

## Editorial Team

### **EDITOR IN-CHIEF**

- Assoc. Prof. Dr. Drs. Abubakar Ajalil, M.Si, SCOPUS ID. [58634461600](#), Universitas Serambi Mekkah, Indonesia

### **MANAGING EDITOR**

- Dr. Dian Aswita, S.Pd, M. Pd, Universitas Negeri Makasar, ID SCOPUS : [57202957850](#), Indonesia

### **SECTION EDITORS**

- Prof. Dr. Magdalena Mo Ching Mok, M. Ed, Educational University of Hongkong, ID SCOPUS 7006024212, Hong Kong
- Dr. Asriani, S. Pd., M. Pd, Universitas Serambi Mekkah, Indonesia
- Dr. Hj. Rani Siti Fitriani, S.S., M. Hum, Universitas Pasundan, Bandung, Indonesia
- Dr. Wahyu Khafidah, S.Pd.I, MA, Serambi Mekkah University, Indonesia
- Dr. Usman Effendi, S.Sos., MM, Universitas Persada Indonesia YAI Jakarta, Indonesia, Indonesia
- Dr. Arfriani Maifizar S,E, M.Si., Universitas Teuku Umar Aceh Barat, Indonesia, ID SCOPUS [57210744149](#), Indonesia
- Zhao Jing, M. ED, Gizhou Education University, China, China
- Nurlaili Ramli, S. SiT., MPH, Health Polytechnic of the Ministry of Health in Aceh, Aceh Besar. ID SCOPUS [57195919249](#), Indonesia
- Zaiyana Zaiyana Putri, Universitas Serambi Mekkah, ID SCOPUS [57211267424](#), Indonesia
- Fitri Wulandari, S.Pd., M. Hum, Universitas Islam Riau, ID SINTA 6704089, Indonesia
- JUNAIDI S, PD., M.PD., Universitas Serambi Mekkah, Indonesia
- Said Ali Akbar, S. Pd., M. Si, Universitas Serambi Mekkah, Banda Aceh ID SCOPUS [57190374979](#), Indonesia
- Muhammad Fajrin Pane, SH.I., M. Hum, Politeknik Tanjung Balai, Sumatera Utara, Indonesia
- Anita Noviyanti, S. Pd., M. Pd, Universitas Serambi Mekkah, Banda Aceh, Indonesia, ID SCOPUS [57219092073](#), Indonesia
- Drs. Burhanuddin AG,. M. Pd, Universitas Serambi Mekkah, Aceh Indonesia, ID SCOPUS [57219343469](#), Indonesia
- Drs. Jailani, M. Pd, Universitas Serambi Mekkah ID SCOPUS [57219098536](#) Indonesia
- Drs. Ridhwan Ismail, M. Pd, Universitas Serambi Mekkah ID SCOPUS [57219091724](#), Indonesia
- Dr. Hj. Israwati, M. Si, Universitas Syiah Kuala, ID SCOPUS [57211263956](#), Indonesia

- Drs. Yulsafli - MA, Universitas Serambi Mekkah, ID SCOPUS , Indonesia
- Drs. Anwar S. Pd., M. Pd, Universitas Serambi Mekkah, Banda Aceh ID SCOPUS [58634699300](#), Indonesia
- Drs. Muhammad Isa, M. Pd, Universitas Serambi Mekkah, Aceh ID SCOPUS [57205735891](#), Indonesia
- Prof. Mahendran, P.hD, Universitas Pendidikan Sultan Idris, Malaysia
- Dr. J. Karthikeyan, Ph.D, National College, Tiruchirappali, India
- Sophia Manning, Ph.D, Kean University New Jersey, USA
- Dra. Hj. Ismawirna, M. Pd, Universitas Serambi Mekkah, [ID SCOPUS](#), Indonesia
- Dra. Hj. Armi, M. Si, Universitas Serambi Mekkah, ID SCOPUS [57219094630](#), Indonesia
- Muhammad Aulia, S.Pd., MTSOL,.MA.(Res)., Ph.D. ID Scopus ID [58785862800](#) Universitas Syiah Kuala, Indonesia
- Septhia Irnanda, S.Pd., M.Tsol., Ph.D, ID Scopus [57209573672](#), Universitas Serambi Mekkah, Indonesia
- Dr. Soetam Rizky Wicaksono, M.M, ID Scopus [57209459047](#), Machung University, Indonesia
- Dr. Lutfi Yondri, M.Hum. ID Scopus [24391756000](#), Kajian Budaya dan Arkeologi Indonesia
- Kamarullah, S. Pdi., M. Pd, Universitas Muhammadiyah Mahakarya Aceh, ID Scopus [58577051200](#), Indonesia
- Teuku Afriliansyah, Universitas Bumi Persada, ID Scopus [57200726172](#), Indonesia
- Suci Maulina, MA, Universitas Jabal Ghafur, ID Scopus [57204472764](#), Indonesia
- Dr. Cut Nya Dhin, S. Pd., M. Pd, Universitas Islam Ar-Raniry, Banda Aceh, Indonesia
- WEB AND OJS MANAGER**
- Munawir Munawir, ST,. MT, Universitas Serambi Mekkah, ID SCOPUS [57194214483](#) Indonesia

### **ADMINISTRATOR OFFICE TEAM**

- Dra. Ismawirna M. Pd, Universitas Serambi Mekkah, Banda Aceh, Indonesia. ID. SCOPUS 57463492600,. ID SINTA 6167918, Indonesia
- Dra. Armi M, Si, Universitas Serambi Mekkah, Aceh. Indonesia ID SCOPUS [57219094630](#), Indonesia
- Said Ali Akbar, S. Pd., M. Si, Universitas Serambi Mekkah, Banda Aceh ID SCOPUS 57190374979, Indonesia

### **ENGLISH LANGUAGE ADVISORS**

- Septhia Irnanda, S.Pd., M.Tsol., Ph.D, Universitas Serambi Mekkah, Aceh ID SCOPUS 5720957372, Indonesia
- Sabrina, S. Pd., M. Appling., M. Tran, Universitas Serambi Mekkah, Banda Aceh, Indonesia
- Muhammad Aulia, S.Pd., MTSOL,.MA.(Res)., Ph.D, Syiah Kuala University, Aceh, ID SCOPUS, [58785862800](#), Indonesia

### **LAYOUT EDITORS**

- Samsuddin Samsuddin, Program Studi Teknik Komputer - Universitas Serambi Mekkah
- Dr. Nasir Ibrahim, SE., M. Si, Universitas Serambi Mekkah, Indonesia
- Dr. Hj. Elvitriana, Universitas Serambi Mekkah, Indonesia
- Firdaus, Designer Grafis Zoom Printing, Aceh, Indonesia

## **PROOFREADERS**

- Prof. Dr. Asnawi Abdullah, BSc.PH, MHSM, MSc.HPPF, DLSHTM, Ph.D, Universitas Muhammadiyah, Aceh, ID SCOPUS : 57202957850, Indonesia
- Ery Utomo, P.hD, Universitas Negeri Jakarta
- Muslem Daud, S. Ag., M. Ed., Ph.D, Universitas Serambi Mekkah, Aceh, Indonesia, Indonesia
- Dr. Faradiba Sari Harahap, S. Pd., M. Pd, Politeknik Tanjung Balai, Sumatera Utara, Indonesia
- Dr. Muhammad Subhan, Ph.D., M.Sc., B.Eng., MLogM, Aff.M.ASCE, King Abdul Aziz University, Saudi Arabia
- Muhammad Aulia, S.Pd., MTSOL,.MA.(Res)., Ph.D, Syiah Kuala University, Aceh, ID SCOPUS [58785862800](#), Indonesia
- Exkarach Denang, M. Ed., Ph,D, Udom Tani University, Thailand
- Sabrina, S. Pd., M. Appling., M. Tran, Universitas Serambi Mekkah, Banda Aceh, Indonesia
- Prof. Yunisrina Qismullah Yusuf, S. Pd., M. Ed., Ph.D, Universitas Syiah Kuala, Aceh, ID SCOPUS : [55351138500](#), Indonesia
- Dr. H. Muhammad Alfatih Suryadilaga, S.Ag., M. Ag, Universitas Islam Negeri Sunan Kalijaga, Depok, Indonesia
- Sukri Adani, S. Pd., M. Pd, STKIP Muhammadiyah Abdiya, ID Sinta 5984339, Indonesia

## **Indexed by :**

1. SINTA 4,
2. Base,
3. Copernicus,
4. ONESearch,
5. Demession,
6. Moraref
7. Garuda,
8. Crossref,
9. Copernicus,
10. WordCat,
11. CiteFactor,
12. ISJD
13. Scilit,
14. Europub,
15. Advance Science Index

## The Analysis of English Teacher-Made Tests for The Students in SMP Muhammadiyah Palopo

Asman<sup>1</sup>, Husnani Aliah<sup>2</sup>, Nasriandi<sup>3</sup>

<sup>1</sup>Asman is A Students of University of Muhammadiyah Palopo, Palopo, Indonesia  
Email: [asman.taslim99@gmail.com](mailto:asman.taslim99@gmail.com)

<sup>2</sup>Husnani Aliah is A Lecturer of University of Muhammadiyah Palopo, Palopo,  
Indonesia

Email: [husnani@umpalopo.ac.id](mailto:husnani@umpalopo.ac.id)

<sup>3</sup>Nasriandi is A Lecturer of University of Muhammadiyah Palopo, Palopo, Indonesia  
Email: [nasriandi@umpalopo.ac.id](mailto:nasriandi@umpalopo.ac.id)

### Abstract

This study aimed to analyze the quality of English teacher-made tests used in the second semester exam of the 2023/2024 academic year at SMP Muhammadiyah Palopo. The validity, reliability, difficulty level, differentiation index, and distractor strength of the test were analyzed. Data were collected through teacher-made multiple-choice tests as well as student answer sheets. This study used the ANATES program to analyze the test results. The results show that most of the questions were in the very easy category with a percentage of correct answers of more than 85%. Only a few questions have a medium or very difficult level of difficulty. In addition, the differentiating power of most questions was low, indicating the lack of ability of questions to distinguish between students with high and low abilities. Distractors on many questions also proved ineffective in directing students to choose the wrong answer. Overall, these findings indicate the need for improvements to the exam questions, especially in terms of differentiating power and the effectiveness of distractors to improve the overall quality of the test.

Keywords: test quality, teacher-made test, multiple choices

### INTRODUCTION

Teaching and learning process at school has to follow several things to get successful. Before teaching and learning process, the teacher has to consider about designing curriculum, determining teaching and learning objectives, need analysis, and preparing materials (Indriyani et al., 2023) . The next process is the teaching and learning process, including designing lesson plans and teaching practice. The final process is the learning and teaching evaluation. In order to measure students' understanding, teacher needs to conduct a test. Testing can be conducted before, during, and after teaching. (Agnes Meilina, 2020) defines test as an instrument or systematic procedure for observing and describing one or more characteristics of a student using

either a numerical scale or a classification scheme. The test is scored by adding the students' points from each question. In this term, students are described using numerical scale.

After conducting a test, it is very important to conduct item and test analysis. These analyzes evaluate the quality of the items and of the test as a whole. Such analysis can also be employed to revise and improve both items and the test as a whole (Iksan & Dirham, 2018; Irvy, 2020). Nowadays, teacher-made tests are also used as a complementary aspect that determines whether students can pass or not in the national examination (Meni Handayani, SS et al., 2020). The teacher-made tests are considered to have more ability to show students understanding, but then it gets some critical comments, because some researchers claimed that the test is not appropriate for use. Therefore, we need to conduct item and test analysis to see and prove the quality of the test. From a conceptual standpoint, distinct assessments fulfill distinct goals, and it is thought that no one assessment can be created and applied for several reasons (Meni Handayani, SS et al., 2020). There are several kinds of tests; depending on their approach, goal, and type of response that pupils can get. The reading comprehension exams' multiple-choice (MC) format is one of them. (Juniardi, 2023).

It is very important for the teacher in particular an English teacher to know how to construct a good test. Constructing a good quality English test instrument, especially for multiple-choice tests, is definitely not easy. A trial run must be applied on the freshly designed test before the instrument is used. Therefore, the analysis will always be needed to evolve the quality of the English test. To review and revise the English tests instrument mostly are designed and used by the teachers as stated above, therefore, the researcher was interested to conduct a research under the title "The Analysis of English Teacher-Made Tests for the Students in SMP Muhammadiyah Palopo."

## METHOD

This study used survey research with descriptive method. This study aims to provide an overview of the quality of the English test used in the exam test in the second semester of the 2023/2024 school year at SMP Muhammadiyah Palopo, including validity, reliability, item difficulty level, differentiating power index, and differentiating power. This researcher analyzed the data and then described the research findings quantitatively. This research variable is test quality. In terms of test quality, it consists of several sub-variables. The overall quality of the test is analyzed based on validity and reliability, item difficulty level, differentiating index, and the test's power. The population of this study is teachers who teach in the second year at SMP Muhammadiyah Palopo. There were five English teachers in Muhammadiyah Palopo Junior High School. This study used random sampling technique by taking one teacher as the sample.

The data collected from teachers consists of teacher-made tests, namely multiple-choice tests and student answer sheets. The quality of the test was analyzed using a software called ANATES. ANATES is a computer application program that can

be used to analyze multiple choice tests. ANATES is easy to learn and use. The facilities owned by this program are a scoring system, determination of upper and lower groups, reliability analysis, validity analysis, distractor analysis, calculating item difficulty, and differentiator index.

In determining test validity, the coefficient of each test item was compared with r-table to find out the degree of significance. If the correlation is bigger than r-table value, it means that correlation is significant. But if the correlation is smaller than r-table, it is assumed to be insignificant. Furthermore, the test is assumed to be highly reliable if its coefficient is greater than 0.70. If its coefficient is lower than 0.70, it means the test has low reliability. In order to get the detailed information about the quality of the test, item difficulty of each test was analyzed. The level of item difficulty is too easy, easy, fair, difficult, and too difficult items. A good test should contain easy, fair, and difficult items, not just focus on one level of difficulty. Too easy or too difficult items are not acceptable to be used in a test. The following criterion is used to determine item difficulty.

Table 1  
Item Difficulty Interpretation

| Item difficulty | Interpretation |
|-----------------|----------------|
| 0 – 15%         | Too difficult  |
| 16% - 85%       | Acceptable     |
| 86% - 100%      | Too easy       |

Source: Thesis of Husnani Aliah, 2014

The next analysis was item discrimination. It was determined the effectiveness of each item to differentiate students based on their knowledge. The maximum item discrimination difference is 100%. It is occurred when all students in upper group answered correctly and all students in lower group answered incorrectly. The detailed criteria that were used to determine the interpretation of discrimination index can be seen in the following table.

Table 2  
Discrimination Index Interpretation

| Discrimination index | Interpretation |
|----------------------|----------------|
| Negative – 0.29      | Improper       |
| 0.30 – 1.00          | Proper         |

Source: Thesis of Husnani Aliah, 2014

The other important element in determining the quality of the test items is distractor power. In this stage, the performance of each incorrect option was analyzed to find out effective and ineffective distractors. A distractor can be classified as working properly if it is selected by at least 5% of students for 3 answer choices and 3% of students for four answer choices.

## RESULT AND DISCUSSION

Analysis of question items before they are used as guidelines in learning evaluation assessments is the main thing. The reason is, this analysis activity is an

effort to maximize the quality of the questions that will be given to students. It is hoped that the existence of quality questions can cover various levels of student competency abilities. The purpose of this analysis is to examine and examine each question item that will be presented to students as a guide for learning assessment. The results of this research reveal the quality of the test questions by explaining the validity of the test items, the distinguishing power and the level of difficulty of the questions.

### 1. Validity

Validity is determined by the correlation of item scores with the total score and compared with the  $r$  table to see the degree of significance. If an item has a high correlation, it means the item has higher validity. Conversely, if the correlation is smaller than  $r$  table, it means that the validity of the item is low.

Table. 3.  
Result of Validity Test

| School | Grade | Category  | Item Number | percentage |
|--------|-------|-----------|-------------|------------|
| SMP    | IX    | Valid     | 30 items    | 27%        |
|        |       | Not valid | 0 items     | 0%         |

Source : Research results2024

Based on the analysis of the data provided, the overall validity of the test has been evaluated using the discrimination index ( $D$ ) for each item. Students were divided into two groups, namely the upper group (27% of students with the highest score) and the lower group (27% of students with the lowest score). The analysis results show that all items have a discrimination index ( $D$ ) that varies from 0.25 to 0.5, with the following categories: Very Good, Good, and Fair. Overall, of the 30 items analyzed, all of them fall into the valid category, which means that each item in this test is able to differentiate between students with high and low abilities effectively. Thus, the overall validity percentage of this test is 100%. This shows that all the items in this test are of good quality and suitable for use to measure student abilities.

From the data provided, there is no information about the answers per item from each student, so the validity of the items cannot be calculated specifically. However, this theoretical explanation provides an overview of how the process of calculating item validity is carried out. For a more accurate analysis, data on answers per item from each student is needed. With this information, the validity of each item can be determined, which will indicate how well each item in the test is able to differentiate between high and low ability students.

### 2. Reliability

Another important characteristic of a measurement procedure is reliability. Reliability is described as the degree of consistency of a test in measuring student abilities. Reliability is calculated based on student answer sheets. According to

Sudijono (in Jabu, 2008: 124), the acceptable degree of test reliability is 0.70. If the result is greater than 0.70, it means the test is very reliable. Conversely, if the result is smaller than 0.70, it means the test is less reliable. The following table shows the reliability level of each test are average is 22.60, Standard Deviation= 3.21, CorrelationXY= 0.55, Test Reliability= 0.71.

Based on the data, it can be observed that the scores for each subject in the teacher-made tests show significant variations. Although some subjects such as Bilal al-Gifari and Muh. Danis DI recorded a high total score (26 and 25 respectively), there were also other subjects who recorded lower scores such as Baya Ardiansyah with a total score of 15. This variation shows that this test may have varying levels of validity depending on test subjects. To evaluate its validity further, statistical analysis such as calculating the correlation between items and coefficient alpha (Cronbach's  $\alpha$ ) can be carried out. It is important to ensure that the test can consistently measure the intended ability or knowledge across all subjects tested, minimizing the potential for bias or uncertainty in the evaluation results. Another important characteristic of a measurement procedure is reliability.

The results of reliability calculations from the data above using the split-half method are as follows:

- a. Pearson correlation between odd scores and even scores: Correlation value (r): 0.552
- b. Reliability coefficient using the Spearman-Brown formula: Reliability value: 0.711

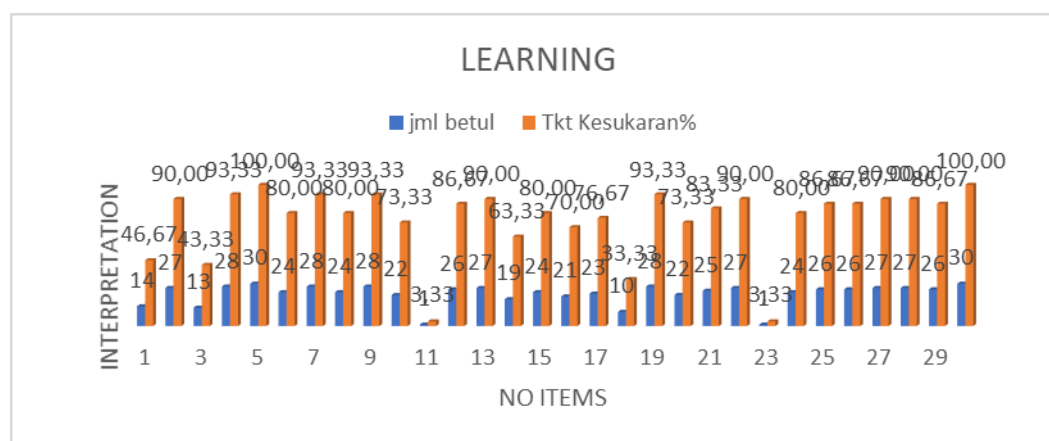
This test has quite good reliability with a reliability coefficient value of 0.711. Although there is room for improvement, the test has demonstrated adequate consistency in outcome measurement.

### 3. Item difficulty

Based on the table of difficulty levels of the questions provided, it can be concluded that of the 30 questions, the majority of the questions, namely 19 questions, are classified as very easy with a percentage of correct answers of more than 85%. The other six questions are relatively easy with a percentage of correct answers between 70% and 85%. Apart from that, there are seven questions that fall into the medium category with a percentage of correct answers between 30% and 70%. Only two questions were classified as very difficult with a correct answer percentage of less than 30%. This distribution of item difficulty levels suggests that exams tend to have too many very easy questions, which can reduce the exam's ability to differentiate between students with different understandings of the material. To improve the quality of the exam, it is recommended to review very easy questions so that the level of difficulty can be increased or replaced with questions that are more diverse in level of difficulty. This will help create better balance in the test and increase its discriminatory power. Number of Subjects= 30, Question Items= 30

Picture 1

Item Difficulty of the Test



Source : Research results2024

Fatimah (2019) also discusses the level of difficulty of questions in the educational context. According to him, the difficulty level of the questions is calculated as a percentage of the number of students who answered the questions correctly or incorrectly. This is similar to the definition from Arikunto (2010), but the emphasis can vary depending on the context or purpose of the question analysis. Based on the data provided, we can analyze the difficulty level of the questions by using the percentage of correct answers for each question number. The following is an analysis of the level of difficulty based on the data provided:

Table.4

Degree of difficulty:

| Interpretation | Percentage of Correct Answers (%) | Number of Questions |
|----------------|-----------------------------------|---------------------|
| Very easy      | > 85                              | 14                  |
| Easy           | 70 - 85                           | 8                   |
| Currently      | 50 - 70                           | 1                   |
| Difficult      | < 50                              | 7                   |
| Very difficult | < 30                              | 1                   |

Source : Research results2024

From this analysis, the majority of questions were considered easy or very easy, with a few exceptions considered moderate or very difficult. This suggests that the majority of students will probably be able to answer these questions well, but there are some areas that may need further attention to increase the level of difficulty and diversity of questions.

#### 4. Discrimination index

The discrimination index shows the ability of an item to differentiate high achievers from low achievers. A high discrimination index is achieved if most of

the questions are answered by high achievers. Conversely, a low discrimination index is found if most of the questions are answered by those with lower achievement. Based on the analysis results, the discrimination index is tabulated as

Number of Subjects= 30, Klp up/down(n)= 8, Question Items= 30. The researcher categorizes each question item based on the differentiated power index (DP) that has been calculated.

Table. 5.  
Distinguishing Power Index Category

| Category     | D Value Range           | Percentage of D Grades |
|--------------|-------------------------|------------------------|
| Very Good    | $D > 0.40$              | > 40%                  |
| Good         | $0.30 \leq D \leq 0.40$ | 30% - 40%              |
| Fair         | $0.20 \leq D < 0.30$    | 20% - 30%              |
| Medium       | $0.10 \leq D < 0.20$    | 10% - 20%              |
| Low          | $0 \leq D < 0.10$       | 0% - 10%               |
| Negative/Bad | $D < 0$                 | Negative               |

Source : Research results2024

#### Analysis

- Item with Very Good Discriminating Power: Item 20.
- Items with Good Discriminating Power: Items 8, 10, 15, 17.
- Items with Sufficient Discriminating Power: Items 12, 21, 25, 27, 29.
- Items with Medium Discriminating Power: Items 1, 3, 16, 22, 24, 28.
- Items with Low Discriminating Power: Items 4, 6, 7, 9, 11, 13, 19, 23, 26.
- Items with Very Low Discriminating Power: Items 2, 5, 18, 30.
- Items with Negative/Poor Discriminating Power: Item 14.

From this analysis, we can see that some items have low or even negative differentiating power. The items may need to be corrected or replaced to improve the overall quality of the test.

#### 5. Distractor Power

The strength of the distractor is known from the student's response to each distractor item. Distractors are acceptable if they can confuse less knowledgeable students and choose them. Each test option is analyzed to check for effective and ineffective items. The findings show that each test has ineffective distractors in it. The following table will summarize the distracting power of each test.

Number of Subjects= 30

Question Items= 30

Analysis of the quality of distractors is very important in assessing the effectiveness of an item in a test. A good distractor is a distractor that is able to attract a number of respondents who answer incorrectly, so that not only a few students choose the distractor incorrectly. Rather, distractors

Analysis of the quality of distractors is very important in assessing the effectiveness of an item in a test. A good distractor is a distractor that is able to

attract a number of respondents who answer incorrectly, so that not only a few students choose the distractor incorrectly. On the other hand, distractors that are not chosen by anyone or are chosen by only a few students indicate that the distractor is ineffective.

#### Overall Analysis

From the result, we can identify several trends and insights regarding the quality of distractors in the test items:

1. Very Good and Good Distractors : Items 6 (D), 10 (C, D), 15 (A, D), 17 (C), 20 (A, B, C), 22 (A, C), 24 (A, B), 26 ( B, D), 27 (A) has an effective distractor because it attracts many students who answer incorrectly.
2. Bad and Very Bad Distractors : Many items have ineffective distractors, such as items 1 (B, D), 3 (B, D), 4 (A, B, D), 5 (A, B, D), 7 (A, C), 13 ( B, D), 19 (B, C, D), 28 (A, C), 29 (A, B, C), 30 (B, C, D), which indicates the distractor is unattractive or unable to mislead students who answered incorrectly.
3. Negative Distractors : Item 14 has a negative distractor in option B, which could indicate a problem in the question formulation or distribution of answers.

#### Recommendations

1. Revise Ineffective Distractors : For items with very poor or poor distractors, it is recommended to revise the answer choices to make them more misleading and effective.
2. Maintain Effective Distractors : Answer choices that successfully attract students who answered incorrectly should be maintained and used as examples in preparing other questions.
3. Review Item Content : For items with negative distractors or not selected, it is necessary to review the question content and answer options to ensure relevance and quality.

By carrying out this analysis and revision, the quality of test questions can be improved, thus providing a more accurate evaluation of student abilities. Based on the results of the analysis of the questions, it can be concluded that most of the questions are classified as easy or very easy. Of the 30 questions, 19 of them are in the very easy category with a percentage of correct answers above 85%, and the other 8 items are classified as easy with a percentage of correct answers between 70% and 85%. There are only 4 questions that fall into the medium category, and 2 questions that are very difficult with a percentage of correct answers below 50%. This suggests that the questions may not be challenging enough for the students.

Discriminating power analysis shows that many questions have low or very low discriminating power. Of the 30 items, only one item has very good discriminating power (above 40%), 4 items have good discriminating power (between 30% and 40%), 5 items have sufficient discriminating power (between

20% and 30%), and 6 questions with moderate discriminating power (between 10% and 20%). On the other hand, there were 9 questions with low differentiating power (between 0% and 10%) and 5 questions with very low or negative differentiating power. This shows that many questions are less effective in differentiating between students with high and low ability. In terms of distractor quality, many distractors are ineffective in misleading students who answer incorrectly. However, there are some distractors that are effective because they are able to attract many students who answered incorrectly, indicating that the distractor is good.

Overall, these items require improvement, especially in terms of discriminatory power and distractor quality. Although most questions are easy to answer correctly, their effectiveness in measuring students' overall abilities needs to be improved. Therefore, before being used as exam questions, further revision and improvement is highly recommended, especially on questions with low discriminating power and ineffective distractors. In addition, it is necessary to add more challenging questions to ensure a variety of levels of difficulty and to assess students' abilities more accurately.

## DISCUSSION

Théophile Muhayimana, Lambert Kwizera, and Marie Rose Nyirahabimana in their journal entitled " *Using Bloom's taxonomy to evaluate the cognitive levels of Primary Leaving English Exam questions in Rwandan schools*" found that the six class exam questions in 574 exam questions stated that they were less effective in measuring students' abilities due to the ease level of the questions which reached 98.79%. (Muhayimana et al., 2022) This is in line with researchers' findings which stated that the quality of the questions from Muhammadiyah Palopo Middle School was less effective in measuring students' abilities. The difference between the researcher's research and Theophile's research is the use of applications to test the questions. Researchers use anates while Theophile's research uses Bloom's Taxonomy.

Zakiyah and Jamilah in their journal entitled " *Analysis of Items, Choices, and Reliability of English Achievement Tests during Covid-19*" stated that the results of students' English language test questions were declared less effective because the questions were young at 46% and the questions at distraction were 18% so there were questions that needed to be revised. This is in line with research by researchers who state that the quality of the questions from Muhammadiyah Palopo Middle School is less effective in measuring students' abilities because of the 30 questions, 19 of them are in the very easy category with a percentage of correct answers above 85%, and the other 8 items are classified as easy with a percentage of correct answers between 70% and 85%. There are only 4 questions that fall into the medium category, and 2 questions that are very difficult with a percentage of correct answers below 50%. This suggests that the questions may not be challenging enough for the students. The difference between the researcher's

research and Zakiah's is the sample level where the researcher used 30 samples while Zakiah used 287 and the researcher used tests while Zakiah used the Quest program.

Herlia Alfiana, Hari Karyono, and Wawan Gunawan in their journal entitled "*Analysis of Test Items for Critical Thinking Skills and Procedural Knowledge of English Language Grammar*" stated that the results of the differentiating power test grouped the test items into good and very good categories. The validity test proves that all test items are valid. Reliability test states that the test was designed is reliable. Thus, it can be seen that the test items designed do not need improvement or can be used to measure students' critical thinking abilities and procedural knowledge in English grammar. This is because the level of difficult category questions and distracting questions is balanced with light and medium questions. After reading the journal, I found that the questions used were 20 questions, but all of them were carefully looked at and looked at carefully for the number of questions. This could be a good suggestion for teachers to pay more attention to making test items for grade promotion.

## CONCLUSION

The test analysis shows that no tests are perfectly good. It is found that almost all tests need to be revised. Test quality as whole shows 27% of all test items are valid. Furthermore, it is found that the test is highly reliable. Based on the findings, the majority of questions were considered easy or very easy, with a few exceptions considered moderate or very difficult. This suggests that the majority of students will probably be able to answer these questions well, but there are some areas that may need further attention to increase the level of difficulty and diversity of questions. From the result of discrimination index analysis, we can see that some items have low or even negative differentiating power. The items may need to be corrected or replaced to improve the overall quality of the test. In terms of discriminatory power and distractor quality. Although most questions are easy to answer correctly, their effectiveness in measuring students' overall abilities needs to be improved. Therefore, before being used as exam questions, further revision and improvement is highly recommended, especially on questions with low discriminating power and ineffective distractors. In addition, it is necessary to add more challenging questions to ensure a variety of levels of difficulty and to assess students' abilities more accurately.

## REFERENCES

- Agnes Meilina. (2020). *7 things you need to pay attention to when evaluating learning* . Work. <https://blog.kejarcita.id/7-hal-yang-perlu-diperhatikan-saat-tidak-produk-pembelajaran/>

- Esmaeeli, B., Shandiz, E.E., Norooziasl, S., & Shojaei, H. (2021). The optimal number of choices in multiple-choice tests : A systematic review. *Med Edu Bull* , 2 (5), 253–260. <https://doi.org/10.22034/MEB.2021.311998.1031>
- Fauzie, M., Pada, AUT, & Supriatno, S. (2021). Analysis of the difficulty index of item bank according to cognitive aspects during the Covid-19 pandemic. *Journal of Educational Research and Evaluation* , 25 (2), 162–169. <https://doi.org/10.21831/pep.v25i2.42603>
- Hoogmoed, A.H. Van, & Ham, A. Van Den. (2022). Exploring the reliability, validity, and dimensionality of the 'Kieler kindergarten test for mathematics'. *Pedagogische Studien* , 99 (304–324).
- Husnani Aliah. 2020. The Analysis of Junior High School Teacher-Made Test for The Students in Enrekang. *Jurnal IDEAS*.
- Indriyani, R., Arnina, Nasruddin, I., & Noviani, D. (2023). The Nature of Curriculum in the World of Education. *As-Shuffah (Journal of Islamic Studies)* , 11 (1), 1–10.
- Juniardi, W. (2023). *Definition of Diagnostic Tests, Objectives, Functions, Benefits, Types and Examples* . Quipper. <https://www.quipper.com/id/blog/info-guru/tes-diagnostic/>
- Kazemitabar, M., Moghadamzadeh, A., Habibi, M., Hakimzadeh, R., & Garcia, D. (2020). School health assessment tools: A systematic review of measurement in primary schools. *Maryam Kazemitabar1 , Ali Moghadamzadeh1, Mojtaba Habibi2,3 , Rezvan Hakimzadeh1 and Danilo Garcia4 , 8* . <https://doi.org/10.7717/peerj.9459>
- Meni Handayani, SS, MS, Dra. Yufriidawati, MS, & Novrian Satria Perdana, SE, ME (2020). *THE IDEAL TEST IN THE FUTURE* (MS Meni Handayani, SS, MS Dra. Yufriidawati, & ME Novrian Satria Perdana, SE (eds.)). pskp. kemdikbud. [https://pskp.kemdikbud.go.id/assets\\_front/images/produk/1-gtk/buku/1629813314\\_Puslitjak\\_10\\_Ujian\\_yang\\_Ideal\\_di\\_MasaMentanda.pdf](https://pskp.kemdikbud.go.id/assets_front/images/produk/1-gtk/buku/1629813314_Puslitjak_10_Ujian_yang_Ideal_di_MasaMentanda.pdf)
- Mochammad Noor Akhmadi. (2021). Analysis of Evaluation Question Items for Theme 1 Class 4 Sdn Plumbungan Using the Anates Program. *Ed-Humanistics: Journal of Educational Sciences* , 6 (1), 799–806. <https://doi.org/10.33752/ed-humanistics.v6i1.1464>
- Muhayimana, T., Kwizera, L., & Nyirahabimana, M.R. (2022). Using Bloom's taxonomy to evaluate the cognitive levels of Primary Leaving English Exam questions in Rwandan schools. *Curriculum Perspectives* , 42 (1), 51–63. <https://doi.org/10.1007/s41297-021-00156-2>
- Purwati, LM, Arianty, R., Syakilah, DM, Ridlo, S., & Susilaningsih, E. (2021). Analysis of Multiple Choice Test Questions Based on Higher Order Thinking Skills using the Anates Windows Application Version 4.0.9 For Windows. *UNIGA Journal of Education* , 15 (2), 460. <https://doi.org/10.52434/jp.v15i2.1287>
- Reffiane, F., Sudarmin, Wiyanto, & Saptono, S. (2021). *Developing an Instrument to Assess Students' Problem-Solving Ability on Hybrid Learning Model Using*

- Ethno-STEM Approach through Quest Program* . 11 (4), 1–8.  
<https://doi.org/10.47750/pegegog.11.04.01>
- Test, K. (2020). *Middle School Teacher-Made Analysis of Student Examinations in Enrekang* . 2 , 122–138.
- Ukobizaba, F., Nizeyimana, G., & Mukuka, A. (2021). Assessment Strategies for Enhancing Students' Mathematical Problem-solving Skills: A Review of Literature. *Eurasian Journal of Mathematics, Science and Technology Education* , 17 (3), 1–10. <https://doi.org/10.29333/ejmste/9728>

**Copyright © 2025,** Asman, Husnani Aliah, Nasriandi

The manuscript open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.