

CHAPTER II

REVIEW RELATED LITERATURE

A. Language Test

1. Definition of Test

In order to know how well the learning process, the teacher must evaluate it. There are many ways that use to evaluate the teaching learning process. One of the ways is test. A test is a method of measuring a person's ability, knowledge or performance in a given domain (Brown, 2000:384). In this definition, Brown wants to show that people's knowledge can be explored through testing. Ur (1996:34) said that the test used as a means to motivate students to learn or review specific material. Through the test, the teacher does not only motivate and measure the students' ability, but also improve the lesson in teaching learning process. Arikunto (2012:66) defines test is tool or procedure used to know or measure student's performance by using a set of ways and rules. When we talk about test we will be familiar with the term of testing. Testing is the way teacher give the test to the students. Testing is very important for teaching and learning process.

Johnson (2001:191) states that testing important for almost all the people involved in the education process. The learners want to know how well he is doing, and wants the 'piece of paper' at the end of the course that will help open professional doors. The teacher wants to know not only how the learner is progressing, but also how he, the teacher, is succeeding in his job. Then there are

parents, educational authorities and countless others who have some interest in the learner's progress.

2. The Types of Test

There are some types of test commonly used in education; Johnson (2001:192) classifies those into five types:

a. Achievement test

Achievement tests are concerned with how well a learner has done in relation to a particular course or program. They usually come at the end of programs, and they deliberately based on the content covered in it.

According to Hughes (2003:13), achievement test are directly relate to language course, their purpose being to establish how successful individual students, group of students, or the courses themselves have been in achieving objectives. There are two kinds of tests: final achievement tests and progress achievement tests.

- 1) Final achievement tests are those administered at the end of a course of study.
- 2) Progress achievement tests, as their name suggests, are intended to measure the progress that students are making.

Gronlund (1968:4) states about the basic principles to have a good achievement test form, a test maker should consider that achievement as following below:

- a) Achievement test should measure clearly defined learning outcomes that are in harmony with the instructional objectives.

- b) Achievement tests should measure an adequate sample of the learning outcomes and subjects matter content included in instruction.
- c) Achievement test should include the types of tests items, which are most appropriate for measuring the desired learning outcomes.
- d) Achievement tests should be designed to fit the particular uses to be made of the results.
- e) Achievement test should be made as reliable as possible and should then be interpreted with caution.
- f) Achievement tests should be used to improve student learning.

b. Diagnostic test

Diagnostic tests are also concerned with achievement, or lack thereof. They are often geared to assess the success of some small stretch of teaching, and perhaps to suggest where remedial work will be required.

c. Proficiency test

Proficiency tests do not relate to any specific content or program. They are tests of what level has been reached in the language, and stand independent of any course.

d. Placement tests

Placement tests are a particular sort of proficiency test. They are given at the beginning of language-teaching programs, to help with the coming teaching program in mind.

e. Aptitude tests

An aptitude test looks at ‘how well you would do’. It has a strong predictive element to it. According to Frankel and Wallen (2003:135), aptitude test assess intellectual abilities that are not, most cases, specifically taught in school. Aptitude tests are intended to measure an individuals’ potential to achieve; in actually, they measure present skills or abilities. They differ from achievement tests in their purpose and often in content, usually including a wider variety of skills or knowledge. The same test may be either an aptitude or an achievement test, depending on the purpose for which it is used.

A language aptitude test is designed to measure a person’s capacity or general ability to learn a foreign language and to be successful in that undertaking. Aptitude tests are most often used to measure the suitability of a candidate for a specific program of instruction. Thus, these tests given before the students begin to study to select them in appropriate section or level of their ability.

3. The Categories of Tests

According to Brown (1996:2), there are two categories of tests can help language teachers to make much better decisions about theirs student:

a. Norm-Referenced Tests

In brief, a norm - referenced test (NRT) is designed to measure global language abilities (for instance, overall English language proficiency, academic listening ability, reading comprehension, and so on)- Each student’s score on such a test is interpreted relative to the scores of all other students who took the test. Such

comparisons are usually done with reference to the concept of the normal distribution (familarly known as the bell curve). The purpose of an NRT is to spread students out along a continuum of scores so that those with “low” abilities in a general area such as reading comprehension are at one end of the normal distribution, while those with “high” abilities are at the other end (with the bulk of the students falling near the middle). In addition, while students may know the general format of the questions on an NRT (for example, multiple-choice, true-false, dictation, or essay), they will typically not know what specific content or skills will be tested by those questions.

b. Criterion-Referenced Tests

In contrast, a criterion-referenced test (CRT) is usually produced to measure well-defined and fairly specific objectives. Often these objectives are specific to a particular course, program, school district, or state. The interpretation of scores on a CRT is considered absolute in the sense that each student’s score is meaningful without reference to the other students’ scores. In other words, a student’s score on a particular objective indicates the percent of the knowledge or skill in that objective that the student has learned. Moreover, the distribution of scores on a CRT need not necessarily be normal. If all the students know 100% of the material on all the objectives, then all the students should receive the same score with no variation at all. The purpose of a CRT is to measure the amount of learning that a student has accomplished on each objective. In most cases, the students would know in advance what types of questions, tasks, and content to expect for each objective because the

question content would be implied (if not explicitly stated) in the objectives of the course.

4. The Types of Test Item

There are 12 types of the test items, such as; questions and answers, true/false, multiple-choice, gap filling and completion, matching, dictation, cloze, transformation, rewriting, translation, essay, and monologue. The tests are support by Ur in Fauzi (2011), that there are 12 kinds of test item:

a. Questions and Answers

Simple questions, very often following reading, or as part of an interview; may require short or long answers.

Example:

What is the (family) relationship between David Copperfield and Mr. Murdstone?

b. True/False

A statement is given which is to be marked true or false. This may also be given as a question, in which case the answer is, yes or no.

Example:

Addis Ababa is the capital of Egypt.

Is Addis Ababa the capital of Egypt?

c. Multiple-Choice

The question consists of a stem and a number of options (usually four), from which the tester has to select the right one.

Example:

The person who writes books is called

- a. A booker
- b. An author

d. Gap-filling and Completion

The testee has to complete a sentence by filling a gap or adding something. A gap may or may not be signaled by a blank or dash; the word to be inserted may or may not be given or hinted at.

Example:

They (go) to Australia in 1980.

They to Australia in 1980.

e. Matching

The testee is faced with two groups of words, phrases or sentences; each item in the first group has to be linked to a different item in the second.

Example:

Small Big

Many Little

f. Dictation

The testee dictates a passage or set of words; the tester writes them down. This mainly tests spelling, perhaps punctuation, and perhaps surprisingly on the face on it, listening comprehension: people can only usually writes word down accurately from dictation if they understand them. If does not, however, test other writing skills or speech, and involves very little reading.

g. Cloze

Words are omitted from a passage at regular intervals (for example, every seventh word). Usually the first two or three lines are given with no gaps.

h. Transformation sentence is given; the tester has to change it according to some given instruction.

Example:

Put into the past tense:

I go to school by bus.

i. Rewriting

A sentence is given; the tester rewrites it, incorporating a given change of expression, but preserving the basic meaning.

Example:

He came to the meeting in spite of his illness. Although

j. Translation

The testee is asked to translate expressions, sentences or entire passages to or from the target language.

k. Essay

Essay the testee is given a topic, such as 'childhood memories', and asked to write an essay of a specific length.

1. Monologue

The taster are given a question and asked to speak about it for a minute or two. This test oral fluency in 'long turn' – something not everyone can do in the mother tongue it also tests overall knowledge of pronunciation, grammar, and vocabulary.

5. The Characteristic of a Good Test

While a teacher considers making a test, she or he needs to recognize the various test of each case that related to the purpose, time and subject. But, the most important consideration to characterize a test being a good test needs some qualifications, they are: validity, reliability and practicality. Arikunto (2009:58) states that a test which is good as the measuring instrument must meet the test requirements, namely to have validity, reliability and usability.

Yet, in this case the writer would like to explain more about validity.

B. Validity

Based on the previous explanation, the writer mentions that one of characteristics a good test is validity. Heaten in Johnson (2001:302) defines validity as "the extent to which (a test) measures what it is supposed to measure and nothing else." If a test is valid, the outsider who looks at an individual score knows that it is a true reflection of the individual's skill in the area the test claims to have covered. Validity is the extent to which the test actually measures what it is intended to measure (Brown, 2000:387) it is also the extent to which inferences made from

assessment results are meaningful, and useful in term of the purpose of the assessment.

Validity also can defined as the extent to which the instrument measures what it should measured, so the test should test what the writer or teacher wants to test the students. The expert should look into whether the test content is representative of the skills that are supposed to be measured. This involves looking into the consistency between the syllabus content, the test objective and the test contents. If the test contents cover the test objectives, which in turn are representatives of the syllabus, it could be said that the test possesses content validity (Brown, 2002: 23).

Brown's idea is supported by Hughes (2005:26), who stated that a test is said to have content validity if its content constitutes a representative sample of the language skills, structures, etc. in which it is meant to be concerned. It means that a test will have content validity.

Linn and Gronlund (1995:307) state that validity refers to the adequacy and appropriateness of the interpretation made from tests, with regard to a particular use. An information data can be said is valid in accordance with actual circumstances.

According to Sukardi (2009:38), there are many factors that affect the evaluation of the test results are not valid, such as:

- 1. Internal factor**

- a. Test instructions meaning are unclear so reduce the validity of the test.

- b. The words used in the structure of the evaluation instrument are too difficult.
So, student cannot understand the test
- c. The test items' construct are bad.
- d. The level of difficulty of test items is not appropriate with students learning material.
- e. The time allotted is not appropriate, it is make the time for doing the test are too fast of too long
- f. The number of test items too little, so cannot represents a sample of the learning material.

2. The external factor

- a. The time that are used to do the exercises are to short. This condition make student are not careful in doing the test
- b. There are many students are not honest in doing the test
- c. The instructions from the invilgilator are not clear.
- d. The scoring techniques are not consistens
- e. Student can not understand the instruction in the test
- f. There are joki that do the student's test

3. Factor that come from the student's answer

For many times, test items are not valid because of the answer of students. For example, student feels nervous before examination. This condition maybe because of the teacher is fierce that make student feel nervous and fail in the test.

Another example the condition of the classroom is noise that disturbed student concentrate.

According to Johnson (2001:301), there are four types of validity namely content validity, face validity, and construct validity, empirical validity and predictive validity.

1. Content validity

Content validity is about what actually goes into the test. To have content validity, a test's content must be seen as representative of the subject area being covered. Arikunto (2012:59) said a test is said to have content validity if it measure the particular goals with is equal to the material or the content of the lesson.

The content validity is concerned with how the test measures the subject matter and behavior under consideration. The definition of content validity states that the test items must be a representative sample of the domain of possible content or behavior. Fernandes (1984:44), He said that content validity is the most appropriate method for evaluating the validity of achievement tests.

Sukardi (2009:32-33) said that content validity is the degree in which a test measure the scope of substances that want to be measured. To get the content validity requires two important aspects, valid content and valid sampling techniques. Valid content includes, matters relating to whether evaluation items describe the measurements in the scope of what wants to be measured. While the validity of sampling techniques generally relates to how good a sample test items present a total coverage of content.

Content validity also has a very important role in achievement test. Content validity is generally determined through consideration of the experts. There is no mathematical formula to calculate and no way to show for sure. However, to know how a test is validated by using the content validity, expert consideration is done in a way as follows. First, the experts are kindly requested to observe carefully all items in the tests validated. Then they asked to correct interpretation of the items that have been made. At the end of the repair, they are also asked to provide considerations about how good the test interpretation the evaluation describes the scope of content to be measured. The consideration of the experts also regarding whether all aspects to be measured have been covered through the interpretation of the questions in the test items. Or in other words a comparison made between what to put with what it wants to be measured which has reflected into the purpose of the test.

2. Face validity

Face validity, on the other hand, is to do with what the world thinks of the test. It relates again to that all-important figure in the testing world, the outsider. He must be able to look at a test and, as a layperson, be convinced that it is a test giving valid information about language use.

3. Construct validity

Construct validity deals with the relationship between a test and a particular view of language and language learning.

According to Bachman and Palmer (1996:22) construct validity is the on-going process of demonstrating that a particular interpretation of test scores is

justified, and involves, essentially, building a logical case in support of a particular interpretation and providing evidence justifying interpretation.

Fernandes (1984:46) in his book *Testing and Measurement* construct validity is usually established by examining the pattern of relationships of the test in question with tests measuring other constructs. Construct validity can be tested in the following different ways:

- a. The correlation between different tests which are assumed to measure the same variable.
- b. The correlation between single items or different parts of the test. The parts of the test must have a high intercorrelation if the test is to be regarded as measuring a unitary variable.
- c. The study of how the test results are influenced by changes in individuals or environment which, according to the theory, should respectively influence or fail to influence the individuals' positions on the continuum.
- d. The study of differences between groups which should differ according to the theory for the variable.

4. Empirical validity

Empirical validity deals with how the test relates to other testing measures. Test should not yield results that are dramatically at odds with the results of other forms of assessment.

In order to know whether a test has empirical validity or not, it can be traced from ways, first is concurrent validity and second is predictive validity. Concurrent

validity applies if data on the two measures - test and criterion - are collected at or about the same time. Predictive validity applies if there is an intervening period (e.g., three or six months) between the time of testing and the collection of data on criterion. Operationally, this time of criterion data collection is the distinction between the two types of criterion validity. Specifically, the question of concurrent validity is whether or not the test scores estimate a specified present performance; that of predictive validity is whether or not the test scores predict a specified future performance.

5. Predictive validity

Predictive validity where its results yield some information about the future. It may be particularly useful to find out whether placement tests.

C. Review of Previous Study

A review of previous study that is relevant is conducted by Nofiyanti (2011). Her thesis entitles *‘Analysis on The Content Validity of The Summative Test for The First Year Students of Junior High School’*. Faculty of Tarbiyah and Teachers’ State Islamic University Syarif Hidayatulloh, Jakarta. The subjects in this study were the English teacher and the students of first grade SMP N 87 Jakarta. The object of the research was the test items of English summative test first year students of SMP N 87 Jakarta, which consist of 50 items of multiple choices.

This research was aimed in analyzing the English summative test validity used by SMP N 87 Jakarta for even semester of first grade students, it consist of (1)

the English material displayed within the English summative test used at SMP N 87 Jakarta, (2) the English material coverage by English summative test appropriate with the content standard of the latest syllabus suggested, and (3) the validity of English summative test at SMP N 87 Jakarta.

In this study, the writer employed comparative analytic method, in the sense, the writer compared the contents of the test to the syllabus in used to find out whether or not English summative test in the first grade of even semester SMP N 87 Jakarta has good content validity.

This findings of the research stated that the content validity of English summative test at SMP N 87 Jakarta for even semester of second grade student reach good validity.