

AN ANALYSIS OF ENGLISH SUMMATIVE TEST ITEMS OF TWELFTH GRADE STUDENT OF MAN 2 MERANGIN ACADEMIC YEAR 2019/2020

Risnita*¹, Netti Zurnelli², Wahyuni Fitria³, Makrub⁴

¹risnita@uinjambi.ac.id, ²nettizurnelli@uinjambi.ac.id, ³wahyunifitria@uinjambi.ac.id,
⁴makrubaja17@gmail.com

UIN Sulthan Thaha Saifuddin Jambi

ABSTRACT

The purpose of this research is to evaluate the validity, reliability, level of difficulty, discriminating power, and distractor efficacy of the English summative test items for the 12th grade MAN 2 Merangin students for the 2019-2020 academic year. This is a quantitative descriptive study. The subjects of this research were the questions from the 2019/2020 MAN 2 Merangin 12th grade English summative examination. Questions, answer keys, and student responses were included in the documentation used for the data collection technique. The quantitative technique is used to analyze the validity, reliability, level of difficulty, discriminating power, and distractor efficacy of the dofa. The obtained dofa were analyzed using manual Microsoft Excel calculations. The results indicate that there are 38 valid items (76%) and 12 invalid items (14%); the questions on the English summative test of 12th grade students of MAN 2 Merangin for the 2019-2020 academic year are unreliable with a reliability level of 0.422; there are 12 items with an easy level of difficulty, 34 with a medium level of difficulty, and 4 with a difficult level of difficulty; and there are 22 poor items.

Keywords: Analysis Test Item, English Summative Test, MAN 2 Merangin

A. INTRODUCTION

Evaluation plays a crucial function in education and related activities. Arikunto (2013: 3) defines evaluation as gathering data to assess the achievement of educational goals. Evaluation serves as a baseline for decision-making about the assessed thing. Evaluation provides valuable insights into the success of an effort, enabling future improvements. Evaluation is crucial to teaching and learning. In formal education, teachers are crucial in determining the effectiveness of the learning process. A competent teacher not only teaches

but also evaluates effectively. During the teaching process, teachers assess student progress in mastering lessons over time. The evaluation results reveal teacher quality and student ability. Tests play a crucial part in evaluating the teaching and learning process. According to Shohamy (1985:3), a test should accurately represent knowledge as a sample. Testing should only assess a sample of behavior or knowledge, not the entire lesson delivered by teachers and pupils, as it is impossible to assess all talents. A exam should be well-organized for effective utilization. To be considered a good test, it must meet the following characteristics: validity, reliability, practicality, difficulty level, discrimination power, and effective distracters for multiple choice items. Many schools worldwide utilize tests to assess students' abilities and knowledge.

Test objects should be evaluated to determine test quality. Prepare the test object well, as its quality will impact the test results. Test quality depends on item quality. The teacher must prioritize the quality of exam items by conducting item analysis. Assessing item quality helps teachers determine which items meet requirements, which should be removed, and which should be revised. To assess student achievement, teachers typically provide tests. Teachers can administer accomplishment tests after each chapter or at the end of the semester. Achievement tests measure student learning through a structured process. Two types of accomplishment tests exist: formative and summative.

Summative tests generate grades or figures for evaluating student achievement. This test is administered after completing the learning experience unit or all topic matter. Summative tests establish award categorization at the end of a course or program. A formative test monitors student learning and provides feedback for teachers and students to enhance their learning. Formative tests let students discover strengths and weaknesses, focus areas for improvement, and enable professors handle issues promptly. Summative test items were employed in this study. The researcher evaluates the validity, reliability, and other quality aspects of the item test. Future school years will use the analysis results to guide teachers in teaching unfamiliar content to students.

The English teacher at MAN 2 Merangin did not assess the final test item for validity, reliability, etc. Based on observations and interviews, it is unclear if the test items are excellent or not. The researcher found samples to support and strengthen this analysis, including: The first research by Muspira Humaerah (2016) analyzed the English Summative Test for Second Grade Students at MAN 1 Tanete Bulukumba. The researcher used a quantitative descriptive method to analyze data from an English summative test.

Anis Yunita Sari (2017) conducted a study on the item analysis of English mid-term test items for 7th graders at SMP Negeri 2 Wonosari in the 2015/2016 academic year. The data collection method is a study document. Data analysis can be done using qualitative methods such as multiple choice items or quantitative methods such as Arikunto formulas. The researcher selects MAN 2 Merangin as the research location and aims to analyze English summative test items for twelfth grade students.

B. METHOD

In this sub chapter, the location and duration of the investigation are detailed. Following is an explanation of the location and time of the research. MAN 2 Merangin is located in Tabir, Rantau Panjang, Merangin, where the investigation will be conducted. The school has nine classrooms. Each classroom contains thirty-five students. This research will be conducted during the second semester of the 2019/2020 academic year.

C. FINDINGS AND DISCUSSION

1. Finding

The validity, reliability, difficulty, and differentiating power of English summative test items for twelfth grade students in MAN 2 Merangin academic year 2019/2020 can be assessed. The following indicators are discussed:

- **Validity**

This research calculated item validity using the point-biserial correlation formula. Calculations were compared to at a 5% significance level. The test was given to 33 students, with a standard of 0.34. If the result was higher than, the item test was legitimate. If the result was lower, the item test was invalid. The research resulted in 38 valid items (76%), and 12 invalid items (24%). Fix invalid items and reuse valid ones in the question bank. 36% 46% 6% 12% Distractor Efficiency rating: very nice, good, fair, bad, very terrible.

The English summative test items for twelfth grade students at MAN 2 Merangin in 2019/2020 show good validity, with over 50% of the total questions being legitimate. Anas Sudijono (2012: 163) defined item validity as the precision of measuring items for their intended use. Following the investigation of Question Item validity, the following steps might be taken: A valid item can be added to the question bank for future use in the next semester's test. Discard invalid items and replace with questions based on material indicators.

- **Reliability**

Reliability refers to the consistency and stability of phenomenon measurements (Carmines and Zeller, 1979). Repeatability is part of reliability. A test is dependable if it yields consistent results when repeated under constant settings (Moser and Kalton, 1989).

Remeasuring reliability multiple times yields consistent results that do not vary. The KR-20 formula is used to calculate the reliability of English summative test items for twelfth grade students in MAN 2 Merangin for the 2019/2020 academic year. Calculations are done manually in Excel. If the reliability coefficient (r_{11}) is greater than 0,70, the item being evaluated is highly dependable, while a lower value indicates low reliability or unreliability. The calculated findings yielded $r_{11} = 0.422$.

The calculations show that the English summative test items for 12th graders in MAN 2 Merangin for the 2019/2020 academic year have low reliability ($r_{11} < 0.70$). A test is dependable if it consistently yields consistent results across multiple groups and people

(Zainal Arifin, 2012: 258). Nana Sudjana (2011: 16) states that the reliability of an evaluation instrument indicates its consistency in judging. The explanation suggests that the English summative test items for twelfth graders in MAN 2 Merangin for the 2019/2020 academic year are of low reliability.

- **Level of difficulty**

The difficulty level of an item is determined by the ratio of right answers to the total number of test takers. Items are considered decent if they fall into the middle category, meaning they are neither too difficult nor too easy. Easy problems do not motivate students to improve their problem-solving skills. However, excessively difficult questions can depress pupils and reduce their motivation to attempt again.

The data reveals that 24% of multiple choice questions are easy, 68% are medium, and 8% are challenging. A good question has a moderate difficulty level (0.31-0.70) (Suharsimi Arikunto, 2013: 225). Suharsimi Arikunto (2013) defines a good question as having a medium difficulty level of 0.31-0.70 (p. 225). From the level of difficulty, the English summative test items for twelfth grade students in MAN 2 Merangin Academic Year 2019/2020 are of good quality. A medium degree of difficulty is present in 34 questions, accounting for 50% of the total questions.

The study found that most questions were classified as medium hard. Questions in the medium category can be saved in a question bank for future evaluation purposes. Re-examine simple or tough items to determine their reason, revise, and test on the next examination. Medium problems must be maintained.

- **Power distinction**

The Distinguishing Power formula is used to calculate the English summative test items of twelfth grade students in MAN 2 Merangin for the 2019/2020 academic year. For distinguishing power, subtract the proportion of correct answers in the higher group from the

proportion in the lower group. The discriminating power of an item is classified as poor if the index is 0.00-0.19, satisfactory if 0.20-0.39, good if 0.40-0.69, and very good if 0.70-1.00. A negative index indicates no distinguishing power. Suharsimi Arikunto (2013): 232 The research reveals 22 poor multiple-choice questions (44%), 17 passable (34%), 0 good (0%), and 11 superb (22%).

According to Zainal Arifin (2012: 273), differentiating power measures how well an item may distinguish between students who have learned the topic and those who have not, depending on specific criteria. The English summative test items for twelfth graders in MAN 2 Merangin's 2019/2020 academic year have strong differentiating power, with over 50% of questions able to differentiate between upper and lower groups. Higher differentiating power indicates better items, whereas lower power indicates worse ones. If most brilliant students answer a question properly, it has strong distinguishing power. The study found that questions with excellent discriminating power were kept, whereas those with poor power needed to be revised.

- **Distracter effectiveness**

To determine distractor efficiency, count the number of testees who chose a, b, c, d, e, or did not choose. Distractors' efficiency indicates their proper functioning. The good distractor will be chosen by at least 5% of the total testees. The twelfth grade of MAN 2 merangin academic year 2019/2020 has 33 students, and a distractor is considered operational if at least 5% of them, or 1,65 (2 students), select it. The English summative test results for twelfth graders at MAN 2 Merangin indicate 18 items with very good distractor efficiency (36%), 23 items with good distractor (46%), 3 items with fair (6%), and 6 items with bad distractor (12%).

To evaluate the effectiveness of each item's distractor, utilize the following criteria taken from the Likert Scale: The distractor is highly effective when all four distractors are working. The effectiveness of a distractor is considered good when there are three working distractors. The efficiency of a distractor is considered high when there are two functioning

ones. The effectiveness of a distractor is inadequate when only one distractor operates. The effectiveness of a distractor is weak when all distractions fail. English summative test items for twelfth graders in MAN 2 Merangin academic year 2019/2020 have good divergent efficiency, with over 50% of test items scoring very good, good, or fair. Anas Sudijono (2012: 417) suggests the following follow-ups after examining the efficiency of the distractor: Store items with effective distractors in the question bank for future testing. Items with malfunctioning distractions can be rectified or replaced with alternative methods.

- **Analysis of Test Items: Validity, Reliability, Level of Difficulty, Distinguishing Power, and Distractor Efficiency**

After analyzing each criterion, the items were evaluated for validity, reliability, difficulty, distinguishing power, and distractor efficacy to assess the quality of English summative test items for 12th graders in MAN 2 Merangin for the 2019/2020 academic year. English summative test results for twelfth grade students at MAN 2 Merangin in 2019/2020 were evaluated for validity, difficulty, differentiating power, and distractor efficiency.

2. Discussion

The analysis of the English summative test items for twelfth-grade students at MAN 2 Merangin during the academic year 2019/2020 reveals a positive assessment of the test quality. The evaluation encompassing validity, reliability, level of difficulty, distinguishing power, and distractor efficiency yields the following conclusions:

- **Validity:** The validity analysis indicates that 76% of the English summative test items (38 items) can be considered valid, while 24% (12 items) fall under the invalid category.
- **Reliability:** The reliability index for the English summative test items is 0.422, signifying that the items can be classified as unreliable due to a reliability index lower than 0.70.

- Level of Difficulty: The distribution of items across difficulty levels reveals that 24% of items are categorized as easy, 68% as medium, and 8% as difficult. The majority of items fall within the medium category, indicating a balanced level of difficulty.
- Distinguishing Power: The distinguishing power analysis demonstrates that 22% of items exhibit excellent distinguishing power, 44% show poor performance, 34% are satisfactory, and none fall into the good category.
- Distractor Efficiency: The distractor efficiency assessment indicates that 36% of items have very good distractor performance, 46% perform well, 6% have fair performance, and 12% perform poorly as distractors.

In a comprehensive assessment of the English summative test items, considering criteria such as validity, reliability, difficulty level, distinguishing power, and distractor efficiency, it can be concluded that the majority of items (more than 50%) demonstrate good quality. Specifically, 32% of items meet all criteria for the good category, 26% fall into the medium category, and 42% have characteristics aligning with the bad category. This overall analysis affirms the effectiveness of the English summative test items in evaluating the twelfth-grade students' proficiency at MAN 2 Merangin for the academic year 2019/2020, suggesting that a substantial portion of the items meet or exceed the standards for good question design.

D. CONCLUSION

English summative test items for twelfth graders at MAN 2 Merangin in 2019/2020 have good quality, with over 50% of items meeting good criteria. Analyzing validity, reliability, difficulty, distinguishing power, and distractor efficiency yields the following conclusions:

- Validity results show r_{table} 0,344 with 5% significance. In the English summative test, 76% of the items were valid, while 24% were invalid.
- The English summative test items for twelfth graders at MAN 2 Merangin in 2019/2020 have a reliability index of 0.422. The English summative test items are not reliable due to a reliability index of r_{11} below 0.70.
- Based on difficulty level, easy category has 12 things (24%), medium category has 34 items (68%), and difficult category has 4 items (8%). English summative test items have a good difficulty level, with over 50% in the medium group.

- The English summative test had 22 items in the bad category (44%), 17 in the satisfactory category (34%), 0 in the good category (0%), and 11 in the excellent category (22%).
- Using distractor efficiency, The results indicate that 18 items had very good distractor work (36%), 23 items had high functioning (46%), 3 things had medium functioning (6%), and 6 items had awful distractor work (12%).

Test analysis reveals that 16 items in the good category meet all criteria (32%), 13 items in the medium category meet 3 criteria (26%), and 21 items in the bad category meet ≤ 2 criteria (42%). Overall study of English summative test items for 12th graders at MAN 2 Merangin in 2019/2020 indicates good questions with over 50% meeting criteria.

REFERENCES

- Alfabetta Creswell, J. W. (2012). Educational Research (4th Ed).
- Anas Sudijono.(2012). Pengantar Evaluasi Pendidikan. Jakarta: PT Raja Grafindo Persada.
- Arikunto, Suharsimi. (2010). Pendidikan Dasar – Dasar Evaluasi. Jakarta : Aksara Sugiyono. (2011). Metode Penelitian Kantitatif Kualitatif dan R&D. Bandung.
- Dian Nurliyanto, (2015). Test Item Analysis of The Final Examination on Economics Subject in Grade Xii Ips Sma Negeri Banyumas Academic Year 2014/2015
<https://eprints.uny.ac.id/25211/1/15.%20FULL%20SKRIPSI.pdf>.
- Harmer, J. (2001). The Practice of English Language Teaching (3rd Ed). Pearson Education Limited.
- Nana Sudjana.(2014). Penelitian Hasil Proses Belajar Mengajar. Bandung: PT Remaja.
- Pearson Education. McMillan, J., & Schumacher, S. (2014). Research in Education: Evidence-Based Inquiry (7th Ed). Pearson Education.
- Ratna, C, W., (2016). Analisis Butir Soal Ujian Akhir Semester Gasal Mata Pelajaran Pengantar Akuntansi Dan Keuangan Kelas X Akuntansi Smk Swadaya Temanggung Tahun Ajaran 2015/2016
https://eprints.uny.ac.id/38661/1/skripsi_12803241006_RATNAfull.pdf.
- Rosdakarya Rusmiana, F.D, (2015). The Test Item Analysis of 1st Semester Final Test of the Accounting Theory for Vocational Education: Case Study of Smk Ypkk 1 Sleman

Academic Year of 2014/2015

[https://eprints.uny.ac.id/15530/1/Undergraduate%20Thesis_Rev.setelah %20ujian-full.pdf](https://eprints.uny.ac.id/15530/1/Undergraduate%20Thesis_Rev.setelah%20ujian-full.pdf)

- Suwartono, (2014). Dasar-dasar metodologi pendidikan. Yogyakarta: ANDI Yogyakarta
- Siyoto, s., Sodik, A,. (2015). Dasar Metodologi Penelitian. Yogyakarta: Literasi media publishing.