

# Validity And Reliability Of The English Summative Test For Second Semester Of The Fifth Grade In Academic Year 2019/2020

Khairunnisa<sup>1\*</sup>, Anni Holila Pulungan<sup>2</sup>, Rahmad Husein<sup>3</sup>

<sup>1</sup> Basic Education Program, Universitas Negeri Medan.

<sup>2</sup> English language applied linguistics program  
Universitas Negeri Medan.

<sup>3</sup> English language applied linguistics program  
Universitas Negeri Medan Medan, North Sumatra 20221, Indonesia.

\*Corresponding author:

Email: [kkhairunnisa800@gmail.com](mailto:kkhairunnisa800@gmail.com)

---

## Abstract.

*The validity and reliability of the test is very important before the test is used to measure student achievement. This study aims to analyze the validity and reliability of the summative English test for fifth grade public elementary schools 058238 Bukit Rata in the academic year 2019/2020, and to investigate the quality level of the summative test of English in elementary schools. This research is a quantitative descriptive study. The subjects in this study were students of grade V public elementary school 058238 Bukit Rata in the academic year 2019/2020. The data analysis technique used is quantitative analysis by testing the validity, reliability, level of difficulty, distinguishing power, and tricking effectiveness. The data obtained were analyzed using the ANATES version 4 program and manual calculations using Microsoft Excel. The results showed that, the level of validity of the English test in grade 5 elementary school before the revision based on material aspects, construct aspects and language aspects obtained a percentage of 73% with a good category. After being revised, based on material aspects, construct aspects and language aspects obtained a percentage of 90% with a good category. The reliability test was carried out twice, namely the questions before expert validation and after expert validation. The results of the reliability test on the first question were the value of -0.084 in the unreliable category because the value was  $\leq 0.700$ . Furthermore, the second test of English test reliability was carried out. The results of the second test reliability test, namely the KR 20 formula with the help of the Excel program obtained a value of 0.413 in the sufficient category. The level of validity and reliability of the English test in grade 5 elementary schools. Before the revision, the number of valid questions was 4 valid questions or 16% of the total questions. Furthermore, the number of questions that can be accepted is 17 questions, but 13 questions are not valid. So it is necessary to improve the overall problem. After the revision was made, the number of valid questions was 16 valid questions or 64% of the total questions. Furthermore, the number of questions that can be used is 15 questions, because one significant question has a low difference power, so the question is not used. The validity and reliability of the English test in fifth grade public elementary school 058238 Bukit Rata should be revised because reliability is used as an indicator in believing the value of a test because it has consistency. While the test must be valid because it is used as a development and evaluation of a test. In addition, validity is also needed to determine the feasibility of items.*

**Key words:** validity, reliability.

---

## 1. INTRODUCTION

The tool used to evaluate learning is a test. The test is an important component in the teaching and learning process. The test is used as information that shows an

increase in student learning outcomes. The test is seen from its use in measuring students' abilities. Tests are conducted by educators to monitor the process, progress, and improvement of student learning on an ongoing basis [1]. Assessment as a systematic process of collecting, analyzing, and interpreting information / data to determine the extent to which students have achieved the objectives of the learning process. In addition, assessment should be viewed as an important factor that determines the success of the learning process and outcomes, not just as a means used to assess learning outcomes. [2]. Item analysis is the assessment of test questions in order to obtain a set of questions that have adequate quality [3]. States that item analysis aims to identify good, bad, and bad questions. three things related to the analysis of the questions, namely the level of difficulty, the power of discrimination, and the effectiveness of the trickster [1]. Based on this opinion, it can be concluded that the assessment of the question instruments was to determine the quality of the items so that the learning outcomes could be accurate and meaningful for students.

This research will be conducted in fifth grade public elementary school 058238 Bukit Rata. This school was chosen because it has good public schools. However, learning English in schools is only limited to local content. Based on the results of observations made, the teacher never develops worksheets or tests to improve student learning outcomes. So that the quality of learning English in schools is still low. This is indicated by only 30% of students who have good basic English skills. In addition, when students take school exams in grade six, students find it difficult to answer the English test.

States that validity refers to appropriateness, meaningfulness, and usefulness of a conclusion obtained from the interpretation of test scores.1 Validity leads to accurate interpretation of the results of using an evaluation procedure in accordance with the measurement objectives [4]. Validity is one of the characteristics that marks a good learning outcome test. To be able to determine whether the learning outcome test has validity or measureability, it can be done from two aspects, namely: the validity of the test and the validity of the item [5]. Also states that a test is valid if it measures accurately what it is intended to measure [6]. validity is the degree of accuracy between the data that occurs in the research object and the power that can be reported by the researcher [7]. It can be concluded that validity is accuracy in measuring objects that should be measured and in accordance with certain criteria. Validity is also a condition when an evaluation instrument can measure what should be measured precisely. A low validity test will produce data that is not relevant to the measurement objective.

There are two kinds of logical validity, namely content validity and construct validity. Content validity is a test-making concept that emphasizes the aspect of the extent to which the tests made can be representative of the material being taught. The validity of the contents of a learning outcome test can be obtained after analyzing, tracing or testing the content contained in the learning outcome test. Construct validity

can be interpreted as validity that is judged in terms of structure, framework, or fiction [5]. Empirical validity is the validity obtained on the basis of field observations. A learning outcome test can be said to have empirical validity if the results of the analysis carried out in field observations prove that the learning outcome test can accurately measure learning outcomes which must be measured through the learning outcome test. The formula used to determine the validity of the objective form test item is to use the point biserial correlation technique. When the correlation coefficients for all items have been calculated, it is necessary to determine the smallest number that can be considered “high enough” as an indicator of the consistency between the item scores and the overall scores. The interpretation of the validity of the test item in the objective form is determined by the following standards: (a) If  $rpbi \geq rtabel$  means the item being tested is valid; (b) Apabila  $rpbi < rtabel$  means the item under test is invalid. To interpret the magnitude of the correlation coefficient is as follows: (a)  $0,80 < r_{xy} < 1,00$  very high validity (very good); (b)  $0,60 < r_{xy} < 0,80$  high validity (good); (c)  $0,40 < r_{xy} < 0,60$  moderate validity (sufficient); (d)  $0,20 < r_{xy} < 0,40$  low validity (less); (e)  $0,00 < r_{xy} < 0,20$  is very low validity (ugly); (f)  $r_{xy} < 0,00$  not validity (Arikuto, 2013: 89).

The number that shows the difficulty and ease of a question is called the *difficulty index*. The magnitude of this difficulty index between 0,00 up to 1.0. This difficulty index shows the level of difficulty of the question. Problem with difficulty index 0,0 indicates that the question was too hard, otherwise the index of 1.0 indicates that the question was too easy [6]. The percentage of the difficulty index difficulty index is interpreted on the criteria according to the following table : (a) 0 - 15% Very difficult, should be thrown away; (b) 16% - 30% tough; (c) 31% - 70% moderate; (d) 71% - 85% Easy; (e) 86 % - 100% Very easy, should be thrown away.

Explains that the distinguishing power of question items is the ability of a question item to be able to differentiate between high-skilled students and low-ability students so that it can be seen that students who have high ability to answer more questions answer correctly, while students who have low ability to answer the items mostly cannot answer the items correctly [5]. With distinguishing power classification; (a) D: 0.00 - 0.20: poor; (b) D: 0.20 - 0.40: sufficient; (c) D: 0.40 - 0.70: good; (d) D: 0.70 - 1.00: very good. The criteria for each question are as follows: (a) Negative - 9% = very bad (dumped); (b) 10%-19% = bad (should be thrown away); (c) 20%-29% = good enough (revised); (d) 30% - 49% = good; (e) 50% and above = very good.

Distractor function tests are carried out only on multiple choice items. Each item of multiple choice questions issued in a learning outcome test has been equipped with several possible answers or what is often known as an option or alternative [5]. The cheater that is not chosen by the tetter at all means that the deceiver is bad, too blatantly misleading. On the other hand, a distractor (distractor) can be said to function well if the distractor has a great appeal to test followers who do not understand the concept or lack mastery of the material [1].

Reliability is a series of measurements or a series of measuring instruments that have consistency when measurements are made using the measuring instrument repeatedly. The test reliability test aims to measure the reliability and consistency of the test in measuring data. revealed that reliability is the level or degree of consistency of an instrument. The reliability of the test relates to the question whether a test is accurate and reliable in accordance with predetermined criteria[8]. The reliability of the assessment tool is the determination or consistency of the tool in assessing what it is judged on." This means that whenever the assessment tool is used it will give relatively the same results. Thus, reliability can also be defined as consistency or stability [3]. Test reliability tests can use the KR 20[9]. To interpret the magnitude of the correlation coefficient is as follows: (a)  $0,80 < r_{xy} < 1,00$  very high reliability (very good); (b)  $0,60 < r_{xy} < 0,80$  high reliability (good); (c)  $0,40 < r_{xy} < 0,60$  moderate reliability (sufficient); (d)  $0,20 < r_{xy} < 0,40$  low reliability (less); (e)  $0,00 < r_{xy} < 0,20$  reliability is very low (bad); (f)  $r_{xy} < 0,00$  not reliability.

Evaluation as an action or process, generally has three main functions, namely measuring progress, supporting planning, and improving or making re-improvements. Or the evaluation function in general, more details are as follows: (1) To determine the progress and development and success of students after experiencing or carrying out learning activities for a certain period of time; (2) To determine the success rate of the teaching program; (3) For guidance and counseling purposes; (4) For the purposes of developing and improving the school curriculum in question[10].

Learning evaluation is very important to see student achievement in learning. Learning evaluation can be effective if you use the right measuring tool. There are two techniques in evaluating student learning outcomes in schools, namely evaluation with non-test techniques and test techniques. The test is a measuring tool in evaluating learning. Tests are conducted by educators to monitor the learning process and the end of learning. The test also functions (1) as an instrument to measure the progress that students have achieved after they go through the teaching and learning process with certainty. (2) as an instrument to measure the success of the teaching and learning process. Through this test, the teacher will be able to find out how far the programmed material can be reached by students.

## II. METHODS

This type of research is quantitative descriptive to find information and data that can be used to describe the quality of summative questions. This research will be conducted in the fifth grade of state elementary school 058238 Bukit Rata. Jl. PaluhTabuhanBatu Hundred. West Brandan District, Langkat Regency. Public elementary school 058238 Bukit Rata was chosen because it is a school that is accredited B and has a large number of students. This school has 6 parallel classes. Each class learns English as local content at the school. Population is defined as a generalization area consisting of: objects / subjects that have certain qualities and

characteristics that are determined by the researcher for study and then draw conclusions [7]. sampling is part of the number and characteristics of the population. The sampling technique in this study was carried out by *purposive sampling* of the sample class.

The sample in this study were students of class V, which consisted of 30 people consisting of men = 15 people and women = 15 people. Grade V students were chosen because students according to English lessons in class V had already studied the use of higher tenses. In conducting the item analysis, the researcher asked for help from 2 experts consisting of 2 lecturers from language and arts faculty majoring in Indonesian and English literature. Experts / validators assess and provide input using the validation sheet provided for the four instrument packages that have been compiled. states that the item validity of an item validity test is the accuracy of measuring an item (which is an integral part of the test as a totality), in measuring what should be measured through these items. The formula used to determine the validity of the objective form test item is to use the point biserial correlation technique with Anates version 4. the test reliability test can use the Kr 20.

### III. RESULT AND DISCUSSION

This test was validated by two experts, namely WildansyahLubis as a lecturer at the Postgraduate Program at the State University of Medan and MasitowarniSiregar as a lecturer in English Language and Literature at the Faculty of Language and Arts, State University of Medan. The results of the validation of the English test at the first meeting of the two experts based on material aspects, construct aspects and language aspects obtained a percentage of 73% with a good category. The results of this validation will be revised according to the suggestions of the validator. The suggestions from the validator are 1) images should be in color; 2) the answer option must be downward, not sideways; 3) drawing made of source; 4) writing should use 1.5 spaces; and 5) the writing must not be anything less than a letter. the results of the validation of the English test at the first meeting of the two experts based on material aspects, construct aspects and language aspects obtained a percentage of 90 % with a good category. Based on the results of expert validation, it can be said that the English test in the second semester can be used after being revised.

The test validation carried out in this research is the item validation test. Testing the validation of the items on the even semester test for class V was carried out using the biserial point correlation formula with the help of the Anates v.4 program. The calculation results are as attached in the appendix and then consulted on the  $r_{table}$  at a significant level of 5% and  $N = 30$ , then the  $r_{table}$  is 0.361. If  $r_{count} > r_{table}$  then the item is said to be valid, conversely, if  $r_{count}$  is less than or equal to  $r_{table}$ , then the item is said to be invalid. Based on the results of anates calculation from the English question even semester 12 valid questions or 48% of the total questions. The number of valid questions is 4 valid questions or 16 % of the total

questions. Furthermore, the number of questions that can be accepted is 17 questions, but 13 questions are not valid. So it is necessary to fix the problem as a whole.

Based on the results of anates analysis, a distractor can be said to be functioning if at least 5% of the test followers are referred to (Arikunto, 2013: 234). The results of the distractor function analysis will be presented in Table 1 as follows:

**Table 1. English Summative Test Distractor Function**

Distractor s	Information				
	Very good	Good	Not good	Bad	Very bad
A	8	4	2	4	2
B	6	5	3	4	1
C	5	3	5	1	0
D	9	6	5	2	0

Based on data distractor functions listed in Table 1, it is known that the distractor A has a very good title of 8, well as much as 4, less well by 2, bad and very bad by 4 as 2. distractor B has a very good predicate sebangak 6, well as much as 5, bad as many as 3, bad as many as 4, and very bad as many as 1. Distrktr C has a very good predicate as many as 5, good as many as 3, less good as many as 5, bad as many as 1 and very bad as many as 0. Diktraktor D has a very good predicate as many as 9, as many as 6 good, 5 as bad as bad, as many as 2 bad, and as many as 0 very bad.

Through the validity of the items, it can be seen the difficulty level of the questions. The difficulty level of the questions can be categorized as easy, medium, and difficult. The results of the difficulty level of the questions will be presented in table 2 as follows:

**Tabel2 Problem Difficulty Level**

No	No Test	Level of Dificulty	Total
1	1,2,5,6,7,8,10,16,21	Easy	9
2	3,4,9,11,12,13,14,15,17,18,19, 20,22,23,24,25	Medium	16

Based on the results of the analysis of the items presented in table 2, it can be seen that the difficulty level of the questions is only in the category of the difficulty level of the easy questions and the medium questions. Of the 25 items, 9 items in the easy category. There are 16 questions that have a moderate category.

The test validation carried out in this research is the item validation test. Testing the validation of the items on the even semester test for class V was carried out using the biserial point correlation formula with the help of the Anates v.4 program. The calculation results are as attached in the attachment then consulted on the  $r_{table\ p}$ , there is a significant level of 5% and  $N = 60$ , then the  $r_{table\ is\ 0,2542}$ . If  $r_{count} >$

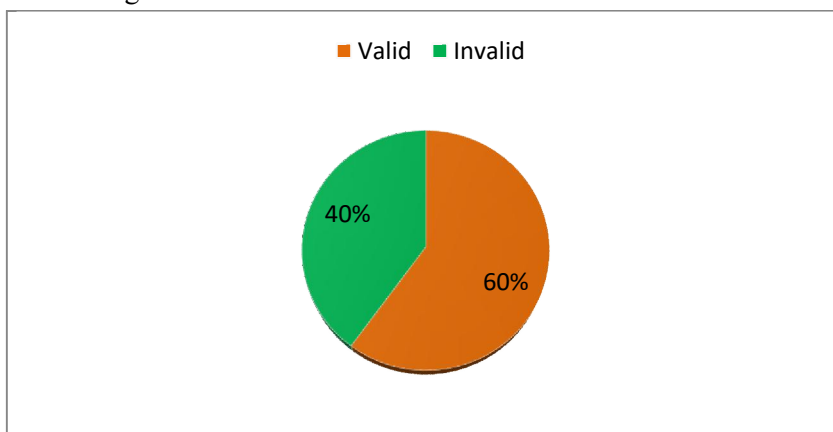
$r_{table}$  then the item is said to be valid, on the other hand, if  $r_{count}$  is less than or equal to  $r_{table}$ , then the item is said to be invalid. Based on the results of anates calculations from the English language questions even semester 16 valid questions or 64 % of the total questions.

The number of valid questions is 16 valid questions or 64% of the total questions. Furthermore, the number of questions that can be used is 15 questions, because one significant question has a low difference power, so that question is not used. The questions that can be used in this research are questions number 4,6,10,11,12,13,14,15,16,17,19,20,21,23,24. The data will be presented in table 3 as follows:

**Table 3 Distribution of Semester Exam Questions for English Subjects in Class V Elementary Schools**

No.	Validity Index	Item Questions	total	Percentage
	> 0.2542 (Valid)	4,6,10,11,12,13,14,15,16, 17,19,20,21,23,24	15	60%
	≤ 0.2542 (Invalid)	1,2,3,5,7,8,9,18,25	9	40%

Based on table 3, it can be seen that after the questions were validated by the expert, the data was obtained empirically validating that the number of valid tests that could be used in the study were 15 questions with a percentage of 60%. While the number of invalid tests were 9 questions with a percentage of 40%. The data will be presented in figure 1 as follows:



**Figure 1. Distribution of Problems Based on Validity**

Based on the results of anates analysis, a distractor can be said to be functioning if at least 5% of the test followers are referred to (Arikunto, 2013: 234). The results of the distractor function analysis will be presented in Table 4.4 as follows:

**Table 4 Functions of English Summative Test Distractors**

Distractors	Information				
	Very good	Good	Not good	Bad	Very bad
A	8	5	4	2	1
B	13	2	2	3	0
C	5	7	1	1	0
D	9	5	5	2	0

Based on the distractor function data shown in table 4, it is known that distractor A has a very good predicate as many as 8, good as many as 5, less good as many as 4, bad as many as 2 and very bad as much as 1. Distractor B has a very good predicate as many as 13, good as many as 2, less good as many as 2, bad as much 3, and very bad as 0. Distractor C has excellent predicate s ebanyak 5, well as much as 7, is not good in 1, bad and very bad in 1 total 0 Diktraktor D has a very good predicate as many as 9, well as much as 5, less good as many as 5, bad as many as 2, and very bad by 0.

Through the validity of the revised items, it can be seen the difficulty level of the questions. The difficulty level of the questions can be categorized as easy, medium, and difficult. The results of the difficulty level of the questions will be presented in table 5 as follows:

**Table 5 Problem Difficulty Level**

No	No Test	Level of Dificulty	Total
1	1,2,5,6,7,8,10,16,21	Easy	9
2	3,4,9,11,12,13,14,15,17,18,19, 20,22,23,24,25	Medium	16

Based on the results of the analysis of the items presented in table 5, it can be seen that the difficulty level of the questions is only in the category of the difficulty level of the easy questions and the medium questions. Of the 25 items, 9 items in the easy category. There are 16 questions that have a moderate category.

Reliability testing of even semester final exam questions on English subjects in grade V public elementary school 058238 Bukit Rata. The reliability test was carried out manually using the KR 20 formula with the help of the Excel program. The reliability test was carried out twice, namely the questions before expert validation and after expert validation. The results of the reliability test on the first question were the value -0.084 in the unreliable category because the value was  $\leq 0.700$ .

After the questions were validated by the experts, then the questions were tested on the students. Furthermore, the second test of English test reliability was carried out. The results of the second test reliability test, namely the KR 20 formula with the help of the Excel program obtained a value of 0.413 in the sufficient category.

**IV. CONCLUSION**

The level of validity of the English test in grade 5 elementary school before the revision based on material aspects, construct aspects and language aspects obtained a



percentage of 73% with a good category. After being revised, based on material aspects, construct aspects and language aspects obtained a percentage of 90% with a good category. The reliability test was carried out twice, namely the questions before expert validation and after expert validation. The results of the reliability test on the first question were the value -0.084 in the unreliable category because the value was  $\leq 0.700$ . After the questions were validated by the experts, then the questions were tested on the students. Furthermore, the second test of English test reliability was carried out. The results of the second test reliability test, namely the KR 20 formula with the help of the Excel program obtained a value of 0.413 in the sufficient category.

The level of validity and reliability of the English test in grade 5 elementary schools. Before the revision, the number of valid questions was 4 valid questions or 16% of the total questions. Furthermore, the number of questions that can be accepted is 17 questions, but 13 questions are not valid. So it is necessary to fix the problem as a whole. After the revision was made, the number of valid questions was 16 valid questions or 64% of the total questions. Furthermore, the number of questions that can be used is 15 questions, because one significant question has a low difference power, so that question is not used.

The validity and reliability of the English test in fifth grade public elementary school 058238 Bukit Rata should be revised because the results of the validity prior to the revision show that the average test validation result in the material aspect is 70% with a good category. The results of the validation on the construction aspect obtained an average of 69% with a good category. The results of the validation on the language aspect obtained an average of 80% with a good category. After being revised it is known that the average result of the test validation on the material aspect is 90% with the very good category. The results for the construction aspect average 88% with very good category. The results of the validation on the language aspect obtained an average of 95% with the very good category. The English test in Grade Five must be revised so that it is valid and reliable. Reliability is used as an indicator in believing the value of a test because it has consistency. While the test must be valid because it is used as a development and evaluation of a test. In addition, validity is also needed to determine the feasibility of items.

## V. ACKNOWLEDGMENTS

The authors say thank you to Medan State University. The authors thank you to Department of Basic Education which has provided support. Thank you for Mrs. Anni Holila Pulungan, and Mr. Rahmad Huseinas supervisor.

## REFERENCES

- [1] Arikunto, S. *Basics of Educational Evaluation*. **Earth Literacy**. 2005. pp. 33.
- [2] Arifin, Z. *Learning Evaluation*. **Remaja Rosda karya**. 2012. pp4.

- [3] Sudjana, N. *Research on Teaching and Learning Process Results*. **Remaja Rosdakarya**. 2014. pp135.
- [4] Kusaeri and Supranato. *Educational Measurement and Assessment* . **Graha Science**. 2012. pp75.
- [5] Sudjiono, A. *Introduction to Educational Evaluation* . **Raja Grafindo Persada**. 2008. pp163.
- [6] Hughes, A. *Testing for Language Teachers*. **Cambridge University Press**.2003. 8.
- [7] Sugiono, S. *Educational Research Methodology*. **Rineka Cipta**. 2013. pp363.
- [8] Arifin, Z. *Learning Evaluation* . **Remaja Rosda karya**. 2012. pp258.
- [9] Sinulingga, S. *Third Edition Research Methodology* . **Universitas Sumatera Utara press**. 2015. pp251.
- [10] Purwanto. *Evaluation of Learning Outcomes* . **Learning Library**.2011. pp5.