Inspiring: English Education Journal
Vol. 6 No. 1 (March 2023) pp. 46-57
ISSN 2721-2130

THE QUALITY OF SUMMATIVE TEST MADE BY EFL TEACHER

Kalsum[1], Ahmed Sardi[2], Nur Andini[3]
Institut Agama Islam Negeri Parepare[1,3], STKIP Darud Da'wah Wal Irsyad Pinrang[2]
kalsum@iainpare.ac.id[1], sardihere@gmail.com[2], nurandini@gmail.com[3]

Article History
Received:
January 18, 2023
Revised:
February 21, 2023
Accepted:
February 22, 2023
Published:
March 3, 2023

Abstract

This study aimed to determine the reliability, validity, and degree of difficulty of a summative test created by an EFL teacher for the eighth grade of SMPN 3 Watang Pulu Sidenreng Rappang. Twenty pupils and one summative test result from the teacher served as the study's samples. The investigation was carried out using a quantitative approach. The summative English test was used as the source of the data collection technique. Validity, reliability, and difficulty level analyses were used to analyze the data. The validity result was on an rtable 0,344 with a 5% significance level. The summative English test has 14 valid items with a percentage of 70%, and 6 invalid items with a percentage of 30%. The category reliability label indicated an index of 0,82 that was higher than 0,70. In terms of level of difficulty, the total number of simple category things is 2 (10%), the total number of medium category items is 12 (60%), and the total number of difficult category items is 6 (30%). According to the aforementioned finding, the medium category of the English summative exam items has an excellent category score of more than 50%. The study's findings were helpful to the teacher and the students since they provided accurate information regarding the caliber of the summative test that the EFL teacher had developed.

Keyword: *EFL Teacher, Summative Test Quality, Evaluation*

Introduction

The teacher commonly offers the students numerous questions in the form of a test to assess the students' understanding of the information that has been presented. Achievement tests are exams that teachers might administer at the end of each chapter of the course material or at the end of the semester. An achievement exam is a methodical way to gauge how much a pupil has learnt. There are two different types of achievement tests: summative and formative (Mahshanian et al., 2019). A summative test is a method of evaluation that yields grades or other numerical results that are then

Inspiring: English Education Journal
Vol. 6 No. 1 (March 2023) pp. 46-57
ISSN 2721-2130

used to assess a student's performance. If such experiential learning unit or all subject material has been finished, this test will be given. At the conclusion of a course or program, the classification of prizes is decided using summative assessments.

Regarding this research, the writer chose summative test as the kind of test which administered at the end of a unit or term, semester, or a year of study in order to measure what has been achieved both individual and by groups (Sardi & Mujahidah, 2020). The exam may take the shape of an essay examination in which students are required to express their responses in the form of sentences. Additionally, multiple-choice tests can be administered by teachers to quickly assess pupils' performance. The methods and principles that must be followed in creating a good test must be known by the teacher who creates it.

Within conducting analysis on a test, it may determine the test's quality and whether it is suitable for usage or not (Nurchalis et al., 2021). The test should be redesigned and rearranged if it doesn't meet the criteria for a good test. When teachers don't evaluate the test they used, a problem occurs. Without taking into account the rules and procedures for creating a decent test, the teacher simply created one.

The type of test used in this study, a summative test, is one that is given at the conclusion of a unit, term, semester, or academic year in order to assess student progress on an individual and group level. There are a few reasons why the SMPN 3 Parepare test for the eighth grade was selected. First, it's crucial for the teacher to create an effective test. The test should be made into correct and appropriate word choices since it is stated that proper vocabulary or diction is concerned with choice of word used in conveying a thought, how forms groups of proper words or use the proper expressions and the proper style in each situation. Diction in speaking is being a difficulty for English teachers in the classroom interaction (Sardi et al., 2017). If a test satisfies a number of criteria, it can be called to be a good test. The test should be redesigned and rearranged if it doesn't meet the criteria for a good test. As a result, we must evaluate test quality (Sardi, JN, et al., 2022). Secondly, based on the researcher's conversation with the eighth-grade English teacher at SMPN 3 Parepare. The researcher discovered a flaw in the way tests were

Inspiring: English Education Journal
Vol. 6 No. 1 (March 2023) pp. 46-57
ISSN 2721-2130

always administered without previously being thoroughly examined. Third, since creating quality summative test items takes more effort and time than creating good formative test items. The effectiveness of the pupils in understanding the subject given must be assessed on a summative exam. In order to determine whether or not the pupils are making good progress, the English teacher will conduct an evaluation (Sanjata et al., 2022).

In light of the foregoing context, the researcher came up with the following research purposes: 1) to explore how valid is the summative test created by the English teacher at the eighth grade level of SMPN 3 Parepare and 2) to illustrate how reliable is the summative test created by the English teacher at the eighth grade level of SMPN 3 Parepare. Afterwards, the third is to investigate how challenging is the summative exam created by the English teacher for the eighth grade of SMPN 3 Parepare.

Method

The descriptive quantitative method was used in this study. With the aid of this methodology, the researcher was more equipped to find the answer to the study question. The phenomena was analyzed using this technique by detecting numbers and graphics in the research data. Based on the foregoing description, this study determined the efficacy of the summative test created by an EFL teacher (Khasanah, 2016).

The sample of this research will be the entire summative test designed by English teacher from Eight Grade of SMPN 3 Parepare, the researcher took nine documents from the teacher for the sample of the research. The researcher required some instruments, the kind of instrument was documents, there are some objects reconsidered in obtaining information and one of them is paper or document (Arikunto, 2013) within this research, some documents would be collected and analyzed. They are the question and answer sheet, answer key, and exam paper.

The researcher separated the data analysis technique according to the problem description in order to provide a clear explanation: Firstly, the researcher analyzed the summative test which taken from the teacher, the validity test was tested by certain

Inspiring: English Education Journal
Vol. 6 No. 1 (March 2023) pp. 46-57
ISSN 2721-2130

formulas which used for SPSS application. It states that if the result of *r* in a test item is higher than table of Product Moment (Gay, 2006). It denotes that the thing is regarded as legitimate. Validity ranges from 0.80 to 1.00 for outstanding status and 0.60 to 0.80 for good status; 0.40-0.60 for satisfactory status; 0.20-0.40 for poor and 0.00-0.20 for a very poor status. Secondly, it states that if the result of r in atest item is higher than table of Product Moment (Tuckman, 1975). The reliability test was tested by certain formulas which used for SPSS application. It indicates that the product is regarded as trustworthy. For a very low status, dependability is 0.00 r11 0, 20; for a low status, it is 0.20 r11 0, 40; for a medium status, it is 0.40 r11 60; for a high status, it is 0.60 r11 0,70; and for a very high status, it is 0.70 r11 1.

The third is that using the following criteria, it is possible to distinguish between questions that are of good quality, good enough, and not good by analyzing the difficulty level, discrimination power, and distractor efficiency of the tests that have been obtained: a). If a question meets all three requirements—level of difficulty, power of differentiation, and deception—it is considered to be of high quality; b). If the question only meets two of the three requirements, it is still of pretty good quality; and c). If a question does not satisfy two or all of the requirements, it is of poor quality (Arikunto, 2013).

Results

There are twenty multiple-choice questions in all, and twenty learners were tested. There are two parts to the analysis: validity and reliability. The researcher has the following knowledge of the English summative test items' quality based on the data analysis:

A. The Validity of Summative Test toward English Learning

The point correlation formula can be used to test the validity of multiple-choice questions. The test is taken by 20 eighth-grade pupils from SMPN 3 Parepare. The calculations' outcome was reviewed in order to rtable at a 5% level of significance. The test has a maximum of 20 participants, thus the rtable standard is set at 0.34; if the rpbi

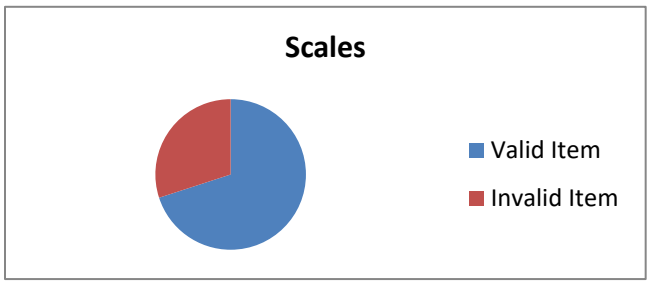
Inspiring: English Education Journal
Vol. 6 No. 1 (March 2023) pp. 46-57
ISSN 2721-2130

is higher than rtable, the item test was valid; if the result is lower than rtable, the item tested was still not valid or invalid.

As noted in the summative test items, there are 14 legitimate questions with a percentage of (70%) and 6 invalid questions with a percentage of (30%). Following is a spread of test items based on the established validity standards:

No	Item	Item Number	Total	Percentage
1	<0,344	1,3,5,7,9,10	6	30%
2	≥ 0,344	2,4,6,8,11,12,13,14,15,16,17,18,19, 20	14	70%

In accordance with the result data above, the researcher showed the diagram below:



The result data which showed on the diagram was the data from the test summative using in the final test in the school, the researcher conducted the data which was distributed from the teacher.

B. The Reliability of Summative Test toward English Learning

The capacity to be trusted depends on how consistent something is. Using KR-20, the question's reliability was assessed. When the interpretation reliability coefficient (r11) is less than or equal to 0,70, the item being tested has a low reliability or unreliability. Conversely, when r11 is greater than or equal to 0,70, the item being tested has a high dependability or reliability. SMPN 3 Parepare eighth grade students' summative test items had a dependability index of 0,422, according to the overall calculation. These findings

Inspiring: English Education Journal
Vol. 6 No. 1 (March 2023) pp. 46-57
ISSN 2721-2130

suggest that the English summative test items given to eighth-grade students in SMPN 3 Parepare who were included in the category are not trustworthy because r11 is less than 0,70, which qualifies the test items as being unreliable.

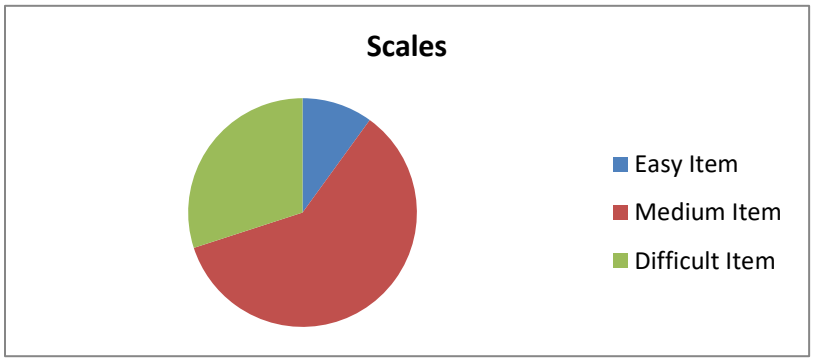
C. The Difficulty Level of Summative Test toward English Learning

This study also identifies the degree of difficulty and the persistence of this investigation in carrying out the quality test. Characterization is used to explain the difficulty calculation results, which are 0,71-1 for questions classified as easy, 0,31-0,70 for questions in the medium category, and 0,00-0,30 for questions in the tough category.

Based on the study, it was determined that there were 12 multiple-choice questions in the easy category with a percentage of 60%, 2 in the medium category with a percentage of 10%, and 6 in the difficult category with a percentage of 30%. According to degree of difficulty, questions are distributed as follows:

No	Difficulty Level	Item Number	Total	Percentage
1	Easy (0,70 – 1)	19,20	2	10%
2	Medium (0,30 – 0,69)	1,2,4,6,8,11,12,13,14,15,16,17	12	60%
3	Difficult (0.00 – 0,29)	3,5,7,9,10,18	6	30%

The researcher displayed the diagram below based on the results data above:



Based on the aforementioned findings, the researcher concludes that the tough item was the one that stood out the most following data analysis, which revealed that medium item data (60%) was the dominating and highest item, followed by difficult item data for 30% and easy item data for 10%.

Inspiring: English Education Journal
Vol. 6 No. 1 (March 2023) pp. 46-57
ISSN 2721-2130

Discussion

This phrase refers to the data discussion that follows the results mentioned above. The discussion is rooted in the findings and the researcher's justification following the analysis of the data in the findings. Some indices of validity, reliability, and level of difficulty can be used to determine the quality of the English summative test items for SMPN 3 Parepare pupils in the eighth grade. The following is a discussion of each of the indicators:

A. The Validity of Summative Test toward English Learning

In this study, the point bacterial correlation formula *(rpbi)* was used to determine the items' validity. The calculations' outcome was reviewed in order to *rtable* at a 5% level of significance. The test has a maximum of 20 participants, thus the *rtable* standard is set at 0,344. If the *rpbi* is higher than the rtable, the item test was legitimate, and if the result is lower than the *rtable*, the item test was not valid or invalid. According to the findings, 14 items were certified genuine with a percentage of (70%) and 6 things were declared invalid with a percentage of (30%). The incorrect items should be corrected, and the right ones can just be utilized again and added to the question bank.

In light of the aforementioned rationale, it can be said that the English summative test items given to SMPN 3 Parepare eighth grade pupils had acceptable value in terms of their validity since there were more valid items than there were total questions. "The item validity of a test is the measuring accuracy that an item has in measuring what should be measured through these items," (Salwa, 2012).

Following up on the findings of the validity study of the question items can be done as follows:

1. A valid item may be added to the question bank for use in the exam the following semester.
2. Reject invalid items and replace them with questions that follow the material indicators.

B. The Reliability of Summative Test toward English Learning

The accuracy of test results across all aspects of the test called reliability. Any

Inspiring: English Education Journal
Vol. 6 No. 1 (March 2023) pp. 46-57
ISSN 2721-2130

good test must have reliability; in order for a test to be valid at all, it must first be trustworthy as a measuring tool. The test ought to yield comparable outcomes if it is administered twice the same student or a group of matched students (Ahmad et al., 2022). The degree to which an assessment of a phenomenon yields a steady and consistent result is what reliability refers to. Repeatability is another aspect of reliability.

A test is deemed dependable if it consistently produces the same result when repeated measurements are conducted. According to the previous explanation, reliability premeasuring will result in constant outcomes no matter how many times it is done. The KR-20 formula is used to determine the reliability of the English summative test items given to eighth-grade students in SMPN 3 Parepare. Excel is used to perform manual calculations. When the interpretation reliability coefficient *(r11)* is less than or equal to 0,70, the item being tested has a poor reliability or unreliability. Conversely, when *r*11 is greater than or equal to 0,70, the item being tested has a high dependability or reliability. The teacher must thus devise a suitable exam to evaluate pupils' participation in classroom interactions (Ermawati et al., 2021).

According to the computations, *r*11 = 0.422 was the outcome. These calculations' findings show that English summative test items for SMPN 3 Parepare eighth grade pupils have low reliability because *r*11 0.70. If a test consistently yields the same findings when administered to the same group at several times and for various subjects, it is said to be trustworthy. It is claimed that a tool's dependability demonstrates how consistently it evaluates the subject it is judging. According to the justification provided, the English summative test item for SMPN 3 Parepare for the eighth grade pupils is a low quality question in terms of reliability.

C. The Difficulty Level of Summative Test toward English Learning

The proportion of students who responded to the question properly to the total number of test-takers indicates the item's difficulty level. Items can be considered good if they fall into the median category—that is, if they are neither too difficult nor too easy (Sardi, Surahmat, et al., 2022). Very simple problems do not encourage students to

Inspiring: English Education Journal
Vol. 6 No. 1 (March 2023) pp. 46-57
ISSN 2721-2130

increase their efforts in problem solving. On the contrary, issues that are very challenging can discourage students and make them lose interest in trying again because they appear unachievable. The analysis's findings indicate that there were 12 multiple-choice questions in the medium-difficulty category (with a percentage of 60%), 2 in the easy-difficulty category (with a percentage of 10%), and 6 in the difficult category (with a percentage of 30%).

A question is deemed to have a good difficulty level if it falls between the range of 0.31-0.70. A decent question has a medium difficulty level, which is between 0.3 and 0.70; the interval of difficulty level stated by (Arikunto, 2013). judged that the English summative test items for SMPN 3 Parepare eighth grade students are well-written questions based on their level of difficulty. This is demonstrated by the fact that there are 12 questions, or 60% of the total, that are of a medium difficulty. The findings indicated that the majority of the questions fell into the category of medium difficulty. The medium category's items can be added to a question bank for future use as an assessment tool. Items that scored as easy or tough can be reviewed to determine the reason why they scored that way, then updated and tested again on the subsequent test. Medium-level issues must be kept up with.

Conclusion

Relying on the findings of the data analysis in the preceding chapter, the conclusion was made on the effectiveness of the summative test administered to SMPN 3 Parepare pupils in the eighth grade, with the following explanation:

1. The calculations' output was reviewed to *rtable* at a 5% level of significance. The test has a maximum of 20 participants, thus the *rtable* standard is set at 0.34; if the *rpbi* is higher than *rtable*, the item test was valid; if the result is lower than *rtable*, and the item test was not valid or invalid. Based on the validity result, there is an *rtable* of 0,344 with a 5% significant level. The overall number of legitimate items on the summative English test are 14, totaling 70%, whereas the total number of invalid items is 6, totaling 30%.

Inspiring: English Education Journal
Vol. 6 No. 1 (March 2023) pp. 46-57
ISSN 2721-2130

2. According to dependability, the English summative test questions given to SMPN 3Parepare eighth grade pupils have a reliability value of 0.82. Because r11 is greater than 0,70, this reliability index demonstrates that the English summative test's items are reliable.
3. In light of the findings of the level of difficulty, the total number of easy category items is 2 (10%), the total number of medium category items is 12 (60%), and the total number of tough category items is 6 (30%). According to the aforementioned finding, the medium category of the English summative exam items has an excellent category score of more than 50%.

Suggestion

The researcher offers two proposals towards the teacher and researchers that are based on the research and conversation. The following is a description of these points:

1. To make the tests more varied, it is advised that English teachers administer tests more frequently.
2. It is advised that English teachers encourage their students to develop opinions about the subject. Consequently, the pupils will be able to construct such concepts as a test result.
3. It is advised that English teachers pay more attention to the rules governing the administration of summative exams.

References

Ahmad, A. K., Ishak, & Afdalia. (2022). Peningkatan Hasil Belajar Matematika melalui Model Pembelajaran Kooperatif Tipe Two Stay Two Stray. Al-Irsyad Journal of Mathematics Education, 1(2), 79–87. https://doi.org/10.58917/ijme.v1i2.23

Arikunto, S. (2013). *Dasar-Dasar Evaluasi Pendidikan*. Bumi Aksara (PT Bumi Aksara).

Ermawati, E., Nurchalis, N. F., & Sardi, A. (2021). Online EFL Teaching and Learning:

Inspiring: English Education Journal
Vol. 6 No. 1 (March 2023) pp. 46-57
ISSN 2721-2130

Different skills, Different Challenges. IDEAS: Journal on English Language Teaching and Learning. Linguistics and Literature, 9(1), 373-382.

Gay, R. . (2006). *Educational Research; Competencies for Analysis & Application* (Eight Edit). Barkley Lehigh Press.

Humaeroah, H., Sardi, A., & Ermawati, E. (2023). Teacher Perspective: Managing Students' Behavior Problem in Teaching English at Primary School. IDEAS: Journal on English Language Teaching and Learning, Linguistics and Literature, 10(2), 2113-2121.

Kalsum, K., Rauf, F. A., & Sardi, A. (2023). Implementation of Reading-Log to Increase Students' Interest on Literacy at Islamic Boarding School. IDEAS: Journal on English Language Teaching and Learning, Linguistics and Literature, 10(2), 1887-1898.

Khasanah, N. (2016). *The Analysis of Students' Perceptions for The Impact of Formative and Summative Test for The fourth Semester Students of English Education Department in IAIN Salatiga in The Academic Year of 2015/2016* [IAIN Salatiga]. http://e-repository.perpus.iainsalatiga.ac.id/1420/

Mahshanian, A., Shoghi, R., & Bahrami, M. (2019). Investigating the differential effects of formative and summative assessment on efl learners' end-of-term achievement. *Journal of Language Teaching and Research*, *10*(5), 1055–1066. https://doi.org/10.17507/jltr.1005.19

Nurchalis, N. F., Ermawati, E., Sardi, A., & Nursabra, N. (2021). Language Laboratory to Overcome the Barrier of Classroom English Learning: Does it Exist and Is it Used in Islamic Schools of Majene?. Elsya: Journal of English Language Studies, 3(3), 183-194.

Salwa, A. (2012). *The validity, Realibility, Level of Difficulty, and Appropriateness of Curriculum of the English TEST* [Diponegoro University]. https://core.ac.uk/download/pdf/19755101.pdf

Sanjata, A. R. M. P., Sardi, A., & Muchtar, J. (2022). Peningkatan Hasil Belajar Melalui Model Pembelajaran Tutor Sebaya Setting Kooperatif. Al-Irsyad: Journal of Education Science, 1(2), 117-124.

Inspiring: English Education Journal
Vol. 6 No. 1 (March 2023) pp. 46-57
ISSN 2721-2130

Sardi, A. (2022). The Building up of Students' Vocabulary Mastery through Knowing by Heart Strategy. LETS: Journal of Linguistics and English Teaching Studies, 4(1), 62-72.

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.

Sardi, A., JN, M. F., Walid, A., & Ahmad, A. K. (2022). An Analysis Of Difficulties In Online English Learning Experienced By The Efl Teacher. Inspiring: English Education Journal, 5(2), 144-154.

Sardi, A., & Mujahidah, M. (2020). Could I Be Illogical?(Cibi Guide) For Non-Native Speaker.

Sardi, A., Surahmat, Z., & Nur, S. (2022). The Washback of Intensive TOEFL Training Program (ITTP) on Student's Learning Motivation. ELS Journal on Interdisciplinary Studies in Humanities, 5(4), 593-597.

Tuckman, B. W. (1975). *Measuring Educational Outcomes Fundamentals of Testing*. Harcourt Brace Javanovich Inc.

Surahmat, Z., Sardi, A., & JN, M. F. (2023). A CHAPTER REVIEW: SELECTING LANGUAGE FOR MATERIALS WRITING: (The Routledge Handbook of Materials Development for Language Teaching-Routledge). Al-Irsyad: Journal of Education Science, 2(1), 15–24. https://doi.org/10.58917/aijes.v2i1.39