

CHAPTER I

INTRODUCTION

A. Background of the Study

Education is fundamental in nurturing students' development and helping the students to achieve full potential. As students are placed in a classroom environment that supports learning, this will promote their cognitive and social growth. According to Woolfolk (2013) a supportive classroom where students feel safe, respected and supported in the learning process can significantly increase motivation and engagement in the learning process. In this learning context, assessment becomes a very important tool to understand whether educational goals have been achieved. According to David et al. (2013) assessment is a systematic process for collecting, synthesizing and interpreting information to support decision-making in educational settings. In conclusion, the success of education is not only determined by the content of the material and teaching methods, but also by a supportive classroom environment and an appropriate assessment system.

The most common form of assessment used in schools is the multiple-choice test. According to Brown, (2004) written assessment, especially multiple choice tests, is one of the most widely used evaluation instruments in education because of its convenience. In line with that, Arikunto (2018) stated that multiple choice is one of the objective tests that are often used in summative evaluation. Supporting by Arikunto's findings reiterate that this

summative test also has limitations, such as the possibility of students guessing answers and not showing students' real thinking processes, so that the results obtained do not necessarily reflect the real understanding of students. Based on these findings, it is important for teachers not only to rely on the practicality of multiple choice tests, but also to pay attention to the quality of each item prepared with awareness in understanding the skills in analyzing and compiling items.

Ministerial Regulation No. 23/2016 provides a framework for educational assessment in Indonesia's Curriculum 2013 (K-13), outlining that teachers should evaluate students' attitudes, knowledge, and skills (Indonesian Ministry of Education and Culture, Article 3, Paragraph 1). Student knowledge is assessed through various methods, including written tests, oral tests, and assignments aligned with specific competencies. Among written assessments, the Multiple- Choice Test (MCT) is one of widely used formats in English classrooms. According Zimmaro (2016) MCTs are effective in measuring knowledge-based outcomes and learning achievements. Similarly, Kolte (2015) highlights MCTs as objective and reliable tools for evaluating student understanding. Teachers often choose MCTs for their efficiency, simplicity, and cost-effectiveness in scoring (Bailey, 1998), as well as their capacity to measure learning across multiple cognitive levels (Zimmaro, 2016). In addition, MCTs are particularly suitable for large-scale testing, such as Indonesia's National Examination and international tests like TOEFL and IELTS. Furthermore, Kolte (2015) notes that the presence of a single correct

answer helps reduce scoring subjectivity. Given their central role in assessing student knowledge, it is essential that MCTs be constructed according to established quality standards.

Several studies have been conducted to evaluate the quality of multiple-choice questions in English assessments at various educational levels by focusing on item difficulty, discrimination index, and distractor effectiveness. The outcomes of these studies have shown considerable variation. For instance, research conducted by Ningsih & Widowati, (2021), which examined 40 multiple-choice items from the final English examination for tenth-grade students at SMK PGRI Wonoasri, revealed that a significant portion of the questions 47% were of moderate difficulty. However, the same percentage of items required revision due to low discrimination indices. Additionally, their analysis showed that the distractors functioned effectively in 90% of the items, successfully diverting students from the correct answers. Similarly, Rejeki et al., (2023) carried out a study analyzing English for Academic Purposes (EAP) exam questions administered to first-year civil engineering students at Universitas Muhammadiyah Surakarta. The results were less encouraging, indicating that most of the questions were too easy, lacked sufficient discriminatory power, and had a high number of ineffective distractors. These findings suggest a recurring issue in test item construction ranging from poorly balanced difficulty levels and weak discriminatory ability to non-functional distractors resulting in tests that fall short of accurately evaluating student competence. Hence, these results emphasize the need for

regular and systematic item analysis as a core component of educational evaluation.

In contrast to studies reporting unsatisfactory outcomes, more positive findings were presented in research by Amalia & Nur, (2020). Their analysis of 40 multiple-choice English test items for ninth-grade students in Ponorogo indicated that most questions had a moderate level of difficulty, high discrimination power, and effective distractors. These results portray a well-constructed set of exam items. Reinforcing this, a study by Memy & Rosidah, (2023) assessed 50 English final exam questions from MTs Nurul Yaqin and found a similar pattern. Their findings showed the majority of items fell within the medium difficulty range, featured strong discrimination indices, and employed distractors that were generally functional. These results suggest that both assessments met key standards of item quality by incorporating appropriately challenging questions, reliable indicators of student performance, and distractors that served their intended purpose.

This present study aims to address a gap between earlier research and the current investigation. It refers to the study by Amalia and Nur (2020) as a relevant foundation. That research examined English multiple choice questions at the junior high school level. In contrast, the current study explores English exam questions for tenth grade students at SMAK St. Bonaventura Madiun. The focus on a higher level of education provides a new contribution to item analysis. This research evaluates test questions based on three core aspects: difficulty level, discrimination index, and distractor effectiveness. The goal is

to provide a deeper understanding of how teacher-made tests function in classroom assessment. Compared to previous research, this study offers insights into a context that has received less academic attention.

Based on the previous studies discussed, which have revealed strengths and weaknesses in the construction of English multiple-choice tests at different levels of education, this study aims to evaluate the quality of teacher-made multiple-choice questions used in the English final exam at SMAK ST Bonaventura Madiun. According to the criteria established in item analysis, this study will focus on three main indicators: difficulty level, differentiating index, and distractor function. In this study, each item will be categorized as easy, moderate, or difficult, assessed based on its ability to differentiate between high- and low-ability students, and examined for the effectiveness of its distractor function. Adopting a descriptive quantitative approach, this study aims to provide constructive insights into assessment practices and support the development of more valid and accurate tools.

B. Delimitation of the Study

This study is limited to analyzing three specific aspects of multiple-choice item quality: difficulty level, discrimination index, and distractor effectiveness. The scope is deliberately restricted to test items used in the English final examination for Grade X students at SMAK St. Bonaventura Madiun. Given this specific focus, the extent to which the findings can be applied to other educational settings is limited. The conclusions drawn are context-specific and reflect only the testing conditions and item characteristics

within the selected school environment. Therefore, the results may not represent assessment practices in other institutions, grade levels, or subject areas. In light of these considerations, the study provides meaningful insights into item construction within a particular setting, although its implications for broader contexts remain limited.

C. Research Question

Based on the background of the study above, the research questions are formulated as follows;

1. What is the difficulty level of the multiple-choice items in the English final examination administered at SMAK St. Bonaventura Madiun?
2. What is the discrimination index of the multiple-choice items in the English final examination administered at SMAK St. Boventura Madiun?
3. How effective are the distractor of the multiple-choice items in the English final examination at SMAK St. Bonaventura Madiun?

D. Purpose of the Study

1. To identify the difficulty level of the multiple-choice items in the English final examination at SMAK St. Bonaventura Madiun.
2. To analyze the discrimination index the multiple-choice item in the English final examination at SMAK St. Bonaventura Madiun.
3. To evaluate the effectiveness of the distractors multiple-choice items in the English final examination at SMAK St. Bonaventura Madiun.

E. Significance of the Study

The findings of the research hopefully will be helpful in few ways :

1. For the School

The findings of this study may contribute to the enhancement of English language assessment practices at SMAK St. Bonaventura Madiun. Through the analysis of final examination items based on difficulty level, discrimination index, and distractor effectiveness, the research is expected to offer practical insights that support the development of more reliable and valid assessments in English language instruction.

2. For the Teachers

An understanding of test item difficulty enables educators to align assessments with students' cognitive abilities and expected learning outcomes. The discrimination index assists teachers in distinguishing students who have achieved mastery from those who require further support, thereby informing instructional decisions. Additionally, effective distractor design promotes critical thinking and helps address common misconceptions. To this end, the research aims to support the improvement of assessment quality at SMAK St. Bonaventura Madiun by encouraging the development of more effective test items that enhance both instructional practices and students' learning outcomes in English education.

F. Definition of Operational Variables

The operational definitions formulated by the researcher in relation to the study are as follows:

1. Multiple Choice Items

Multiple-choice items refer to test questions that consist of a stem (the main problem or prompt) and a set of alternatives, including one correct answer and several distractors.

2. Analysis

Analysis refers to a systematic examination of multiple-choice items to evaluate item quality, effectiveness, and alignment with relevant educational standards.

3. English Final Semester

The English final semester refers to the concluding assessment period for Grade X students in English language instruction, typically conducted to evaluate students' comprehension and language skills.

4. Students

Students refer to Grade X learners at SMAK St. Bonaventura Madiun who participated in the English final semester examination. The responses collected from these students serve as the primary data source for evaluating the quality of the multiple-choice items.

5. Institution

SMAK St. Bonaventura Madiun's English teaching unit is responsible for designing, conducting, and evaluating the final semester exam to align with the school's learning goals and assessment standards.