

CHAPTER I

INTRODUCTION

1.1 Background of Study

Learning evaluation is a very important part of measuring learning outcomes. This is because the results of the learning evaluation will provide information that is very important for student development. Through the evaluation of learning teachers can find out the success and failure of students. Evaluation is also a measuring tool for achieving learning goals. Evaluation as an assessment process in teaching and learning activities as a whole includes all the achievements of learning competencies. One form of learning evaluation that is concrete and numerical can be seen from student learning outcomes. Student learning outcomes can be obtained through assessment.

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 according to Arikunto (2005: 33) 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. Moreover, the test is an attempt to determine the achievement of educational goals. Measuring feedback on learning outcomes for teachers and students in carrying out the next learning process.

Besides being able to show learning outcomes, tests can also show that schools have developed the quality of their education to teach standards or not (Nissaipan, et al 2015). In the teaching and learning process, the teacher must

perform tests. Testing and teaching are closely related to other learning because teaching success can be measured and determined by conducting tests. Thus, the authors concluded that the test is an instrument used by teachers to evaluate the teaching and learning process and the results of the assessment can indicate whether students have developed the quality of their education or not. This refers to the characteristics of a good test.

The function of the test can be categorized into two functions (Sudijono, 2006: 67): (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.

In addition to reliability and validity, good tests need norms . The norm is the set of scores obtained by whom the test is intended. The scores obtained by these groups provide the basis for interpreting individual scores. The recognition of the objectivity of quality in a test is largely responsible for the development of the aroused and objective type tests.

Alsalihi (2019) stated that there are four characteristics of a good test; validity, reliability, objectivity, and norms. Based on characteristics of a good test, the two first are the most important; they are valid and reliable tests. Validity is the extent to which the test measures what it wants to measure. In other words, a valid test can actually measure what it is supposed to measure. For example, if the test is supposed to measure speaking ability, it is constructed, conducted, orally answered. Rajhy (2014) states that the term validity refers to the extent to which

the test measures what it says measures. Hughes (2003, p. 26) also states that a test is valid if it measures accurately what it is intended to measure. Validity refers to the suitability between the test as a measurement instrument and the domain of what it is supposed to measure.

Another important characteristic of a good test is reliability. Reliability refers to the consistency of the score. The test is considered a good test and can be trusted as a measure if the test scores are consistent in the number of times the test is performed. Consistency doesn't mean the scores have to be the same, but it can be a steady change in score.

There are three types of tests: Diagnostic, Formative, and Summative. Diagnostic tests identify students' current knowledge of a subject, their skills and abilities, and to clarify misunderstandings before teaching takes place. Knowing students' strengths and weaknesses can help teachers better plan what to teach and how to teach it. Formative tests provide feedback and information during the teaching process, when learning is taking place, and when learning is taking place. Formative tests measure student progress, but can also measure the progress of the teacher as an instructor. In other words, the main focus of formative testing is to identify areas that may need improvement. Summative test is one of the tests categorized by function in the teaching and learning process.

As research conducted by. This shows that summative tests and formative tests must be carried out continuously so that the success of learning English in schools can be evaluated for its success. As learning English in elementary schools is a foreign language for students. Hardjono Rayner (2001: 25) argues

that English is an international language so that it is the language most widely used throughout the world. We can see the position of English as an international language with the presence of anglofon speakers (English speakers) spread across five continents. English is not only used by anglophone speakers, but is used by people in the world, especially people who tend to be modern. This is also due to various advantages in English, including in the richness of its idioms (special expressions), which are more varied and always evolving than other European languages .

English in elementary schools uses the Education Unit Level Curriculum. In one academic year, it consists of two semesters. Semester one and semester two. In the first semester there are two exams, namely the midterm and the first semester final exams . Likewise in the second semester, midterm examinations and second semester final exams. So, in one teaching year there are four exams, namely two midterm exams and two final semester exams. So the semester exam questions lead to summative tests .. Each question, three forms of multiple choice questions, answers and descriptions. Multiple choice questions give a score of 1 for each question, entry questions are given a score of 2 for each question and essay questions are given a different score, depending on the problem level of the question. In this study, researchers focused on multiple choice test items.

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.

Based on the results of interviews conducted with English teachers in fifth grade public elementary school 058238 Bukit Rata. The fifth grade was chosen because based on information obtained from the homeroom teacher, the fifth grade students were very active in learning. Class five also has a homogeneous average ability. Based on the results of the interview, it is known that the questions being tested were made by the Langkat district education office. The number of questions is 40 questions consisting of 25 multiple choice questions, 10 questions for answers and 5 questions in essay. Question making is done based on competency standards and basic competencies as well as competency achievement indicators. The teacher distributes questions made by the service during the semester exams. In the process of analysis, the teacher only compares the previous year's questions to be used as learning references. However, the teacher did not analyze the items, validity, reliability, difficulty level, and item difference. So it can be concluded that the English test used in fifth grade public elementary school 058238 Bukit Rata is invalid. A test can be said to be valid if the test performs its measuring function, or provides precise and accurate measurement results in accordance with the purpose of the test. A test that produces data that is not relevant to the purpose of the measurement is said to be a test that has low validity.

Problems made by the department , after the analysis was carried out, there were still deficiencies in the guidelines for making questions. This can be seen in question number 8. Number 8. The old woman is my ... this question uses an unclear sentence (Old woman). Problem number 12. The answer choices are not uniform. In the rule of making multiple choice questions, there must be uniformity between the distractors and the answer choices. The choice should be uniform. (C. Mortuary is a morgue). Problem number 15. My sister should be my sister. Number 18. The choice of answers is not uniform. (D.Kitchen). question number 19. The answer choices are not uniform. You should be able to use school-related words. (B. Lawyer. And d.architect).

In addition, English questions should include four basic competencies, namely speaking, reading, listening and writing. In the exam, the four competencies should be raised, so that the evaluation can be measured properly. The results of this analysis indicate that the questions made by the Langkat Regency education office also do not meet the standard for making questions. So it needs to be revised so that students can understand the meaning of the questions well.

This has an impact on the quality of education in public primary schools. Based on the results of observations of students' parents, many parents prefer to send their children to Islamic -based private schools that have excellent English learning facilities. So that parents' trust in public schools decreases. This shows that the test is very important to determine learning success and increase public confidence in schools.

Based on the results of the interview, the teacher did not do item analysis because it took a long time, was complicated, and 70% of the teachers did not understand how to do item analysis. This is in line with research conducted by Salwa (2012). The findings are in the form of index values of validity, reliability, difficulty level, and differentiation in terms of quantitative analysis. Whereas in qualitative analysis, the findings are in the form of the percentage of questions that match the curriculum, and some errors that exist on both tests. From the findings, it can be concluded that the quality of the two questions is good in terms of quantitative. The values of validity, reliability, level of difficulty, and differentiation are both balanced. However, from a qualitative point of view, question 1 is better than question 2. This is due to the fact that there are more errors found in test item 2 than in test 1. The writer suggests that the maker of test 2 must be careful and pay attention to the provisions of making good questions. on the preparation of the next test.

Sugianto (2011) Based on the result of the data analysis, the writer found that the English formative test conducted on October 2008 belongs to high validity in its content validity and excellent validity in its construct validity. The English formative test conducted on November 2008 belongs to low validity in its content validity and high validity in its construct validity. While the English formative test both conducted on October 2008 and November 2008 are unreliable.

As the problems have been described, the researchers are interested in analyzing the validity and reliability of the English test items in fifth grade public elementary school 058238 Bukit Rata. This aims to determine the quality of a

good test. As research conducted by Bernasela (2014) Based on the whole analysis of test items, it can be concluded that there are 33 good test items which still can be used to the next summative test, 6 test items should be discarded or changed by the other test item and 11 test items should be revised. So this study contributes that the tests used in learning are only valid tests. The goal is that student learning outcomes are truly measurable because they use valid and reliable tests on foreign language education in public elementary schools 058238 Bukit Rata.

The validity in this study was measured based on the content validity and the instrument validity. Content validity was carried out by 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 . While the validity of the instrument was carried out based on the item validity test, the level of difficulty, the difference in the items, and the distractor function test. The results of the instrument validity analysis were assisted by using *Annater Version 4 software* . While the reliability of the test in this study was analyzed using the KR 20 formula with the help of *Microsoft Excel 2010* . This research only leads to a multiple choice summative test in the fifth grade of the state elementary school 058238 Bukit Rata.

1.2 Identification of Study

Based on the background of the problem, several problems that have been identified can be described as follows:

1. The teacher did not analyze the items so that the quality of the questions tested on students could not be known.
2. The reason the teacher did not do item analysis was because the process was very complicated
3. As many as 70% of students did not understand the item analysis.

1.3 Problems of Study

Based on the problems that have been limited, the formulation of the problems in this study are:

1. What is the level of validity and reliability of the English language test in fifth grade public elementary school 058238 Bukit Rata in the 2019/2020 academic year?
2. How is the suitability of the level of validity and reliability of the English test in fifth grade public elementary schools 058238 Bukit Rata in the 2019/2020 academic year?
3. Why should validity and reliability be revised ?

1.4 The Objectives of Study

This research aims to:

1. Knowing the level of validity and reliability of the English language test in fifth grade public elementary schools 058238 Bukit Rata in the 2019/2020 academic year.
2. Knowing the suitability of the level of validity and reliability of the English test in fifth grade public elementary schools 058238 Bukit Rata in the 2019/2020 academic year.

3. Knowing the validity and reliability must be revised

1.5 Scope of Study

In order not to deviate from the problem and be more focused, this research limits the scope, namely the summative test. Summative test is a test that is used to determine student mastery of all the amount of material presented in a certain period of time such as quarterly or semester. In teaching practice summative tests are known as final or quarterly examinations depending on the time unit used to complete the material (Purwanto, 2011: 67). The focus of research is only on the validity and reliability of multiple choice tests on the English subject.

1.6 Significant of Study

Based on the background of the problem and the formulation of the problem, this research is expected to provide useful theoretical and practical contributions.

1. Theoretical Benefits

Theoretically, the results of this study are useful for learning English in public elementary schools so that teachers can analyze the items, namely the validity and reliability of summative test questions before they are distributed to students. It is intended that students can easily understand the contents of the questions and can improve student learning outcomes.

2. Practical Benefits

In practical terms, the results of this study are especially useful for:

- a. Teacher

The results of this research can be used as a reference in making summative test questions. So that teachers are more enthusiastic in designing tests.

b. Students

Students will find it easier to understand the meaning of the questions if the questions have been tested for validity and reliability accurately.

c. Further researchers

As for researchers, this study can be a reference to enrich knowledge about the theory and practice of related topic tests and their analysis.

