ITEM ANALYSIS OF ENGLISH SUMMATIVE TEST AT VOCATIONAL HIGH SCHOOL

Risma Feby Wahyuningasih[1*], Taufiqulloh[2], Yuvita[3]

[123] *English Education Department, Universitas Pancasakti Tegal, Indonesia*

*)Corresponding author: Email: rismafeby01@gmail.com

Abstract

The objective of this Research Project is to find out the quality of test items in terms of item difficulty, discrimination power, and distractor effectiveness. This research also aimed to find out the result validity and reliability of summative tests at tenth-grade vocational high schools. This research was a descriptive quantitative study. The object of this research was test instruments of the summative test, including documents of teacher-made tests, key answers, and student responses of tenth-grade students at four different schools in Tegal Regency. The data was collected by documentation method and analyzed using the software program SPSS 27 version. The result of this research is that all the schools have good item difficulty index. Meanwhile, the quality of discrimination power for school C did not satisfy enough to distinguish higher and lower student performance. However, the items can still be accepted and improved in the future. The item quality of distractor effectiveness of school A 0,79 are accepted, school B 0,71 are accepted, school C 0,61 are accepted, and school D 0,94 are accepted. Based on the validity result, it was found that school A and school C have low valid points. Fortunately, all the school's test items are reliable, which means the tests are consistent. In summary, the items in the English summative test of the first semester for tenth-grade Vocational High School in Tegal Regency in the academic year 2023/2024 could still be used as a test instrument with some revisions and improvement from the test maker.

Keywords: *Item Analysis, English Summative Test, Tenth Grade Vocational High School.*

1. INTRODUCTION

Test play an important role in measuring students' understanding and progress in every subject, providing feedback to both students and teachers. The teacher will need a test as a tool of evaluation in order to evaluate the students (Nurbaeti et al., 2019). One of the types of tests is multiple-choice questions. Multiple-choice questions are widely used in educational assessments due to their efficiency and objectivity in measuring student knowledge and understanding across various subject areas (Haladyna & Rodriguez, 2013). However, the quality of these multiple-choice items is important to ensure their effectiveness in accurately measuring students' knowledge and skills. Ensuring the quality of multiple-choice questions requires an item analysis for the questions distributed to students.

The item analysis process involves statistical techniques to provide valuable information about the quality of individual test items. A good test should have the following criteria: high validity, reliability, objectivity, practicality and clear instruction (Fatimah et al., 2020). Analysis of multiple-choice items includes evaluating the characteristics of the items. These characteristics include item difficulty, discrimination power, distractor effectiveness, and content validity.

Item difficulty is one of the important aspects of analyzing multiple-choice questions; it refers to the difficulty level of each question for the students, whether it is categorized as hard, medium or easy level. Items that are consistently answered correctly by most students may be considered too easy and may not effectively discriminate between students with different levels of knowledge or skill. Conversely, items with low correct response rates may be overly difficult or ambiguous, leading to poor measurement accuracy. Understanding the difficulty of each item helps in ensuring that the test appropriately targets the intended knowledge or skills for the specific grade level. According to (Brown, 2004) is a statistical calculation to examine the percentage of students who answer a given item correctly. The proper item difficulty will provide an accurate measurement of student comprehension. Meanwhile, insufficient information about student understanding will be generated if the questions are too easy. On the other hand, students may not be able to answer correctly if the questions are difficult. Thus, it

will not give information about students comprehension accurately.

Discrimination power aims to distinguish between high achievers or upper group and students who do not perform well or lower group (Kumar et al., 2021). High discrimination power indicates that the item successfully differentiates between students who have good knowledge or skills and those who do not. In contrast, items with low discrimination power fail to differentiate significantly between students who achieve well and those who achieve poorly, according to (Brown, 2004). Discrimination power primarily, measures the ability of questions to differentiate students who have good understanding and those who are not able to understand yet. Thus, items with good discrimination power can help identify students with a deeper understanding of the material.

Distractor analysis assesses the effectiveness of incorrect response choices designed to confuse students. The option items considered to non-functional distractor when the options selected by less than five percent of the students (Kumar et al., 2021). The primary objective of analyzing distractors is to identify distractors that students frequently choose, as this indicates a lack of understanding. Effectiveness of distractor on the test items is when both of students of upper group and few students of lower group cannot answer correctly by choosing the most incorrect option, otherwise, the distractor items are ineffective when both of students of upper group and students of lower group able to answer correctly. Effective distractors should be reasonable to lower group, yet they must also be different from the correct response. If a significant number of individuals in the higher group can answer correctly while only a small number of the lower group can answer correctly, the test item will also be considered of good quality.

This research attempts to examine the item analysis of multiple-choice questions in the English summative test for tenth grade vocational high school in Tegal Regency.

2. METHODOLOGY

This study used descriptive quantitative research design. Four different vocational high school were selected as object of the study in academic year 2023/2024. The object consisted of one class student responses and the multiple-choice item teacher-made test for each school. The researcher involves several steps to analyze data. In supporting the analysis, the process of data analysis used a computing application named spss 27 version to generate item analysis results, validity and reliability scores. The analysis consists of item difficulty, discrimination power, distractor effectiveness, also validity, and reliability of the items to see if they meet the characteristics of effective test, which reflects to the item quality. Each school have a different number of questions and participants. The collected data were computed first to produce the statistical information needed. Then, the data proceeded descriptively to present the interpretation of statistical data. Shortly, the researcher presented the data in form of tabulation and interpreted it descriptively.

There are several steps to analyzing data as follows:

1. In the first step, the researcher collected all of the students' multiplechoice responses from four different schools.
2. In the second step, the researcher filled test information such as its key answer, number of questions, and correct and wrong points to the sheets.
3. In the third step, all the collected responses were computed in the SPSS 27 version.
4. In the last step, the program generated the result of item difficulty, discrimination power, validity and reliability score, then the result will appear.

3. RESULTS

To find out the result of item analysis, this research was collected the summative test in four different schools. Each school have a different number of questions and participants. The first school have 25 multiple-choice questions and 21 student participants, the second school have 20 multiple-choice questions and 36 student participants, the third school have 20 multiple-choice questions and 30 student participants, and the fourth school have 25 multiple-choice questions and 34 student participants. In total, there are 90 items and 121 student participants.

Table 1. Item Difficulty Result

Characteristic	Item				Percentage			
	School A	School B	School C	School D	School A	School B	School C	School D
Too diffic	7, 9, 15	10 an	1, 2, 3,	9, 16,	20 %	10 %	30 %	12 %

ult (< 0.30)	, 19 , and 2	d 13	6, 8, and 10	and 20				
Good (0.31 - 0.70)	1, 4, 6, 8, 10 ,11, 12 ,13, 16 ,17, 18 ,21, and 22	2, 4, 5, 6, 12 ,15, 16 , and 18	9, 13 ,14, 15 , and 16	1, 3, 4, 8, 10 ,11, 12 ,14, 18 ,21, 23 and 24	56 %	40 %	25 %	48 %
Excel lent (0.50 - 0.60)	2, 3, 5, 14 , and 20	8, 11 ,19, and 20	44 ,5, 7, and 18	6, 15 ,17, 19 ,22, and 25	20 %	20 %	20 %	24 %
Too easy (> 0.70)	25	1, 3, 7, 9, 14 , and 17	11 ,12, 17 ,19, and 20	2, 5, 7, and 13	4 %	30 %	25 %	16 %
Total	25	20	20	25	100 %	100 %	100 %	100 %

As a result of the data finding, Vocational High School A item difficulty result shows 14 items of good category dominates the items, Vocational High School B and Vocational High School C are almost evenly distributed, yet the good and excellent categories are higher. The Vocational High School D majority items are excellent. It can be indicated that all the schools have good item difficulty.

Table 2. Discrimination Power Result

Characteristic	Item				Percentage			
	School A	School B	School C	School D	School A	School B	School C	School D
Poor (< 0.20)	8, 9, 10 ,12, 17 ,18, 19 ,20, and 25	1, 4, 7, 11 ,12, 15 , and 19	1, 2, 3, 4, 5, 6, 7, 8, 10 ,12, 13 ,14, 15 ,17, and 18	3, 12 ,15, 17 ,18, 19 ,20, 21 ,22, and 23	36 %	35 %	750 %	40 %
Acceptable (0.21-0.24)	-	-	9 and 20	-	0 %	0 %	10 %	0 %
Good (0.25-0.34)	1, 14 ,21, and 22 .	9, 13 ,17, 18 , and 20	16 and 19	2, 4, 5, 7, 9, 12 ,14, and 16	16 %	25 %	10 %	28 %
Excellent (> 0.35)	2, 3, 4, 5, 6, 7, 11	2, 3, 5, 6, 8, 10 ,1	11	1, 6, 8, 10 ,11, 13	48 %	40 %	5 %	32 %

	,13,15,16,23, and 24	4, and 16		,24, and 25				
Total	25	20	20	25	100 %	100 %	100 %	100 %

High discrimination power means that the item successfully differentiates between higher group who have knowledge of the subject and lower group students who struggle, in contrast, items with low discrimination power fail to differentiate the groups. It was found that 48% of the items for Vocational High School A and 40% of the items for Vocational High School B were excellent, which means the items can distinguish the groups. However, Vocational High School C and D still have issues with the quality of the items to differentiate higher group and lower group students. It is important to note that discrimination power alone does not guarantee the quality of a test item, and it should be considered alongside other factors like item difficulty and overall test reliability.

Table 3. Distractor Effectiveness Result

Characteristic	Total Distractor				Percentage			
	School A	School B	School C	School D	School A	School B	School C	School D
Accepted	99	71	61	94	79 %	71 %	61 %	94 %
Revised	18	14	19	4	14 %	14 %	19 %	4 %
Discarded	8	15	20	2	6 %	15 %	20 %	2 %
Total	125	100	100	100	100 %	100 %	100 %	100 %

According to the findings, more than half of distractor items are accepted amounting to 79% for Vocational High School A, 71% for Vocational High School B, and 61% for Vocational High School C. Suppose a significant number of individuals in the higher group can answer correctly, and only a small number of the lower group can answer correctly. In that case the test item will also be considered of good quality. Effective distractors should be reasonable to low achievers, yet they must be different from the correct response.

Table 4. Validity and Reliability Result

School	Validity		Reliability	
	Score	Category	Score	Category
A	0,48	Valid	0,858	Reliable
	0,52	Invalid		
B	0,65	Valid	0,816	Reliable
	0,35	Invalid		
C	0,30	Valid	0,505	Reliable
	0,70	Invalid		
D	0,60	Valid	0,799	Reliable
	0,40	Invalid		

However, many invalid items were found on the test. Vocational High School A and C have more invalid questions than valid questions. Meanwhile, Vocational High Schools B and D have more valid items than invalid items. Fortunately, the results of reliability from four different vocational high schools show that all the test instruments are reliable., meaning the test instrument has consistency. If the same individuals are assessed multiple times with the same test or if different observers evaluate the same behavior or attribute, the results should be relatively consistent.

4. CONCLUSIONS

The average score of item difficulty for vocational high school a was 0,43, vocational high school b was 0,58, vocational high school c was 0,45, and vocational high school d was 0,52. The results indicate that overall items are good. It can be concluded that all of the items of questions meet the good test criteria. The discrimination power average point for vocational high school a was 0,31; vocational high school b was 0,32; vocational high school c was 0,06; and vocational high school d was 0,24. In short, only three of the schools had a good ability to distinguish higher

skilled and poor skilled groups. Fortunately, some questions could still be accepted and improved. More than half of the distractors on the items of the summative test at four schools could distract students' answers based on the accepted items. In term of validity analysis, vocational high school a and vocational high school c have low validity score which means the test are invalid. And for vocational high school b and vocational high school d validity score are fulfil the validity score to be said as valid test. In analyzing reliability, it was found that all of the schools having reliable score, means the test are consistent.

From the result of the analysis, it can be concluded that Vocational High School A have good index at overall analysis result and its reliability, but this school test does not have good validity score. In another school, Vocational High School B have good item analysis score, its validity and reliability score. For the next school analysis, Vocational High School C has strength index on its item difficulty and distractor effectiveness, and reliability result, yet the discrimination power average score is not meet satisfy result to distinguish more higher and lower group, aside from that the items also do not have good validity score. Last school, Vocational High School D have good result on the overall analysis and its validity and reliability.

The writer draws a conclusion that the items in the English summative test for Tenth grade Vocational High School in Tegal Regency in academic year 2023/2024 can still be used as a test instrument with some revisions and improvement.

REFERENCES

Brown, H. D. (2004). *Language Assessment: Principles and Classroom Practices*.

Fatimah SMPN, S., Bernhardo Elzamzami SMPN, A., & Slamet, J. (2020). *Item Analysis of Final Test for the 9 th Grade Students of SMPN 44 Surabaya in the Academic Year of 2019/2020* (Vol. 2, Issue 1).

Haladyna, T. M., & Rodriguez, M. C. (2013). *Developing and validating test items*.

Kumar, D., Jaipurkar, R., Shekhar, A., Sikri, G., & Srinivas, V. (2021). Item analysis of multiple choice questions: A quality assurance test for an assessment tool. *Medical Journal Armed Forces India*, *77*, S85–S89.

Nurbaeti, A. P., Taufiqulloh, T., & Sulistyawati, A. E. (2019). Items Analysis of The Achievement Tests in EFL Classrooms. *English Focus: Journal of English Language Education*, *3*(1), 32–42.