

An Analysis of The English Summative Test: EFL Teacher- Made Test

P-ISSN 2721-1096

E-ISSN 2721-1916

Yessicka Noviasmy

Institut Agama Islam Negeri Parepare

Risma

Institut Agama Islam Negeri Parepare

Abstract

This study aims to determine the quality of the multiple-choice summative test items for senior high school EFL teachers in Parepare. The analysis consisted of validity, reliability, level of difficulty, and distractors. The type of research used in this research was quantitative descriptive. The subjects in this study were 33 students. The data in this study was collected using documentation techniques. Data analysis in the form of item analysis was carried out using the Biserial point correlation. The results of the study showed that the valid category was 80% and invalid are 20%. Included in the unreliable test with $KR-20 < 0.70$ or 0,6. The difficult category was 1 item, the moderate category was 8 items, and easy category was 1 item. Distractors function very good was 70%, good category was 20%, and the quite good category was 10%. It can be resumed that the final multiple-choice test made by EFL teacher has very good quality of item validity, unreliable test, good quality level of difficulty, and very good quality distractor. Based on the result analysis of test items, it can be concluded that several good test items can still be used for the future summative test, and several test items should be revised.

Key words

Item Analysis, Summative Test, Teacher-Made Test

Corresponding author:

Yessicka Noviasmy, Institut Agama Islam Negeri Parepare, Parepare, Indonesia

Email: yessickanoviasmy@iainpare.ac.id

INTRODUCTION

The implementation of evaluation in the learning process is an important component in the teaching and learning process. Evaluation will reflect the success of teaching and more importantly the achievement of the students. Therefore, EFL teachers should develop a good test to measure students' achievement. Test is a way of measuring and assessing in the field of education in the form of giving assignments or a series of tasks that must be done by students so that students' achievement scores can be known (Putri, 2009). The quality of a test can be seen from the characteristics of the questions used, a test with good quality will have good items. To find out whether the questions have good quality, each item must be analyzed. An analysis is a systematic procedure that will provide very specific information on something very detailed, such as an item analysis of what has been compiled. Quality questions are questions that can provide information as accurately as possible so that it can be seen that they have mastered the material and those who have not.

The teacher should make semester exam questions, and the questions that are made must be proportionally appropriate. The items have good and functioning instruments, meaning that with these instruments (Kalsum et al, 2023), the measuring instruments used contain the material to be measured and follow the compilers of the questions. Therefore, teachers must pay attention to the quality that will be tested on students to obtain the correct results so that the quality of students can be guaranteed.

Teacher rarely does trials on the questions to be used, including analyzing the quality of each item to be tested, so most of them cannot identify good or bad tests. This is due to the lack of teacher time and teacher understanding (Sardi et al, 2017). Even though analyzing the items is an activity that must be carried out by a teacher. According to Nurgiyantoro, the reason for doing item analysis is because it will produce quality questions on subsequent tests and to find out the strengths and weaknesses of items so that items can be selected, revised, and the problem with the item found immediately (Nurgiyantoro, 2010). If the tests made by the teacher are not in accordance with what has been determined, it will have an adverse effect on students; namely, the interests, talents, and understanding of students cannot be measured, so the teacher cannot classify students who have high abilities and students who have low abilities.

According to the findings of an interview with an English teacher at UPT SMKN 3 Parepare named St. Ruwaedah, she never does a test analysis of the final test that has been based on validity, reliability, level of difficulty, discrimination index, and distractors. In addition, the teacher also does not really understand the problem of the test instrument, and the teacher does not have time to analyze the instrument test that has been made to determine whether the exam questions that have been made meet the standards, terms, and conditions contained in the assessment tool.

This research is very important because it is to find out the quality of the test made by the teacher. The quality of the test in question is whether the test is valid, reliable, the questions are of good quality, the questions are easy to answer or difficult

to answer, and the distractor functions properly. Based on the description above, it is necessary to analyze the items in detail quantitatively to test results for validity, reliability, level of difficulty, discrimination index, and distractors.

LITERATURE REVIEW

Maya Marsevani conducted a study aimed at determining the quality of multiple-choice questions in a public elementary school in terms of difficulty level, discrimination power, and distractor. To gather data and assess MCQs on student assessments, this study used a cross-sectional survey. 40 students are the study's subject. This study demonstrated that the item discrimination power was strong and that the majority of the difficulty index was acceptable. Two objects were entirely effective as a distractor.(Marsevani, 2022)

Maya's research employed a cross-sectional study to obtain information and evaluate multiple-choice questions in the students' tests, while the research conducted by the prospective researcher used documentation and quantitative descriptive. Maya's research does not used validity and reliability, while the research conducted by the prospective researcher used validity and reliability. The similarity of research conducted by prospective researchers is the focus of research on the quality of the English teacher-made test.

The novelty in this research is to analysis of the English Teacher-Made test for second grade students at UPT SMKN 3 Parepare based on 5 aspects namely: validity, reliability, level of difficulty, discrimination index, and distractors.From some of the research results that have been described above, there are quite clear differences between previous research and the research that carried out by researchers. Therefore, this time the researcher tries to examine An Analysis of the English Teacher-Made Test for Second Grade Students at UPT SMKN 3 Parepare based on validity, reliability, level of difficulty, discrimination index, and distractors.

According to Suharsimi Arikunto, a teacher-made-test is a test written and made by a teacher in the school, so the validity and reliability of the test is not like a standardized test.(Suharsimi, Arikunto, 2013) The effectiveness of this type of test depends on the skill and ability of the teacher in designing the test. This test is based on materials and specific goals formulated by the teacher for his own class. Rarely does the teacher analyze and revise test items that have been tried, so the teacher does not know the validity, reliability, level of difficulty, discrimination index, and distractors. Test is said to have good quality when have validity, reliability, high discrimination index, medium difficulty, and a working distracter.(Dwipayani, 2013)

Item validity is the ability to measure exactly test made English teacher. A test or measurement device is said to have high validity if the test performs the measuring function correctly or provides measurement results that are in accordance with the purpose of the measurement. An indicator of how much a measure device can be trusted or relied upon is called reliability. If a test yields the same findings when

administered to same group of students over different period, it is said to be reliable. Level of difficulty is a measure that indicates how difficult or easy a test is. Good questions are questions that are not too easy and not too difficult. Discrimination index is the ability of a question to distinguish between students who are able and students who are less able to answer the question. The discriminatory power of questions will examine test in terms of the test's ability to distinguish students who are in the high and low achievement categories. The effectiveness of the distractor is the distribution in determining the answer choices on multiple-choice questions. The pattern of the answer to the question can determine whether the distractor is functioning properly or not.

RESEARCH METHOD

The type of research used in this research was quantitative descriptive. Quantitative Descriptive is research aims to describe the situation precisely and accurately, not to look for a relationship between the independent variable and the dependent variable or to compare two or more variables to find causation.(Paramita et al., 2021) The research location was UPT SMK 3 Parepare in Jl. Karaeng Burane No.16, Ujung District, Parepare, and South Sulawesi.

The population of this research was all students of class XI Multimedia at UPT SMKN 3 Parepare for the 2022/2023 academic year, consisting of 101 students. In this study, the researcher used a purposive sample. Purposive sample is a sample selected based on certain considerations with the aim of obtaining a sample that has the desired characteristics.(Setiawan, 2005) The researcher selected XI Multimedia 1 with 33 students as sample.

This final exam consists of fifteen (15) questions, 10 of which are multiple-choice and 5 of which are essay questions. In order to find the data, and the students' answers can be more easily corrected, the researcher only analyze the final test in the form of multiple-choice. To get the data in the form of test scores, the researcher carried out the result of the students' works. The students's result work are the students' answers which recorded in the answer sheets.

The data in this study was collected used documentation and the formula used in calculate the validity, reliability, level of difficulty, discrimination index, and distractors. In this study, the technique used by researchers is documentation. Documentation is data collection by viewing data or information by studying written or recorded data. The documentation method used to obtain data is test made by an English teacher, answer key, and student answers.

FINDINGS AND DISCUSSIONS

a. Validity

In testing validity, researchers used the Biserial Point Correlation formula. This calculation used criteria that are below 0.30 when entering invalid criteria and above or equal to 0.30 when entering valid criteria. Based on the results of the analysis with 10 multiple-choice questions and 33 students in the XI Multimedia Class English subject at

UPT SMKN 3 Parepare, the results of the validity test showed that 2 or 20% of the items are invalid and 8 or 80% of the items are valid.

Table 1. The result of an English Teacher-Made Test Analysis Based on the Validity.

No	Category	Item	Total	Percentage
1.	Valid $r_{pbis} \geq 0.3$	1, 2, 3, 5, 6, 7, 9, and 10.	8	80%
2.	Invalid $r_{pbis} < 0.3$	4 and 8	2	20%

Based on the results analysis validity of the multiple-choice items for English subjects made by an English teacher at UPT SMKN 3 Parepare and which have been answered by 33 XI Multimedia 1 students, it shows that the multiple-choice questions that are included as valid questions are 8 or 80% items, and 2 or 20% of the items are invalid. This means that most of the items are valid. The results of this study were almost the same as the results of research from Lukmanul Hakim and Irhamsyah which showed that there were 23 valid items (92%) and 2 items (8%) included in the invalid category.(Hakim & Irhamsyah, 2020)

b. Reliability

In calculating reliability, this study used the KR-20 formula. Testing the reliability of the questions based on the guidelines: if the value of the calculated reliability is greater than or equal to 0.70, then the questions tested have high reliability, but if the calculated reliability values are less than 0.70, then the questions tested have low reliability or are not reliable.

Based on the KR-20 formula for calculating reliability, the result of multiple choice test analysis made by one of the teachers at UPT SMKN 3 Parepare is 0.6. Based on the results of these calculations, it can be concluded that the multiple-choice test is unreliable. The results of this study are the same as the research conducted by Taufiq Effendi and Illza Mayuni which showed reliability results with a value of 0.48 which means unreliable.(Effendi & Mayuni, 2022)

c. Level of difficulty

Based on the formula used in calculating the level of difficulty, there are criteria for calculating these items. These criteria are: questions with a difficulty level of 0.00–0.30 are difficult questions; questions with a difficulty level of 0.31-0.70 are moderate questions, and questions with a difficulty level of 0.71–1.00 are easy questions.

Table 2. The result an English Teacher-Made Test Analysis Based on Level of Difficulty.

No	Category	Item	Total	Percentage
1.	Difficult (0.00 – 0.30)	8	1	10%
2.	Moderate (0.31 – 0.70)	1, 2, 3, 4, 6, 7, 9, and 10	8	80%
3.	Easy (0.71 – 1.00)	5	1	10%

The results of this study are the same as those conducted by Mutiara Kusumawati and Samsul Hadi, whose results of the analysis of the level of difficulty are more dominant in the medium category with a score of 60%. The test that is used for final test purposes was supposed to use questions with a moderate level of difficulty. Suharsimi Arikunto stated that good questions are questions that are not too easy or too difficult. Questions that are too easy will not stimulate students to think or enhance their efforts to solve problems in each item. On the other hand, questions that are too difficult will cause students to become discouraged and have no enthusiasm to try again because they are out of reach.

d. Discrimination index

Based on the formula used to calculate the discrimination index, there are several criteria, including: if the discrimination index is 0.70–1.00, then it is included in the excellent category; 0.40–0.69 is good, 0.20–0.39 is satisfactory, < 0.20 is poor, and < 0.00 (-) is the worst category.

Table 3. The result of an English Teacher-Made Test Analysis Based on Discrimination Index

No	Category	Item	Total	Percentage
1.	Excellent (0.70 – 1.00)	3 and 9	2	20%
2.	Good (0.40 – 0.69)	1, 2, 5, 6, 7 and 10	6	60%
3.	Satisfactory (0.20 – 0.39)	4	1	10%
4.	Worst (< 0.00 / -)	8	1	10%

Based on the results above, it can be seen that the final exam for English subject Class XI Multimedia 1 UPT SMKN 3 Parepare for the academic year 2022/2023 consists of 10% or 1 item in the worst category, 1 item in the category of satisfactory,

and good categories as much as 6 or 60% of items, and items that fall into the category of very good as much as 20% or 2 items. The good category is more dominant than the other categories. The results of this study are different from the research conducted by Maya Marsevani with the results of the excellent category being more dominant with a score of 80%.

There are criteria to determine whether the distractor is working or not. These criteria are: the effectiveness of the distractors is said to be very good if the four distractors work, the effectiveness of the distractor is said to be good if there are three distractions that function, the effectiveness of the distractor is said to be quite good if there are two distractions that function, the effectiveness of the distractor is said to be poor if there is one distractor that functions, the effectiveness of the distractor is said to be bad if all four (all) distractors don't work.

Table 4. The Result of an English Teacher-Made Tests Analysis Based on Distractors

No	Category	Item	Total	Percentage
1.	Very good	1, 2, 3, 5, 6, 8 and 10	7	70%
2.	Good	4 and 9	2	20%
3.	Quite good	7	1	10%

Based on all the explanations above, it can be concluded that the final test for English Class XI Multimedia 1 UPT SMKN 3 Parepare for the 2022–2023 academic year, based on the effectiveness of the distractor, are included in the category of very good questions, namely 70%. So it needs to be maintained in the very good category for the next test. This is related to the statement according to Sudijono which states that distractor that has been obtained function very well can be used again on future tests, while the distractors which are not functioning properly should be repaired or replaced with another distractor.

CONCLUSION

Valid category is 8 items and invalid are 2 items. Included in the test that is unreliable. Difficult category is 1 item, moderate category is 8 items, and easy category is 1 item. Worst discrimination index category is 1 item, satisfactory category is 1 item, good category is 6 items and excellent category is 2 items. Based on the result showed that 85% of multiple choice of summative test fulfilling criteria as good tests items. Distractors function very good is 7 items, good category are 2 items, and quite good category is 10 items. It can be concluded that the final multiple-choice test for class XI Multimedia 1 at UPT SMKN 3 Parepare has very good quality of item validity,

unreliable test, good discrimination index, good quality level of difficulty and very good quality distractor.

The suggestions can be filed as follows: for teachers, questions that are less qualified can be corrected according to the indicators that cause failure so that they become quality questions. Questions that are not qualified should be discarded and not used again. For schools, it is necessary to conduct training about test quality analysis for teachers so that their skills in making test questions can be further improved.

REFERENCES

Al-Qur'an Al-Karim

Adom, Dickson, Jephtar Adu Mensah, and Dennis Atsu Dake. "Test , Measurement , and Evaluation : Understanding and Use of the Concepts in Education" 9, no. 1 (2020): 109–19. <https://doi.org/10.11591/ijere.v9i1.20457>.

Asrul, Rusydi Ananda, and Rosnita. *Evaluasi Pembelajaran*. Medan: Citapustaka Media, 2014.

B, Nurgiyantoro. *Penilaian Pembelajaran Bahasa*. Yogyakarta: BPFE_Yogyakarta, 2010.

Bachman, Lyle. "Statistical Analyses for Language Assessment (Cambridge Language Assessment) (Lyle F. Bachman) (z-Lib.Org).Pdf," 2004.

Brown, H.Douglas. *Language Assessment Principle and Classroom Practices*. New York: Pearson Education Inc, 2004.

Burud, Ismail, Kavitha Nagandla, and Puneet Agarwal. "Impact of Distractors in Item Analysis of Multiple Choice Questions" 7, no. 4 (2019): 1136–39.

Dwipayani, Anak Agung Sri. "Analisis Validitas Dan Reliabilitas Butir Soal Ulangan Akhir Semester Bidang Studi Bahasa Indonesia Kelas X.D SMAN 1 Terhadap Pencapaian Kompetensi," 2013, 1–18.

Effendi, Taufiq, and Ilza Mayuni. "Examining Teacher-Made English Test in a Language School" 2, no. 2 (2022): 67–76.

Hakim, Lukmanul, and Irhamsyah. "The Analysis Teacher-Made Test for Senior High School at State Senior High School 1 Kutacane" 21, no. 1 (2020): 10–20.

Hartati, Neti, and Hendro Pratama Supra Yogi. "Item Analysis for a Better Quality Test," 2003, 12.

Haryanto, Kurniawan. "Analisis Butir Soal Pilihan Ganda Ulangan Akhir Semester Genap." Universitas Sanata Dharma, 2016.

- Hingorjo, Mozaffer Rahim, and Farhan Jaleel. "Analysis of One-Best MCQs: The Difficulty Index, Discrimination Index and Distractor Efficiency." *2012* 62, no. 2 (n.d.): 142–47.
- Ismail, Fajri, Mardiah Astuti, and Hani Atus Sholikhah. *Evaluasi Pembelajaran Berbasis Riset*. Palembang: Karya Sukses Mandiri, 2020.
- Kalsum, K., Sardi, A., & Andini, N. (2023). THE QUALITY OF SUMMATIVE TEST MADE BY EFL TEACHER. *Inspiring: English Education Journal*, 6(1), 46-57.
- Kusumawati, Mutiara, and Samsul Hadi. "An Analysis of Multiple Choice Questions (MCQs): Item and Test Statistics from Mathematics Assessments in Senior High School" 4, no. 1 (2018): 70–78.
- Magdalena, Ina, Septy Nurul Fauziah, Siti Nur Faziah, and Fika Sulaehatun Nupus. "Kesulitan Dan Daya Beda Butir Soal Ujian Akhir Semester Tema 7 Kelas III SDN Karet 1 Sepatan" 3 (2021): 198–214.
- Marsevani, Maya. "Item Analysis of Multiple-Choice Questions: An Assessment of Young Learners" 10, no. 2 (2022): 401–8.
- Paramita, Ratna Wijayanti Daniar, Noviansyah Rizal, and Riza Bahtiar Sulistyan. *Metode Penelitian Kuantitatif*. Widya Gama Press, 2021.
- Putri, Yohana Fransisca Dewi Risna. "An Analysis of Teacher-Made English Final Second Semester Test for the Year Eleven Students Of SMAN 1 Ambarawa In The Academic Year Of 2008/2009 Based on the Representativeness of Content Standard." Semarang State University, 2009.
- Qodir, Abdul. *Evaluasi Dan Penilaian Pembelajaran*. Edited by Ngalimun. Yogyakarta. K-Media, 2017.
- Rahmadi. *Pengantar Metodologi Penelitian*. Edited by Syahrani. Banjarmasin: Antasari Press, 2011.
- Rahman, Arief Aulia, and Cut Eva Nasryah. *Evaluasi Pembelajaran*. Uwais Inspirasi Indonesia, 2019.
- Riinawati. *Pengantar Evaluasi Pendidikan*. Banjarmasin: Thema Publishing, 2021.
- Ropii, Muhammad, and Muh Fahrurrozi. *Evaluasi Hasil Belajar*. Edited by Syukrul Hamdi. Lombok: Universitas Hamzanwadi Press, 2017.
- Sardi, A., Haryanto, A., & Weda, S. (2017). The Distinct types of diction used by the efl teachers in the classroom interaction. *International Journal Of Science and Research (IJSR)*, 6(3), 1061-1066.
- Setiawan, Nugraha. "Teknik Sampling," 2005, 25–28.
- Sudijono, Anas. *Pengantar Evaluasi Pendidikan*. Jakarta: PT Raja Grafindo Persada,

2012.

Sudjana, Nana. *Penilaian Hasil Proses Belajar Mengajar*. Bandung: Remaja Rosdakarya, 2014.

Sugiyono. *Metode Penelitian Pendidikan Pendekatan Kuantitatif Dan R&D*. Bandung: Alfa Beta, 2010.

Suharsimi, Arikunto. *Dasar-Dasar Evaluasi Pendidikan*. 2nd ed. Jakarta: Bumi Aksara, 2013.

Sujarwadi, Sri. "Validitas Dan Reliabilitas Instrumen Penelitian," 2011, 1–23.

Wahyuningsih, ErwinTri. "Analisis Butir Soal Tes Objektif Buatan Guru Ulangan Semester Ganjil Mata Pelajaran Ekonomi Kelas." Universitas Negeri Yogyakarta, 2015.

Widiyanto, Joko. *Evaluasi Pembelajaran*. Madiun: Unipma Press, 2018.

Yani, Abet, Ali Fikri Asri, and Ahmad Burhan. "Analisis Tingkat Kesukaran, Daya Pembeda Dan Fungsi Distraktor Soal Ujian Semester Ganjil Mata Pelajaran Produktif Di SMK Negeri 1 Indralaya Utara," 2013, 98–115.

Zainal, Arifin. *Evaluasi Pembelajaran*. Bandung: PT Remaja Rosdakarya, 2014.

Zubair, Muhammad Kamal et. all. *Penulisan Karya Ilmiah Berbasis Teknologi Informasi*. Parepare: IAIN Parepare Nusantara Press, 2020