

AN ANALYSIS OF THE VALIDITY AND RELIABILITY OF ENGLISH EXAMINATIONS: A STUDY AT A SENIOR HIGH SCHOOL IN KEFAMENANU

Dionesia Nurani Arcelibur¹, Thresia Trivict Semiun², Anselmus Sahan³
^{1,2,3}*Universitas Timor*

Abstract: This study analyzed the validity and reliability of the English language school exam at a senior high school in Kefamenanu during the 2023/2024 academic year. A mixed-method approach was employed, with content validity assessed by matching test items to the blueprint and reliability measured using the Kuder-Richardson 20 (KR-20) formula. The results indicated that the test had a validity rate of 55%, which was classified as moderate due to some items not aligning with the specified indicators. Meanwhile, the reliability coefficient was 0.89, demonstrating a very high correlation and strong consistency of test results. It is recommended that the school improve the exam blueprint and ensure that all test items accurately reflect the taught material.

Keywords: English school examination, reliability, validity

INTRODUCTION

In education, evaluation plays a crucial role in measuring the teaching and learning process. According to Rahmah et al. (2020), evaluation is an assessment process conducted by teachers to measure student learning outcomes. It serves as a vital component for determining students' understanding of the material taught over a semester. One key purpose of evaluation is to measure students' learning progress and the extent to which they have achieved predetermined learning objectives.

Various methods can be used for evaluation, including the administration of tests. Among these, multiple-choice tests are one of the most common formats used to assess student achievement (Kusumawati & Hadi, 2018). Teachers typically administer multiple-choice tests during midterm exams, final semester exams, school examinations, and national assessments. For a multiple-choice test to be effective, it must meet the criteria of validity and reliability. Harahap and Novita (2020) state that validity involves expert judgment to determine whether the test instruments are appropriate for their intended purpose, while reliability refers to the consistency of test results over repeated administrations. Valid and reliable test items significantly influence the accuracy of measuring student learning outcomes. Specifically, analyzing content validity is crucial because it ensures that the test items fully represent the curriculum objectives and the material that has been taught. A test with low content validity may fail to measure the intended competencies, leading to unfair or inaccurate assessments of students' abilities. Similarly, analyzing the reliability of a test is essential to determine whether the test produces consistent results, providing trustworthiness to the evaluation process. Without high validity and reliability, the results of an examination cannot be considered a true reflection of a student's knowledge and skills.

Despite the importance of these aspects, at various educational levels (junior high school, senior high school, and college), multiple-choice tests often fail to meet the criteria of good item construction fully. Poorly designed multiple-choice questions can affect both validity and reliability, compromising the evaluation's effectiveness. Teachers must, therefore, possess the

skills and expertise necessary to develop appropriate multiple-choice items. Additionally, students' guessing strategies can undermine test validity, as they might select correct answers without a full understanding of the material (Foley, 2016). Hence, multiple-choice questions must be carefully designed to measure not only basic knowledge but also higher-order thinking skills (Marsevani, 2022). Based on these considerations, this study analyzes the quality, in terms of validity and reliability, of a multiple-choice English examination at one of the senior high schools in Kefamenanu. The research was conducted in the school where the researcher completed her teaching practice for six months. The researcher chose the academic year 2023/2024 to ensure that the data analyzed is current and relevant to the present educational context.

Several previous studies are relevant to this research. First, Furwana (2019) investigated the validity and reliability of English summative tests at Vocational High School 2 Palopo, finding good content validity (80%) and relatively high reliability (0.6132). Second, Jayanti et al. (2019) analyzed the validity and reliability of the English National Final Examination for Junior High Schools in South Jakarta, reporting content and construct validity of 92% and a reliability coefficient of 0.784. Third, Winarti et al. (2021) evaluated the validity, reliability, and authenticity of English achievement tests for twelfth-grade students at SMAN 4 Tebo, Jambi, finding moderate validity and reliability. Last, Sugianto (2016) analyzed the English National Final Examination for Junior High Schools, reporting 100% content and construct validity and a high-reliability coefficient of 0.89. These previous studies provide a strong foundation for the present research and highlight the critical need to ensure that multiple-choice examinations meet the standards of validity and reliability to promote fair and accurate assessment of student learning.

The findings of this study are expected to provide valuable insights for English teachers in designing valid and reliable school examinations. By highlighting weaknesses and strengths in the existing tests, the study aims to encourage improvements in test construction practices. Furthermore, the results may serve as a useful reference for school administrators and curriculum developers in evaluating and enhancing the quality of assessment tools. Ultimately, this research contributes to ensuring that school examinations more accurately reflect students' abilities and support better educational outcomes.

METHOD

This research used qualitative and quantitative methods. Qualitative was used to analyze the content validity of the test school examination of English subjects. According to Sugiyono (2017). The documents collected in this study were the school examination test of English subject for the academic year 2023/2024, the blueprint of the test, and students' scores in the school examination of English subject for the academic year 2023/2024. English school examination test of English subject for the academic year 2023/2024 and blueprint of the test used to analyze the content validity of the test. Meanwhile, students' scores were used to analyze the reliability of the school examination test for English subjects. To determine the content validity of the school examinations of English subjects, the researcher matched the items in the school examinations of English subjects with the blueprint. It is to see whether each item of the test measures the same indicator. After that, the researcher asked the validator to see the analysis conducted by the researcher. For analyzing the content validity, the researcher followed the data analysis conducted by Semiun and Luruk (2020). Meanwhile, the researcher applied the KR 20

formula to analyze reliability. Then, the results calculation was categorized into scales of interpretation of reliability (Jayanti et al., 2019).

RESULTS AND DISCUSSION

In this study, an analysis of the validity and reliability of school exam questions was conducted to ensure the quality of the evaluation instrument used. Validity aims to measure the extent to which exam questions are able to measure the expected competencies. At the same time, reliability refers to the consistency of measurement results when tested on similar groups of respondents. The validity analysis process is carried out by matching test items with the blueprint. If one question item is declared valid, then the presentation is 2.5%, but if one question item is invalid, then the presentation is 0%. Meanwhile, reliability uses the Kuder-Richardson 20 (KR-20) formula. The results of this analysis are presented in the following section. In analyzing validity, researchers match each multiple-choice question item with the blueprint that has been provided. The following is a table of content validity analysis of multiple choice questions on the school examination.

Tabel 1 Analysis of content validity

Basic Competencies	Indicators	Students Ability	Percentage
Distinguish the social function, text structure, and linguistic elements of several and written news texts by giving and requesting information related to simple news from newspapers/radio/TV, according to the context of use.	Identifying the parts of the report text structure and observing how to use them. (1,2,3,4,5,8,9,10)	Reading	20%
	Identifying several information report texts related to subjects (6,7)	Reading	5%
Applying social functions, text structures, and linguistic elements of oral and written transactional interaction texts that involve the act of giving and asking for information relate to assumptions followed by commands/suggestions, according to the context of their use. (Pay attention to the linguistic elements if with imperative, can, should)	Identifying conditional expressions with suggestions from existing examples, seen from the content and how they are expressed. (11,13,15,17,19)	Reading	0%
	Identifying social functions/text structures/linguistic elements used in the interactions studied. (12,14,16,18,20)	Reading	0 %
Applying social functions, text structures, and linguistic elements of oral and written transactional interaction texts involving the act of giving and asking for information related to concessions, according to the context of their use. (Pay attention to the linguistic elements even though, although)	Identify the social function, text structure, and linguistic elements of the conversation. (21,22,23,24,25,26,27,29)	Reading	17,5%
	Determine the social function, text structure, and linguistic elements of the conversation. (28)	Reading	0%
Interpreting the social function and linguistic elements of song lyrics related to the lives of high school teenagers	Identifying social functions, text structures, and linguistic elements of song lyrics, contextually. (31,32,33)	Reading	7.5%
	Finding information in song lyrics.(34)	Reading	0%
	Identifying information in song lyrics.(35)	Reading	0%

Analyzing the social function and linguistic elements of song lyrics, which are being studied contextually.(36)	Reading	0%
Determining the topic of the text review.(37)	Reading	0%
Finding implied detailed information from text review. (38,39)	Reading	0%
Finding messages from song lyrics. (40)	Reading	0%

The English School Examination's validity analysis showed a validity coefficient of 55%. Based on standard validity categories, this score indicates a moderate to low level of validity, implying that several aspects of the test require revision to enhance its feasibility as a reliable measurement instrument.

One of the primary factors contributing to the low validity is the mismatch between the test items and the predetermined blueprint. Several questions did not specifically measure the indicators outlined in the test specifications, leading to a weak correlation between the items and the intended learning outcomes. This finding aligns with Astuti's (2020) analysis, which noted that invalid items often stem from discrepancies between the blueprint and the test items. For example, in this study, questions numbered 1–20 were intended to assess material on commands and suggestions but instead evaluated direct and indirect speech. Similarly, questions 21–30, which should have assessed the use of "although," instead asked about personal letters. As a result, the test instrument does not fully reflect the curriculum objectives, affecting the overall validity and potentially leading to unfair or inaccurate assessments of students' competencies.

Before analyzing reliability, researchers added up student scores (\bar{X}) divided by the number of students and produced 0.079. This 0.079 will later be used as a reduction in the standard deviation table (S2). The following table of PQ shows the results.

Table 2. Table of PQ

Question Number	Number of students who answered correctly	P	Q	PQ
1	27	0.96	0.04	0.038
2	28	1	0	0
3	28	1	0	0
4	27	0.96	0.04	0.038
5	27	0.96	0.04	0.038
6	28	1	0	0
7	27	0.96	0.04	0.038
8	28	1	0	0
9	28	1	0	0
10	27	0.96	0.04	0.038
11	28	1	0	0
12	28	1	0	0
13	27	0.96	0.04	0.038

14	26	0.92	0.08	0.078
15	28	1	0	0
16	27	0.96	0.04	0.038
17	28	1	0	0
18	28	1	0	0
19	28	1	0	0
20	28	1	0	0
21	28	1	0	0
22	28	1	0	0
23	28	1	0	0
24	27	0.96	0.04	0.038
25	27	0.96	0.04	0.038
26	28	1	0	0
27	28	1	0	0
28	28	1	0	0
29	28	1	0	0
30	28	1	0	0
31	26	0.92	0.08	0.073
32	22	0.78	0.22	0.171
33	27	0.96	0.04	0.038
34	28	1	0	0
35	28	1	0	0
36	28	1	0	0
37	20	0.71	0.28	0.198
38	27	0.96	0.04	0.038
39	27	0.96	0.04	0.038
40	28	1	0	0
Total				0.971

Based on the calculation results, the total pq value obtained is 0.97. This value indicates the level of variation in the proportion of correct and incorrect answers to the question items. After finding the pq value, here is the table of standard deviation(S^2).

Table 3. Table of S^2

Students	X	$X - \bar{x}$	$(X - \bar{x})^2$
1.	85,00	$85,00 - 0.079$	$84.921 = 7.211,58$
2.	87,50	$87,50 - 0.079$	$87.421 = 7.642,43$
3.	87,50	$87,50 - 0.079$	$87.421 = 7.642,43$
4.	87,50	$87,50 - 0.079$	$87.421 = 7.642,43$
5.	87,50	$87,50 - 0.079$	$87.421 = 7.642,43$
6.	85,00	$85,00 - 0.079$	$84.921 = 7.211,58$
7.	87,50	$87,50 - 0.079$	$87.421 = 7.642,43$

8.	90,00	90,00 – 0.079	89.921 = 8.085,78
9.	82,50	82,50 – 0.079	82.421 = 6.793,22
10.	80,00	80,00 – 0.079	79.921 = 6.387,36
11.	82,50	82,50 – 0.079	82.421 = 6.793,22
12.	87,50	87,50 – 0.079	87.421 = 7.642,43
13.	85,00	85,00 – 0.079	84.921 = 7.211,58
14.	87,50	87,50 – 0.079	87.421 = 7.642,43
15.	85,00	85,00 – 0.079	84.921 = 7.211,58
16.	87,50	87,50 – 0.079	87.421 = 7.642,43
17.	87,50	87,50 – 0.079	87.421 = 7.642,43
18.	85,00	85,00 – 0.079	84.921 = 7.211,58
19.	85,00	85,00 – 0.079	84.921 = 7.211,58
20.	85,00	85,00 – 0.079	84.921 = 7.211,58
21.	85,00	85,00 – 0.079	84.921 = 7.211,58
22.	87,50	87,50 – 0.079	87.421 = 7.642,43
23.	87,50	87,50 – 0.079	87.421 = 7.642,43
24.	77,50	77,50 – 0.079	77.421 = 5.994,01
25.	80,00	80,00 – 0.079	79.921 = 6.387,36
26.	85,00	85,00 – 0.079	84.921 = 7.211,58
27.	87,50	87,50 – 0.079	87.921 = 7.642,58
28.	87,50	87,50 – 0.079	87.921 = 7.642,58
			204.754,48

The researcher looked for the standard deviation (S^2) using the following formula:

$$S^2 = \frac{\sum (X)^2}{n} - \frac{(\sum X)^2}{n^2}$$

$$S^2 = \frac{204.754,48}{28} - \frac{(840)^2}{28^2}$$

$$S^2 = 7.312,66$$

The result of calculating the value of total variance (S^2) is 7.312,66. The researcher employed the Kuder-Richardson 20 (KR-20) formula, calculating reliability based on students' total scores and the number of correct responses. The analysis yielded a reliability coefficient of 0.89, indicating a very high correlation and excellent internal consistency. This finding suggests that the English School Examination administered at one of the senior high schools in Kefamenanu during the 2023/2024 academic year is highly reliable and capable of consistently measuring students' knowledge in English.

When compared with previous studies, this reliability result is higher than the 0.79 reliability coefficient reported by Astuti (2020) and Azizah et al. (2024). The higher coefficient in the present study indicates better internal consistency of the test items. Although Astuti's study still provided valuable contributions to test development, the difference may stem from improvements in test construction and differences in data collection methods. Nevertheless, to ensure optimal reliability, it is recommended that further studies expand the sample size and apply additional methods such as test-retest

reliability analysis to confirm the stability of the measurement results under different conditions.

The findings of this study demonstrate a lower validity but higher reliability compared to several previous studies. For instance, Jayanti et al. (2019) found a content validity of 92% in their analysis of the English National Final Examination, significantly higher than the 55% validity found in this study. This discrepancy suggests that the present instrument does not yet fully align with the intended constructs, affecting its effectiveness in evaluating student competencies. However, in terms of reliability, the current study's result of 0.89 exceeds the reliability coefficient reported by Astuti (2020), which was 0.79. This indicates that although the validity of the instrument needs improvement, the internal consistency of the test items is commendable.

CONCLUSION

Based on the validity analysis, 22 out of 40 questions were found valid, resulting in an overall validity of 55%, which indicates a medium level of validity. This suggests that while the test has a solid foundation, 18 invalid questions should be revised to better measure the intended competencies. The reliability analysis yielded a high coefficient of 0.89, indicating very high reliability. However, to ensure the test accurately reflects students' abilities, it is recommended that English teachers align questions with the basic competencies and the material taught, ensure clarity in the questions, and use various types of questions for a more comprehensive assessment. Teachers should also involve colleagues in reviewing the exam. The school should revise the question blueprint to ensure it accurately reflects the curriculum and regularly evaluate test instruments to maintain high standards of validity and reliability. These improvements will enhance the test's overall quality and provide more accurate student assessments.

REFERENCES

- Azizah, A. M., Nanda, D. P., & Huda, S. (2024). Assessing the Suitability of the Content and Construct of the End-of-Semester Exams For English Class VII MTS PUTRA-PUTRI SIMO According to the 2013 Curriculum. *MEDIA DIDAKTIKA*, 10(2), 26-29. <https://doi.org/https://doi.org/10.52166/didaktika.v10i2.4456>
- Foley, Brett P. (2016). Getting Lucky: How Guessing Threatens the Validity of Performance Classifications. *Practical Assessment, Research & Evaluation*, 21(3). Available online: <http://pareonline.net/getvn.asp?v=21&n=>
- Furwana, D. (2019). Validity and Reliability of Teacher-Made English Summative Test at Second Grade of Vocational High School 2 Palopo. *Journal of Language and Literature*, 13(2), 112–113. <http://journal.unnes.ac.id>
- Harahap, I. P. P., & Novita, D. (2020). Validitas dan Reliabilitas Instrumen Tes Diagnostik Four-Tier Multiple Choice (4TMC) pada Konsep Laju Reaksi. *Unesa Journal of Chemical Education*, 9(2), 222–227.
- Jayanti, D., Husna, N., & Hidayat, D. N. (2019). The Validity and Reliability Analysis of English National Final Examination for Junior High School. *VELES Voices of*

- English Language Education Society, 3(2), 127.
<https://doi.org/10.29408/veles.v3i2.1551>
- Kusumawati, M., & Hadi, S. (2018). An analysis of multiple choice questions (MCQs): Item and test statistics from mathematics assessments in senior high school. *REID (Research and Evaluation in Education)*, 4(1), 70–78.
<https://doi.org/10.21831/reid.v4i1.20202>
- Marsevani, M. 2022. Item Analysis of Multiple-Choice Questions: an Assessment of Young Learners. *English Review: Journal of English Education*, 10(2), 401–408.
<https://doi.org/10.25134/erjee.v10i2.6241>
- Mohamad, M. M., Sulaiman, N. L., Sern, L.C., & Salleh, K. M. (2015). Measuring the Validity and Reliability of Research Instruments. *Procedia - Social and Behavioral Sciences*, 204, 164–171.
<https://doi.org/10.1016/j.sbspro.2015.08.129>
- Rahmah, N., Yusrizal, & Syukri, M. (2020). Analysis of multiple-choice question (MCQ) of physics final examination in senior high school. *Journal of Physics: Conference Series*, 1460(1). <https://doi.org/10.1088/1742-6596/1460/1/012143>
- Semiun, T. T., & Luruk, F. D. (2020). The quality of an English summative test of a public junior high school, Kupang-NTT. *English Language Teaching Educational Journal*, 3(2), 133. <https://doi.org/10.12928/eltej.v3i2.2311>
- Sugianto, A. (2016). An analysis of English national final examination for junior high school in terms of validity and reliability. *Journal on English as a Foreign Language*, 6(1), 31. <https://doi.org/10.23971/jefl.v6i1.427>
- Sugiyono, (2017). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Bandung: CV. Alfabeta.
- Winarti, M., Faridi, A., & Rozi, F. (2021). Evaluating the Validity, Reliability and Authenticity of English Achievement Test for the Twelfth Grade Students of SMAN 4 Tebo, Jambi. *English Education Journal*, 11(1), 130–138.
<https://doi.org/10.15294/eej.v11i1.441>