

CHAPTER I

INTRODUCTION

This chapter discussed of background of the study, statement of the research problems, objectives of the study, significance of the study, scope and limitation of the study, and definition of the key terms.

A. Background of the Research

Education has an important role as a basis for preparing someone in today's era of competition, especially in English education. As an international language, English skills are one of the abilities that should be mastered. Therefore, quality tests play an important role as a evaluating tool for the student's ability and performance in English. Learning evaluation is an ongoing process that includes assessing performance needs, designing curriculum and learning experiences, facilitating learning, and supporting the transfer of learning outcomes.¹ The goal is to obtain information to improve educational performance. While test is a method to assess how capable, how insightful, how skillful, or how a person performs in a particular field. One of the tests usually applied in schools is the summative test.²

¹ Dwi Astuti Wahyu Nurhayati as cited in Siti Shofiah et al., *Dasar-Dasar Evaluasi Pembelajaran*, ed. Sarwandi (PT. MIFANDI MANDIRI DIGITAL, 2023). p. 25.

² H. Douglas Brown, *Language Testing Book: Principles and Classroom Practice*, Book (Pearson Education, Inc, 2004). p. 3.

A summative test is designed to measure students' mastery of the material taught at the end of a certain period. A test is considered good if the instrument fulfills the standards of validity, reliability, comprehensiveness, and practicality.³ Validity helps ensure that the test measures what it is supposed to measure, while reliability ensures the consistency of the test measurement results. In this study, the researcher focuses on analysis each question. So, the researcher evaluating the test questions' quality by looking at how students answered them. This means that the student's answers reflected the quality of each test question. Analysis each question of the test is the most important factors to ensure the quality of the test.

However, in practice, test made by teachers often do not qualify according to the basic competencies that must be achieved. As the data obtained by Vianti et al., from the final test of semester 1 of class XI students of SMAN 3 Rembang using Arikunto's theory, found that out of 43 question items, only 28 question items were valid and 15 other question items were invalid.⁴ Another study in South Tangerang showed that there were still some question items that had to be revised and some others that needed to

³ Rizky Hilaldy, "Content Validity of English Summative Test" *Journal of Language Learning and Research (JOLLAR)* 4, no. 1 (2021): 37–41, <https://doi.org/10.22236/jollar.v4i1.7708>. p. 38.

⁴ Febriana Vicky Vianti, et al., "An Analysis on the Validity Test Question With the Exam Content Outline in First Semester for the XI Grade Student of SMAN 3 Rembang," *Undergraduate Conference on Applied Linguistics, Linguistics, and Literature* 2, no. 1 (2022): p. 101.

be replaced.⁵ The findings of some of these researchers show the importance of summative tests reflecting the competencies that should be measured.

Most previous research has focused more on the analysis of multiple-choice questions and rarely investigated essay-type questions, even though essay questions play an important role in measuring students' critical thinking skills more deeply. Although research on the analysis of tests has been extensively conducted in various contexts, in the Tulungagung region, studies that comprehensively review summative test were still very limited. Based on data found by several previous researchers, the researcher considers this to be an issue in the field of education, and it was important to analyze the quality of English summative test created by teachers. The purpose of this research is to fill that gap by analyzing the quality of the mid-semester English test created by teachers at MA Darul Falah. This study was expected to provide reflective material that can assist teachers in both public and private schools in developing more structured, fair, and unbiased learning evaluations for students.

MA Darul Falah was chosen as the research location because, based on the results of an interview with Rofik Zamzami, S.Pd.I, that in this school, there has never been a validity test for test questions. This problem is because the validity test team has not yet been formed. So, the nature of the tests they provide is still limited to the teacher only. No standard might be

⁵Sabilar Rosyad, "An Item Analysis of the English Summative Test for the Seventh Grade Students of SMPN 12 Tangerang Selatan" *AKSELERASI: Jurnal Ilmiah Nasional* 6, no. 1 (2024): p. 18, <https://doi.org/10.54783/jin.v6i1.876>.

used as a reference for whether or not the question is feasible. In terms of evaluating the quality of the questions, the school looks at the final score. So, from the vice-principal curriculum team, there will be some kind of standardization. The minimum limit for the accumulation of grades for all subjects is 5.5. So, as a standard, for example, students accumulate scores below 5.5. Then this means there will be remediation that we give to students because they are considered not yet fulfilling the target of the subject.⁶

In addition, interviews conducted with English teacher at MA Darul Falah in the end of 2024, stated that teachers prepared their exams using a grid referencing the Merdeka Curriculum. According to the teacher, the goal of creating essay questions is to limit the culture of cheating while also influencing students to think more critically. This was also based on observations of the final test used in the 2023/2024 academic year, which included a mix of 10 multiple-choice questions, 10 short descriptions, and 5 long descriptions.⁷ This pattern of questions was rare at other schools, where usually multiple-choice questions were dominant than essay questions, hence it was a compelling argument for the researcher to select MA Darul Falah as the research location. With these criteria, the researcher concluded that this school had unique characteristics that were useful to

⁶ Interview Rofik Zamzami, English Teacher at MA Darul Falah, conducted in person, January 30th, 2025.

⁷ Interview Rofik Zamzami, English Teacher at MA Darul Falah, conducted in person, December 10th, 2025.

research in terms of validity, reliability, discrimination power, and difficulty level of the test.

This research focused on the written test for English subjects, especially the mid-term tests. The mid-term test was selected because it is a summative assessment that evaluates students' learning progress over a half-semester and typically uses teacher-prepared questions. This makes a validity and reliability analysis relevant. Furthermore, the mid-term test was chosen since the essential materials, such as the question sheet and student response sheets, were readily available and in good condition, because the exam was held in March 12th 2025. The availability of complete and contextual data contributes to the accuracy and efficiency of the analysis process. Fiktorius suggests that people's assumptions about tests in Indonesia lack authenticity. Therefore, the quality of teacher-made tests is still debated in the educational assessment process.⁸

This study applied document analysis methods. The analysis was conducted by analyzing written test and the students' answers. As such, this study not only provides an overview of test quality but also serves to identify areas that need to be improved to enhance the quality of test.

Test analysis is crucial, especially in schools that have never conducted a validity test with limited learning time. For 10th graders students at MA Darul Falah, there is only about one hour of English learning per week. This

⁸ Naelul Rohmah, "Validity and Reliability Study on Teacher-Made Assessment for English Mid-Term Examination" *Conaplin* (2020): p. 107, no. 254, <https://doi.org/10.2991/conaplin-18.2019.236>.

is a challenge in itself to improve students' abilities. For MA Darul Falah, this analysis provides information about the quality of the questions to improve the effectiveness of evaluation in Islamic boarding schools.

B. Statement of Research Problem

The main problem to be answered through this research is how the quality of the midterm summative test of the English subject made by teachers at MA Darul Falah for 10th graders based on item analysis. The problems presented in this study are:

1. How is the validity of the midterm test made by the teacher in the 2024/2025 academic year for the English subject for 10th graders at MA Darul Falah?
2. How is the reliability of the midterm test made by the teacher in the 2024/2025 academic year for the English subject for 10th graders at MA Darul Falah?
3. How is the difficulty level of the midterm test made by the teacher in the 2024/2025 academic year of the English subject for 10th graders at MA Darul Falah?
4. How is the discrimination level of the midterm test made by the teacher in the 2024/2025 academic year of the English subject for 10th graders at MA Darul Falah?

C. Objective of the Research

Based on the formulation of the problem, the purposes of this study are:

1. To determine the validity of the midterm test made by the teacher in the 2024/2025 academic year of the English subject for 10th graders at MA Darul Falah;
2. To determine the reliability of the midterm test made by the teacher in the 2024/2025 academic year of the English subject for 10th graders at MA Darul Falah;
3. To determine the item difficulty of the midterm test made by the teacher in the 2024/2025 academic year for the English subject for 10th graders at MA Darul Falah;
4. To determine the item discrimination of the midterm test made by the teacher in the 2024/2025 academic year for the English subject for 10th graders at MA Darul Falah.

D. Significance of the Research

The researcher hopes that the results of this study will be useful both theoretically and practically for readers. Theoretically, this research is useful to help improve the quality of English tests. In addition, the research can also be used as a reference for further research. Practically, this research can help teachers and schools to identify areas that need to be fixed so as to improve the quality of tests.

E. Scope and Limitation of the Research

The researcher analyzed the mid-term English summative assessment for 10th graders created by the English teacher based on the Merdeka Curriculum. The summative assessment sheet and the students' answer sheets were used as tools to measure students' abilities in the middle semester period in the 2024/2025 academic year for 10th graders in senior high school. The researcher focused on the validity, reliability, difficulty level, and discrimination index of the midterm summative test item.

F. Definition of the Key Term

To avoid misunderstanding and misinterpretation, the researcher provides the following definitions of key terms that can be used as guidelines in conducting this research.

1. Summative Test

A summative test is the process of summarizing or examining what has been learned and done at the end of the teaching and learning process.⁹ Flucher & Davidson stated that a summative test is usually given at the end of a learning unit to assess the success of a group or individual as a whole. A summative test, unlike a formative test, provides feedback to improve ongoing learning.¹⁰ While a summative

⁹ Ella Aprilia Tiana, "An Analysis of Content Validity on English Summative Assessment of Fifth Grade in The First Semester in Merdeka Curriculum" *Journal of English Teaching and Linguistic Issues (JETLI)* 3, no. 1 (2024): p. 2, <https://doi.org/10.58194/jetli.v3i1.1839>.

¹⁰ Glenn Flucher & Fred Davidson, *Language Testing and Assessment*, Routledge Taylor & Francis Group, 2007, https://doi.org/10.1057/978-1-137-59900-1_39.

test can help highlight areas where students may need more guidance, their main purpose is to measure student learning at a specific point in time.

2. Validity

A valid test requires that the measurement instrument meets the test requirements, specifically validity, reliability, and practicality. Validity refers to the extent to which conclusions drawn from test results are appropriate, meaningful, and useful concerning the purpose of the test.¹¹

4. Reliability

The most accurate measure of test accuracy is reliability, which is the degree to which test results are stable, consistent, and error-variance-free. This is the degree to which a test delivers the same grading to test takers when administered again and measured using the Alpha Coefficient or KR-20.¹²

5. Item Difficulty

According to Thorndike et al., Item difficulty, or question difficulty level, is the percentage of test participants who answer a question correctly. This value indicates how easy or difficult a question is for students.¹³ The level of difficulty is also related to the ability of the

¹¹ H. Douglas Brown, *Language Testing Book: Principles and Classroom Practice*, 2004, p. 314.

¹² Richard J. McCowan & Sheila C. McCowan, *Item Analysis for Criterion-Referenced Tests*, Research Foundation of SUNY/Center for Development of Human Services, 1999, p.45. <http://eric.ed.gov/ERICWebPortal/recordDetail?accno=ED501716>.

¹³ H. Douglas Brown, *Language Testing Book: Principles and Classroom Practice*, 2004, p.18.

question to differentiate between students who understand the material and those who do not. Questions that are too easy or too difficult are usually ineffective in distinguishing students' abilities.¹⁴

6. Item Discrimination

Item discrimination contrasts the number of accurate answers from high and low scores on an item. It refers to the degree to which question distinguish between students in the upper and lower categories. The entire test and each question should measure the same thing. Upper students are expected to answer excellent question accurately, whereas lower students are more likely to answer wrong.¹⁵

¹⁴ Mary Ann Tobin, *Guide to Item Analysis*, 2024, p.3.

¹⁵ Richard J. McCowan & Sheila C. McCowan, *Item Analysis for Criterion-Referenced Tests*, 1999, p. 20.