759/cureus.42492>
- Darmawan, M., Riyanti, D., Yuliana, Y. G. S., & Sumarni. (2022). A Test-Items Analysis of English Teacher-Made Test Corresponding email. *Journal of English Education and Teaching (JEET)*, 6(4), 498–513.
- Dewi, W. A., Lestari, S. F., Mirza, A. N., & Ratmadiyah, R. J. (2022). Item Analysis of Multiple-Choice Questions of Mid-Semester Assessment of Mathematics in the Fifth-Grade Elementary School in Wates District Year 2021/2022. *DIDAKTIKA: Jurnal Pendidikan Sekolah Dasar*, 5(1), 17–24. <https://doi.org/10.21831/didaktika.v5i1.46229>
- Girolamo, T., Ghali, S., Campos, I., & Ford, A. (2022). Interpretation and Use of Standardized Language Assessments for Diverse School-Age Individuals. *Perspectives of the ASHA Special Interest Groups*, 7(4), 981–994. https://doi.org/10.1044/2022_persp-21-00322
- Gonsalves, C. (2023). On ChatGPT: what promise remains for multiple choice assessment? *Journal of Learning Development in Higher Education*, (27). <https://doi.org/10.47408/jldhe.vi27.1009>
- Hasen Allehaiby, W., & Al-Bahlani, S. (2021). Applying Assessment Principles during Emergency Remote Teaching: Challenges and Considerations. *Arab World English Journal*, 12(4), 3–18. <https://doi.org/10.24093/awej/vol12no4.1>
- Logayah, D. S., Ruhimat, M., Arrasyid, R., & Islamy, M. R. F. (2024). Item analysis of National Geography Olympiad multiple-choice questions (MCQs) in Indonesia. *Cogent Social Sciences*, 10(1). <https://doi.org/10.1080/23311886.2024.2354971>
- Noviasmy, Y., & Risma. (2024). An Analysis of The English Summative Test: EFL Teacher-Made Test. *Jelita*, 5(1), 41–50. <https://doi.org/10.56185/jelita.v5i1.618>
- Pendidikan, P. P. (2013). *Panduan Analisis Butir Soal*No Title. Badan Penelitian dan Pengembangan, Kementerian Pendidikan dan Kebudayaan.
- Rame, G., Hoinbala, F. R., Kristen, U., & Wacana, A. (2024). Item Analysis of Multiple-Choice Questions : an. 2(1), 58–65. <https://doi.org/10.1357/l.v2i1.45>
- Rezigalla, A. A., Eleragi, A. M. E. S. A., Elhussein, A. B., Alfaifi, J., ALGhamdi, M. A., Al Ameer, A. Y., Yahia, A. I. O., Mohammed, O. A., & Adam, M. I. E. (2024). Item analysis: the impact of distractor efficiency on the difficulty index and discrimination power of multiple-choice items. *BMC Medical Education*, 24(1), 1–7. <https://doi.org/10.1186/s12909-024-05433-y>
- Suhas Khare, A., Verma, A., Akshay Kadam, A., & Avachar, K. (2023). Evaluation Of Discrimination Index Of Single Best Response Type Of Multiple Choice Questions In Physiology By Item Analysis INTRODUCTION. *Journal of Pharmaceutical Negative Results* , 14(2), 2023. <https://doi.org/10.47750/pnr.2023.14.S02.126>
- Tractenberg, R. E. (2021). The assessment evaluation rubric: Promoting learning and learner-centered teaching through assessment in face-to-face or distanced higher education. *Education Sciences*, 11(8). <https://doi.org/10.3390/educsci11080441>