

## AN ANALYSIS OF FINAL TEST QUALITY ON EFL STUDENTS

Firman Ahmad Faizal<sup>1</sup>, Syahban Mada Ali<sup>2</sup>

*IAIN Parepare*<sup>1</sup>, *Institut Ilmu Sosial dan Bisnis Andi Sapada*<sup>2</sup>  
firmanahmadfaizal@gmail.com<sup>1</sup>, syahban.syan@gmail.com<sup>2</sup>

### Article History

Received:  
March 7, 2022  
Revised:  
September 13, 2022  
Accepted:  
September 19, 2022  
Published:  
September 30, 2022

### Abstract

This research determines the quality of the final test in terms of difficulty level, discrimination level, distracter effectiveness, validity, and reliability for senior high school students in second grade at SMA Negeri 3 Parepare. This research used a quantitative design with content analysis. The investigation was conducted at SMA Negeri 3 Parepare. The subject was 134 English Final Test Form of the second grade. However, the researcher took 58 students from three classes to serve as test subjects, with the sample for this research being retrieved using a random selection group technique. They were XI MIPA 1, XI MIPA, and XI IPS 1. The conclusion of the final test for senior high school at the second grade SMA Negeri 3 Parepare was of medium difficulty (p), meaning that the test was neither too easy nor too tough. On the discrimination (D) scale, there were 6 (20%) items with outstanding classification, 1 (3,33%) item with adequate classification, and 23 (76,6%) items with good classification. Distracter effectiveness is likewise weak, since 12 (40%) of the items have very poor distractions. On the validity 9 (30%) items were invalid and needed to be replaced. However, 21 (70%) items were valid. The final test for senior high school at the second grade SMA Negeri 3 Parepare is reliable since the reliability index 0,972, it was higher than the table value of critical of product moment. It has shown that the final test was reliable.

**Keywords:** *UAS, English Test, Quality of Test.*

### Introduction

One of the most important part in the learning process is a test. The aim of the learning process can be got or achieved when students pass the test and have standards scores that have been established by the teachers (Hasanah et al., 2020). When students cannot pass the test, most of the teachers think that the student's ability is decreasing without regard to quality standards in designing the test. They are only doing their duty to design the test.

Student achievement can be measured by giving a test and the test made should through a validation (Bachman, 2002). However, the accuracy of the

information can only be got by measuring accuracy, which means the test of good. Based on information got informally on February 13, 2020, by interviewing several teachers who taught in a school of Parepare, they only designed, created, and checked the final test without analyzing the test.

Related to it, when the researcher did Field Experience Program (FEP) in one school in Parepare, he found that some teachers shared each other for the item of test through short message service or Whatsapp. They made the test form of 50% for the third grade in senior high school, 30% for the second grade, and 20% for the first grade, but sometimes they didn't make it on their own. They just asked another teacher for the items of the test. The excuse they made for this trend were; a) too many classes and teaching hours, b) some teachers teach over one school, c) tired after doing many activities at school.

According to the researcher's interview, several problems affected the weaknesses of the test made by the teacher. Some of them were better on time, teacher's energy, and cost. However, the main factor of the failure was the teacher's own opportunity to do a trial run for the test or tested the test to measure and analyzed the item of the test.

Generally, because the test is made by the teachers their selves, so they only use it for private, it is meant only for their students and not for the public. That's the reason they rarely test the test. Because they don't test the test so the teacher tests are not equipped with information about the quality of the items, validity, and reliability of the items.

Research on the final test result so far has only been conducted regarding the result as NEM and suspicion about the item of test it's started in Budiharso et al.'s research (2020) about the validity and reliability of English tests for senior high school, Riandy's research (2018) about the item analysis of English final test for junior high school and Astutia's research (2019) about the quality of English final test for junior high school. The three studies show that the items of each question and the items in the English final test are indeed low. We can imagine that not only because of the decreasing ability of students but also the quality of English final tests made by the teacher can be a consideration and included in the factor that causes student scores to decline. Every test given to the students should be proper in every ethnic and district (Pudjastawa et al., 2021), including in Parepare. the Therefore, the researcher is inspired to know about the quality of English teacher final test for senior high school in Parepare.

This current research was conducted in SMA Negeri 3 Parepare in the academic year 2020-2021, which focused on the second-grade students. This school is one of the best senior high schools in Parepare. It is also a school that is noted to be fast using curriculum 2013 or K-13 as their standard competence for teaching and learning since 2013. It proves that the school is quick to adapt to new changes, especially the

new curriculum and which means students can learn well even with the new curriculum.

Here, this research was limited only for analyzing the quality of final test, where the test has been held only for senior high school in the second grade.

### Subject of the Research

The subject of this research was the 58 sheets English final test items used to test the students who are registered as the first-year students in the academic year 2020-2021 for second grade of senior high school SMA Negeri 3 Parepare incorporated in concurrent test execution for the second grade of senior high school SMA Negeri 3 Parepare that comprising three classes. They were XI MIPA 1, XI MIPA, and XI IPS 1, and the total population was 134 students. Yet, the researcher only took the 58 students from the three classes as test participants, where the sample of this research retrieval using a random sampling group technique.

Table 1. The Students Sample of the Second Grade of SMAN 3 Parepare

No	Class/Grade	Students
1	XI MIPA 1	29
2	XI MIPA 2	19
3	XI IPS 1	10
	<b>TOTAL</b>	<b>58</b>

### Method

The researcher used three steps to collect the data. First, the test answer sheet was identified and analyzed for each item of the test with the sign of true and false. True-false identification produced a record of the number of items that are answered correctly and answer wrong. Second, split the answer sheet into two groups, that was the top group and bottom group 29 sheets in each of the groups. This step produced data to identify P and D for each item. After these two steps are done, interpretation was also carried out to see the effectiveness distracter, validity, and reliability.

### Results

#### 1. The Difficulty of The Final Test

They have done the final test years 2020/2021. Then these 58 sheets of English teacher final test, which was used as a sample. The findings show that all test items were at a medium level. For clarity, the researcher provides a table that gave a brief description of the difficulty level of each item.

Table 2. The Difficulty Level Analysis

Item	P	Classification	Difficulty Level
1	0,64	$0,30 < P \leq 0,70$	Medium
2	0,66	$0,30 < P \leq 0,70$	Medium

<b>Item</b>	<b>P</b>	<b>Classification</b>	<b>Difficulty Level</b>
3	0,64	$0,30 < P \leq 0,70$	Medium
4	0,57	$0,30 < P \leq 0,70$	Medium
5	0,55	$0,30 < P \leq 0,70$	Medium
6	0,59	$0,30 < P \leq 0,70$	Medium
7	0,61	$0,30 < P \leq 0,70$	Medium
8	0,64	$0,30 < P \leq 0,70$	Medium
9	0,61	$0,30 < P \leq 0,70$	Medium
10	0,42	$0,30 < P \leq 0,70$	Medium
11	0,64	$0,30 < P \leq 0,70$	Medium
12	0,61	$0,30 < P \leq 0,70$	Medium
13	0,67	$0,30 < P \leq 0,70$	Medium
14	0,61	$0,30 < P \leq 0,70$	Medium
15	0,61	$0,30 < P \leq 0,70$	Medium
16	0,54	$0,30 < P \leq 0,70$	Medium
17	0,69	$0,30 < P \leq 0,70$	Medium
18	0,62	$0,30 < P \leq 0,70$	Medium
19	0,59	$0,30 < P \leq 0,70$	Medium
20	0,51	$0,30 < P \leq 0,70$	Medium
21	0,66	$0,30 < P \leq 0,70$	Medium
22	0,54	$0,30 < P \leq 0,70$	Medium
23	0,57	$0,30 < P \leq 0,70$	Medium
24	0,62	$0,30 < P \leq 0,70$	Medium
25	0,62	$0,30 < P \leq 0,70$	Medium
26	0,59	$0,30 < P \leq 0,70$	Medium
27	0,61	$0,30 < P \leq 0,70$	Medium
28	0,57	$0,30 < P \leq 0,70$	Medium
29	0,64	$0,30 < P \leq 0,70$	Medium
30	0,62	$0,30 < P \leq 0,70$	Medium

There are four columns on the table; the first column shows information about the number of the test, the second column shows information about the result of P level analysis, the last column shows information about the difficulty of classification. And, the fourth column shows information about difficulty level status. Then, to get the P level of the test, the researcher used the formula of Bachman.

For example, at the first item of the test there were 58 students who have taken the test, and the test has 30 questions. All the questions were multiple choice. And, out of 58 students there were 38 students answered correctly. The first item of the test, the index of the difficulty level of the test, was 0,64.

## 2. The Discrimination of the Final Test

The data of findings shows that the final test for second grade SMA Negeri 3 Parepare academic year 2020/2021 have six items were excellent classification, one item satisfactory and the rest of the items have good classification. For clarity, the researcher provides a table that gave a brief description of the discrimination of each item.

Table 3. The Discrimination Level Analysis

Item	D	Classification	Interpretation
1	0,69	Good	The test item has good discrimination power
2	0,66	Good	The test item has good discrimination power
3	0,69	Good	The test item has good discrimination power
4	0,75	Excellent	The test item has excellent discrimination power
5	0,87	Excellent	The test item has excellent discrimination power
6	0,66	Good	The test item has good discrimination power
7	0,76	Excellent	The test item has excellent discrimination power
8	0,69	Good	The test item has good discrimination power
9	0,55	Good	The test item has good discrimination power
10	0,44	Good	The test item has good discrimination power
11	0,62	Good	The test item has good discrimination power
12	0,69	Good	The test item has good discrimination power
13	0,63	Good	The test item has good discrimination power
14	0,76	Excellent	The test item has excellent discrimination power
15	0,69	Good	The test item has good discrimination power
16	0,68	Good	The test item has good discrimination power
17	0,52	Good	The test item has good discrimination power
18	0,65	Good	The test item has good discrimination power
19	0,66	Good	The test item has good discrimination power
20	0,76	Excellent	The test item has excellent discrimination power
21	0,66	Good	The test item has good discrimination power
22	0,83	Excellent	The test item has excellent discrimination power
23	0,69	Good	The test item has good discrimination power
24	0,45	Good	The test item has good discrimination power
25	0,52	Good	The test item has good discrimination power
26	0,31	Satisfactory	The test item has medium discrimination power.
27	0,55	Good	The test item has good discrimination power
28	0,62	Good	The test item has good discrimination power
29	0,69	Good	The test item has good discrimination power
30	0,65	Good	The test item has good discrimination power

The table comprises 4 columns. The initial column contains data about the number of tests. The second column displays the results of the descriptive analysis. The third column contains data on classification D. The fourth column contains data on interpreting comparisons. The researchers used a recipe adapted from Proyonigor.

### 3. The Effectiveness of the Distracter

The data of the findings show that several distractors have scores under 5% and to be a good distractor the score should be more than 5%. For clarity, the researcher provides a table that gave a brief description of the effectiveness of distractors of each item.

Table 4. The Effectiveness of the Distractor Analysis

Item	Option					Information
	A	B	C	D	E	
1	20,33	(38)	10,1	1,6	1,6	
2	16,9	8,47	1,6	(39)	1,6	
3	(39)	18,64	10,16	3,38	0	
4	11,86	8,47	8,47	11,86	(34)	
5	15,25	18,47	8,47	(33)	0	
6	10,16	8,47	16,94	(35)	3,38	
7	13,55	15,25	(36)	5,08	3,38	
8	(39)	6,77	11,86	8,47	5,08	
9	(36)	5,08	15,25	10,16	6,77	
10	20,33	6,77	5,08	11,86	(32)	
11	5,08	6,77	10,16	11,86	(38)	
12	15,25	6,77	10,16	(35)	5,08	
13	(39)	8,47	10,16	6,77	6,77	( ): Kunci Jawaban
14	15,25	13,55	(37)	5,08	1,6	
15	3,38	(37)	15,25	10,16	6,77	
16	15,25	15,25	10,16	(32)	3,38	
17	(42)	6,77	5,08	8,47	6,77	
18	11,86	(32)	6,77	8,47	8,47	
19	11,86	(35)	5,08	16,9	5,08	
20	8,47	22,03	(30)	8,47	8,47	
21	10,16	(40)	10,16	6,77	3,38	
22	13,55	10,16	(33)	6,77	10,16	
23	(34)	11,86	8,47	15,25	5,08	
24	15,25	8,47	(36)	10,16	3,38	
25	(35)	11,86	11,86	8,47	6,77	
26	22,03	3,38	5,08	(32)	13,55	

Item	Option					Information
	A	B	C	D	E	
27	11,86	5,08	(36)	11,86	8,47	
28	13,55	(34)	13,55	8,47	5,08	
29	11,86	6,77	(38)	11,86	3,38	
30	11,86	5,08	11,86	(37)	5,08	

From the table above, there are seven columns; the first column to show the item, second, third, fourth, fifth, and sixth column to show the result of the effectiveness of distracter analysis, and the last column to show the information of the answer key and the answer key has been written by the researcher with closed brackets and open brackets. And, to get the effectiveness score from the test, the researcher used a formula adopted from Sudjiono.

We can conclude that the result of the existing data of the effectiveness of distracter of final test for second grade SMA Negeri 3 Parepare reported that there are several some distractions that need to be replaced because it has low distracting power that is lower than 5%. As we know, a good distracter is it can be avoided by smart students and selected by the less intelligent students. If the distracter is selected 5% by the number of students, it means that the distracter has a good function.

From those data and statement above, the writer could conclude that the distracter function in the final test for second grade student SMA Negeri 3 Parepare have a poor function so its need to be replaced and it have to find a new distracter which has a good function.

#### 4. The Validity of the Final Test

The data of the findings show that twenty one item of the final test were valid and nine item were invalid. For clarity, a table that gave a brief description of the effectiveness of distractors of each item provides below.

Table 5. The Validity Analysis

Item	Correlation	Table	Status
1	0,734	0,678	Valid
2	0,766	0,678	Valid
3	0,695	0,678	Valid
4	0,714	0,678	Valid
5	0,834	0,678	Valid
6	0,671	0,678	Invalid
7	0,871	0,678	Valid
8	0,773	0,678	Valid
9	0,674	0,678	Invalid
10	0,532	0,678	Invalid

<b>Item</b>	<b>Correlation</b>	<b>Table</b>	<b>Status</b>
11	0,744	0,678	Valid
12	0,761	0,678	Valid
13	0,685	0,678	Valid
14	0,808	0,678	Valid
15	0,743	0,678	Valid
16	0,788	0,678	Valid
17	0,625	0,678	Invalid
18	0,755	0,678	Valid
19	0,694	0,678	Valid
20	0,798	0,678	Valid
21	0,752	0,678	Valid
22	0,855	0,678	Valid
23	0,748	0,678	Valid
24	0,536	0,678	Invalid
25	0,571	0,678	Invalid
26	0,399	0,678	Invalid
27	0,639	0,678	Invalid
28	0,643	0,678	Invalid
29	0,765	0,678	Valid
30	0,717	0,678	Valid

There are 4 columns in the table; initial column shares data about the test no, column that shares data about the results of the validity analysis, column that shares data on the table of product-moment critical values with a significance level of 95%, and column that shares data about the validity status. Not only that, to get test validity.

Based on the table above, the author can plan that the results of the available test information tell that twenty-one items are valid and 9 items are invalid. This is quite a reflection of the state of the final test of the second grade students of SMA Negeri 3 Parepare.

Not only that, an item is said to be valid if the correlation coefficient of each item is greater than or equal to the product-moment critical value table with a significance level of 95%. Therefore, questions that have been valid in the test do not need to be exchanged and can be placed in the question bank for future use, but invalid test items need to be exchanged.

## **5. The Reliability of the Final Test**



The data shows that the final test of class II SMA Negeri 3 Parepare is reliable with a reliability index of 0.972. This reliability works with standard indexes. A product is said to be reliable if the correlation coefficient for each point is greater than or equal to the critical value table for product moment with a significance level of 95% (Arikunto, 2010). For more details, the researcher presents the reliability analysis table as follows:

Table 6. Reliability Analysis

<b>Correlation</b>	<b>Table</b>	<b>Status</b>
0,972	0,678	Reliable

The table contains 3 columns. Its initial column shares data about the link. In the second column, it shares data on the critical moment table with a significance level of 95%. The third column shares data about the validity status.

Based on the table above, the author concludes that the reliability score of the item with Spearman-Brown (2013) product moment is greater than the reliability of the final test of class II students of SMA Negeri 3 Parepare, because the reliability index is 0.972. We can plan that the marker is reliable. This is greater than the critical value of the moment table. Reliability is the extent to which the same score or score is opposite when the same test is assessed by 2 or more different examiners or by the same examiner on different occasions. For the test to be reliable, the measurement must be constant.

## **Discussion**

### **1. The Difficulty Level of Final Test**

Based on the findings, the results of the information in the last test difficulty show that all test items are at another level. As shown by the information, the difficulty index is between 0.30-0.70 and more. tier items. This means that the test can be placed in a question bank and can be useful or reused for further tests. The following could be an example for making a new test for class II students of SMA Negeri 3 Parepare.

Not only that, a good test is a test that is not very easy or the opposite is very difficult for students. Should share answer options that can be selected by students and near to the answer key. And the last test carried out by the teacher for class II SMA Negeri 3 Parepare already had these criteria, very easy questions can also be formed into several things that affect the feeling of "success" in low-ability students and are used as warm-up questions, and very difficult items. Can be a challenge for the most highly skilled students. A good test ensures students recognize and note the characteristics of the teacher's test that the test always becomes very easy and very

difficult for them (Ali, 2020). Therefore, the standard mandatory test and meet the characteristics of a good test. The text of the test should fill the need of the students' guidance reading (Ali et al., 2022).

From the statement above, the author can plan that the final test made by the teacher for Grade 2 has the standard criteria of SMA Negeri 3 Parepare. It is not very easy or vice versa, very difficult for students.

## **2. The Discrimination of the Final Test**

Referring to the result of the data of final test for second grade SMA Negeri 3 Parepare academic year 2020/2021 has six items were excellent classification with have score range 0,70-1,00. One item satisfactory to have a score range 0,20-0,40 and the rest of the items have good classification with have score 0,40-0,70.

Based on statement above, the writer could conclude that it is a good result because there were no negative signs in each item when a test item inversely to shows the quality of test items, namely smart students are considered not smart, and less intelligent students are smart. And, the final test for second grade SMA Negeri 3 Parepare.

A good test can be answered correctly by smart students. For example, if a group of smart students can answer the test correctly and the groups of students who are less intelligent answer incorrectly, it means that the test has excellent discrimination power.

Based on statement above, the writer could conclude that the final test for second grade SMA Negeri 3 Parepare has an excellent discrimination level it means that students who master the subject that has been taught by the teacher have answered the test correctly it is showed by the data have a good discrimination level.

## **3. The Effectiveness of the Distracter**

Based on the findings, the outcome of the existing data of the effectiveness of distracter of final test for second grade SMA Negeri 3 Parepare reported that there are several some distractions that need to be replaced because it has low distracting power that is lower than 5%. As we know, a good distracter is, it can be avoided by smart students and selected by the less intelligent students. Also, if the distracter is selected 5% by the number of students, it means that the distracter has a good function.

However, form the statement above the writer could conclude that the distracter function in the final test for second grade student SMA Negeri 3 Parepare have a poor function so its need to be replaced and its have to find a new distracter which has a good function.

## **4. The Validity of the Final Test**

Based on the findings, the outcome of the existing data of the test reported that twenty-one items were valid and nine items were invalid. This simply provides us

a point about the current condition of the final test for second grade student SMA Negeri 3 Parepare.

Besides, an item is stated valid if the coefficient correlation of each item is higher or equal to the table of the critical value of product-moment with the level of significance 95%. Therefore, items that are already valid in the test do not need to be replaced and can be saved in the question bank for future use, but the test items were invalid need to be replaced.

### **5. The Reliability of the Final Test**

Referring to the result of the data, the reliability of the result using the Spearman-Brown (2013) product moment showed that the reliability index of the final test for the sophomore SMA Negeri 3 Parepare was reliable as the index reliability was 0.972, which was higher than the product-moment critical value table. Reliability is the extent to which the same grades or marks are awarded when the same tests are graded by two or more different examiners or by the same examiner on different occasions. In short, to be reliable, a test must be consistent in its measurement. This just gives us a point about the current state of the final exam for the sophomore SMA Negeri 3 Parepare. It is the degree to which a test consistently measures whatever it measures. The researcher realizes that this study contains a weakness. Since it is a quantitative descriptive study that analyzed the final test for the second-year student SMA Negeri 3 Parepare in relation to the difficulty, discrimination level, distracter effectiveness, validity and reliability of the final test, this research will be more useful if the researcher helps the teacher if necessary to reshape the evidence element of the final examination. By this, it is also important that the material of the final test should closely related to the culture of students (Rahman et al., 2022).

### **Conclusion**

Alderson (1995) said, "Language Test Construction and Evaluation describes the process of language test construction clearly and comprehensively". After conducting the research data and doing analysis data, the difficulties of final test items SMA Negeri 3 Parepare have a medium level in difficulty level and it was a good sign because the test had a medium level, not too easy, or vice versa too difficult for students. The discrimination level of final test items SMA Negeri 3 Parepare have a good result there were not a negative sign in each item, and students who master the subject that have been taught by the teacher have answered the test correctly it is showed by the data that have a good discrimination level. The effectiveness of distractions in several items was not effective and several distractions need to be replaced, even removed. For example; distracter D and E in the item number one, two distractions have score 1,6% less than 5% and, we can say the two distractions have no good effectiveness. The validity of the final test SMA Negeri 3 Parepare there were nine items (30%) were in valid, item number 6, 9, 10, 17, 24, 25,26,27, and, 28, these

items need to be replaced and twenty-one items (70%) were valid. The final test items were reliable since the reliability index was 0.972, which was higher than the table value of critical of product-moment, it has shown that the final test items were reliable.

## References

- Alderson, J. C., Clapham, C., & Wall, D. (1995). Language test construction and evaluation. In Cambridge language teaching library.
- Ali, S. M. (2020). ENHANCING STUDENTS' READING ABILITY THROUGH CNA INDONESIA. *JURNAL NALAR PENDIDIKAN*, 8(2). doi: 10.26858/jnp.v8i2.15334
- Ali, S. M., & Hasanah, N. (2022). Applying High Order Thinking Skill (HOTS) On Teaching Reading Skills for the X Grade Students of SMAN 2 Kalukku. *Inspiring: English Education Journal*, 5(1), 11–21.
- Arikunto, S. (2010). Dasar-dasar Evaluasi Pendidikan. Jakarta: Bumi Aksara. 2010. *Manajemen Penelitian*.
- Astutia, N. F. W., & Astuti, R. D. (2019). Impact of undernutrition in early life on adolescents' nutritional status and academic performance: Evidence from Indonesia. *Annals of Nutrition and Metabolism*, 75(3).
- Bachman, L. F. (2002). Some reflections on task-based language performance assessment. *Language Testing*, 19(4). doi: 10.1191/0265532202lt240oa
- Brown, H. D. (2013). Language Assessment Principle and Classroom Practice. In Journal of Petrology (Vol. 369, Issue 1).
- Budiharso, T., & Tarman, B. (2020). Improving quality education through better working conditions of academic institutes. *Journal of Ethnic and Cultural Studies*, 7(1). doi: 10.29333/ejecs/306
- Hasanah, N., & Mada Ali, S. (2020). The Students' Motivation in Writing through Mind Mapping at Tomakaka University of Mamuju. *Edumaspul: Jurnal Pendidikan*, 4(2). doi: 10.33487/edumaspul.v4i2.837
- Pudjastawa, A. W., Cantika, V. M., & Dwijonagoro, S. (2021). Development of Online Exam Questions for Javanese Language Subjects with a Cultural Responsive Assessment Approach. *KnE Social Sciences*. doi: 10.18502/kss.v6i2.9995
- Rahman, A., Ali, S. M., & Badriah. (2022). Integrating Local Culture-Based Instruction in Teaching Speaking Classroom at the Third Semester Students of Universitas Tomakaka Mamuju. *Inspiring: English Education Journal*, 5(1), 64–72.
- Riandi, R., Apriliana, V., & Purwianingsih, W. (2018). *The Analysis of 21st Century Teachers' Ability in Technological Pedagogical Content Knowledge*. doi: 10.2991/icei-18.2018.60