### CHAPTER II

### REVIEW OF RELATED LITERATURE

## A. Language test

### 1. Definition of Language Test

Achievement is one of the purposes for teachers and students have a teaching learning process. How to know the achievement of our goal is using evaluation. We evaluate the process of learning during the period of a teaching-learning, revise the weaknesses and inquiring for the way to increase it.

In teaching learning, we meet with a curriculum as a manual theory of the standardization of our Education. Test becomes a tool to evaluate the achievement and the degree of the successes in teaching – learning process. By conducting this test, be provided that the students are able to evaluate themselves when they get a bad score and they are motivated to get better score in the next test.

There are some explanations from some experts about the definition of a test. Henry Braun and friends argue that testing refers to the process of administering a test to measure one or more concepts, usually under standardized conditions.<sup>1</sup> In addition, test or assessment is among an instructor's most essential educational tools, when properly developed and interpreted, assessment can help teachers better understand what their student are learning.<sup>2</sup>

Harmer also argues that test is conducted to measure the student's abilities to see if they can enter a course or institution, sometimes it is to see how well they increase their ability, and sometimes it is because the students themselves want a qualification.<sup>3</sup> In the other way, Macalister in his book states that testing is only one way of gaining information about the progress of learners and the effectiveness of the course.<sup>4</sup>

The other definition, test is the practice of making objective judgments regarding the extent to which the system (device) meets, exceeds, or fails to meet stated objectives. Besides, testing is about verification that what was specified is what was delivered: it verifies that the product (system) meets the functional, performance, design, and implementation requirements identified in the procurement specifications and also testing is about managing risk for both the acquiring agency and the system's vendor.<sup>5</sup> The

\_

<sup>&</sup>lt;sup>1</sup>Braun, Henry, Anil Kanjee, Eric Bettinger, and Michael Kremer. (2006). *Improving Education through Assessment, Innovation, and Evaluation*. Cambridge, MA: American Academy of Arts and Sciences P. 9 <sup>2</sup>Haertel, Edward H. (2013). *Reliability and Validity of Inferences About Teachers Based On Student* 

Test Scores. (www.ETS.org accessed on March, 23, 2014)

<sup>&</sup>lt;sup>3</sup>Harmer, Jeremy.(2007). *The Practice of Eglish Language Teaching Fourth Edition*. USA: Longman Press p. 379

<sup>&</sup>lt;sup>4</sup>Language Curriculum design, p. 10

<sup>&</sup>lt;sup>5</sup>Testing for Transportation Management Systems: Q & A (tmcpfs.ops.fhwa.dot.gov/.../Final%20Q&A.pdf- accessed on March, 28, 2014)

last simple meaning of validity, validity is concerned with the meaningfulness of research components. When researchers measure behaviors, they are concerned with whether they are measuring what they intended to measure.<sup>6</sup>

Some of those experts' explanations about test tell about the utility of test as a tool in measuring the goal and achievement of teaching – learning process. From all of those statements about the definition of test above, we can conclude that test is a way to measure the achievement of learning process in the end or in the beginning of learning process. Test is conducted in the beginning of the semester, with a purpose to make better process of teaching – learning after conducting test. In the other hand, when a test is conducted in the end of the semester, it is useful for showing about the achievement of the goal during the teaching – learning process in one semester.

#### 2. The Kind of Test

There are many types of test used to measure student's achievement.

According to Tinambunan, there are four types which are commonly used by teachers in the classroom:

<sup>&</sup>lt;sup>6</sup>Drost, Ellen A.(2011). *Education research and perspective*. California State University, Los Angeles: Vol. 38, No.1, Validity and reliability in social science research p. 114

#### a. Placement test

A placement test is designed to determine the treatment for students based on their performance at the beginning of instruction.

#### b. Formative test

It is used in the end of a unit in the course book or after a lesson designed. The result of this test will also give the students immediate feed back

### c. Diagnostic test

Diagnostic test is intended to diagnose learning difficulties during instruction. Thus, the main aim of diagnostic test is to determine the causes of learning difficulties and then to formulate a plan for a remedial action.

## d. Summative test

The summative test is intended to show the standard that the students have to know reached in relation to other students at the same stage. Therefore it typically comes at the end of a course or unit of instruction. <sup>7</sup>

#### 3. 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. In

<sup>&</sup>lt;sup>7</sup>Tinambunan, Drs Wilmar.1998. Evaluation of students achievement. Jakarta: Depdikbud p. 7-9

the other hand, the most important consideration to characterize a test being a good test needs some qualifications, they are validity and reliability.

In addition, Jeremy harmer states in his book about the characteristic of a good test. He argues that test which effective for the learner can be measured, those are contain of validity and reliability.<sup>8</sup> For brief explanation about those characteristics of a good test, these are explanations of each item:

### a. Reliability

It is difficult to decide that the test is a good test when we only look at the process of test making but do not pay attention to the condition of test taker. Reliability is an extent to which a questionnaire, test, observation, or any measurement procedures produces the same results on repeated trials. In short, it is the stability or consistency of scores over time or across raters. <sup>9</sup> The other expert explains that reliability is one of the most important elements of test quality. It has to do with consistency, or reproducibility, or an examinee's performance on the test. <sup>10</sup>

The other linguist gives an addition characteristic of a good test that after we look at the validity and reliability of a test then we have to know

<sup>&</sup>lt;sup>8</sup>Harmer, Jeremy.(2007). *The Practice of Eglish Language Teaching Fourth Edition*. USA: Longman Press p. 381

<sup>&</sup>lt;sup>9</sup>Michael J. Miller, Ph. D. *Graduate Research Methods*. Western International University,

<sup>&</sup>lt;sup>10</sup>Professional Testing Inc.(2006). How do you determine if a test has validity, reliability, fairness, and legal defensibility? p.1

that a good test should be Practicality. It refers to facilities that available to test developers regarding both administrations and scoring procedures of a test.<sup>11</sup> In addition, we have to consider about the tests that we have made whether they are applicable or not.

### b. Validity

Validity is the level of the measurement based on the test making process. From all appearances, this test has been appropriate with the target of the test. Validity is arguably the most important criteria for the quality of a test. The term validity refers to whether or not the test measures what it claims to measure. The other linguist gives an explanation about validity by giving a short example. Test is designed to be a measure of Algebra and the items are algebraic in nature and have been deemed to be relevant to knowing algebra, the results from the test are said to be valid. 13

Validity is divided into some basics, they are; content, criterion and construct validity. Gebotys has a brief explanation about those three kinds of validity;<sup>14</sup>

1

<sup>&</sup>lt;sup>11</sup>Hossein Farhady, Fundamental concepts in Language testing. University for teacher education, Iran university of Science and Technology. P.26

<sup>&</sup>lt;sup>12</sup>How Do You Determine If A test Has Validity Reliability, Fairness, and Legal Defensibility? Profesional Testing Inc, 2006 p.1

<sup>&</sup>lt;sup>13</sup>Vansickle, Timothy R. (2008). A testing Brief Validity. Apple valey, USA: Questar Assessment Inc p. 2

<sup>&</sup>lt;sup>14</sup>Gebotys, Dr. Robert. (1999). Handout on Validity. Wilfried Laurier University, Canada p. 1

## a. Content Validity

Content Validity is employed when it seems like that test users will want to draw references from observed test scores to performances on a larger domain of a task similar to items on the test. In addition, Vansickle also argues about the definition of a content validity, that this form of validity is not a statistical form of validity but rather a process of building validity into the test from beginning of the design process. A valid test can convey the effect of variation in the attribute one intends to measure. It means that the relation between test scores and attributes is not correlated but causal. A test is valid for measuring an attribute if variation in the attribute causes variation in the test scores. 15 The other expert, Harrison said that content validity is concerned with what goes into the test. The content of a test should be decided by considering the purposes of the assessment and draw up a list known as content specification. The content specification is important, because it ensure as far as possible that the test reflects all the areas to assessed in suitable proportions and also because it represents a balanced sample, without bias towards the kinds of items which are easiest to write or towards the test material which happens to be available. 16

Basically, content validity is considered with how well the sample of test tasks represents the domain of task to be measured. Content validity is

<sup>&</sup>lt;sup>15</sup>Borsboom, Denny, Gideon J. Mellenbergh and Jaap Van Heerden.(2004). *The Concept of Validity.* Washington DC: American Psychological Assosiation p. 1067

<sup>&</sup>lt;sup>16</sup>Harrison, Andrew. (1983). A language Testing Handbook. London: Macmillan Press. P. 11

also called curricular validity, because materials that will be tested are based on curriculum. Therefore, the way to know the content validity is by looking the objective of curriculum and the material of the test. To measurement of content validity is important in employment setting, where tests used to hire and promote people are carefully scrutinized for their relevance to the job.<sup>17</sup> So the test makers must be careful to make a test to be good test based on the syllabus.

# b. Criterion Related to Validity

Criterion related validity is a study of the relationship between test scores and practical performance criterion that is measurable. When a test, X, can be related to a criterion score, Y, criterion – related validity can be determined.

#### c. Construct Validity

Construct validity is appropriate whenever the test user wants to draw inferences from test scores to a behavior domain which cannot be adequately represented by a single criterion or completely defined by a universe of content. Bachman defines construct validity as the extent to which performance on tests is consistent with predictions that we make on the basis of the theory of abilities.<sup>18</sup>

.

<sup>&</sup>lt;sup>17</sup>Jay, Cohen Ronald. (1988). *Psychological Testing and Assessment*. California: Mountain view p. 178.

<sup>&</sup>lt;sup>18</sup>Alavi, Taher (2012). *The Predictive Validity of Final Exams as a Measure of Success in Iranian National University Entrance English Exam.* Teharan, Iran:Allame Tabatabaee University,. (Journal of Language Teaching and Research, Vol. 3, No. 1, p. 224 – 228, January 2012).

## 4. Content validating of the English Placement test (EPL)

The purpose of conducting the placement test is to ascertain a student's level of preparedness in certain basic skill areas to discover whether he or she can perform school level work.<sup>19</sup> Therefore, placement test also called with the Basic competence test.

In the other hand, Content validity is a form of validity based on the degree to which a test adequately and sufficiently measures the particular skills of behaviors it sets out to measure. As Henning articulates that any test developer has to answer the fundamental question: Is the content of the test consistent with the stated of the goal for which test is being administered? In the other words, since congruence of the Basic Competence test and the curriculum refers to the coordination between the content of the test and the curriculum, the test should adequately measure what it is supposed to measure. Brown and Hudson also states that study of content validity requires a systematic investigation of the degree to which the items of a test, and the resulting scores, are representative of relevant samples of whatever content or abilities the test has been designed to measure.<sup>20</sup>

From the statement about the congruence of content validity of the placement test with a curriculum, it is made obvious that the higher content

<sup>&</sup>lt;sup>19</sup>Ewell, Peter T. *About learning assessment policies*. <u>www.sinclair.edu</u> accessed on 29<sup>th</sup>of January, 2015

<sup>&</sup>lt;sup>20</sup>Brown, James Dean and Hudson, Thom (2002). *Criterion-referenced Language testing*. Cambridge University press p. 87

validity a test receives, the more likely it is to be an accurate measure of what it is supposed to measure. According to Henning, the test developer has to avoid the threats to test validity such as misapplication of test, and inappropriate selection of content.<sup>21</sup>

### a) Misapplication of test

Test developer must consider carefully the designated purpose of the test, including the exact content or objectives to be measured, and the type of examinee to whom measurement is given.

## b) Inappropriate selection of content

Inappropriate selection of content commonly occurs when items do not match the objectives or the content of instruction. Also, this may happen when the test items are not comprehensive in the sense of reflecting all of the major points of the instructional program.

Mehrens and Lehmann, proposed some traditional steps of selecting the most suitable test.<sup>22</sup> Those are:

- a. Define the domain of interest
- b. Specify the test selected for possible adoption
- c. Identify a panel of content experts.

<sup>&</sup>lt;sup>21</sup> Henning, G. (1987). *A guide to languange testing: development-evaluation-research.* Rowley, Massachusetts: Newbury house. p.112

<sup>&</sup>lt;sup>22</sup> William A. and Lehmann, Irvin J. Mehrens (1987). *Using Standarized tests in Education*. Longman p.193

d. Have these experts review the test on an item-by-item basis and evaluate the degree to which items match the curriculum.

### **B.** Review of Previous Study

A study about content validity has been arranged since years ago. The Philosopher of science – cross cultural psychologist, Donald Campbell during the mid – 20<sup>th</sup> century, evaluate internal and external validity of research in terms of the extent to which it successfully avoids the many "threats" to validity.<sup>23</sup> The term of validity were going to develop by many experts in education because it sense to be an important thing to measure since the development of Education in the world.

The newest research during 21<sup>th</sup>century also prove the important of determining the content validity as one of the requirement in testing evaluation. In 2012, some researchers from University of Granada, Spain, provide a study about how to conduct a study aimed at obtaining content validity evidence in the test construction adaptation process. This research focused on the content validity analysis of the Spanish adaptation of the State-Trait Cheerfulness Inventory trait form (STCI – T). The data obtained showed that the process of obtaining content validity evidence leads to an

<sup>23</sup>Campbell, Donald T. Relabeling Internal and External Validity for Applied Social Scientist.

improvement of the items created both regarding the formal wording aspects and the theoretical representativeness – relevance of such items.<sup>24</sup>

In the other research, the validity based on test content was investigated by some researchers from University of Massachusetts Amherst in USA. They provided that for Educational tests and other assessment targeting knowledge and skill possessed by examinees, validity evidence based on test content is necessary for building a validity argument to support the use of a test for particular purpose.<sup>25</sup>

Fulcher conducts a research about the content validity in English for Academic Purposes (EAP). Drost also did a research about validity and focus on Social Science Research, and the paper has clarified the meaning of reliability of measurement and the general problem of validity in behavioral research. Those are the result that show and concern about a content validity which is included into a good test.

The contrary, Kaner proves that there is no simple formula or perceptions for generating "good" test cases and the space of interesting test is too complex for this. Against the statement from Kaner, research from McCroskey and Beatty, the research result show strongly supportive of the

<sup>24</sup>Rico, Elena Delgado and friends (2012). *Content validity evidence in test development: An Applied Perspective.* International Journal of Clinical and Health Psychology. Vol., 12, p. 449 – 460.

 $^{25}$ Sireci, Stephen and Bond, Molly Faulkner (2014). *Validity based on test content*. USA: Psicothema Vol, 26, No, 1, p. 100 – 107.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

content validity of the items employed in the PRCA – 24 (Personal Report of Communication Apprehension).

The other result of the research about the relationship between validity and the culture also against the statement from Cem Kaner. The researcher from Japan conducted a research about content validity and cultural sensitivity. They found that the instrument has similarities and differences compared with instruments of western origin. And they suggested that designers of evaluation instrument should consider the probability that the content validity of instruments for evaluating clinical teachers can be influenced by cultural aspects.<sup>26</sup>

In Indonesia, Yusrizal, the researcher from Universitas Negeri Medan (UNIMED) Medan, also had research about the Validity. He focused in studying about the test of construct validity using the factor analysis in instrument of lecturer performance evaluation. The result of his study showed that by using the construct validity, the instrument in evaluating the lecturer performance can be tested. From his research, both two instruments can be important information for the lecturers in increasing their performance in managing the class, lesson plan, teaching and learning process etc.<sup>27</sup> This research showed that there is a close relation between the

<sup>&</sup>lt;sup>26</sup>Kikukawa, Makoto and friends (2014). *An Instrument for evaluating clinical teaching in Japan: content validity and cultural sensitivity.* Japan: BMC Medical Education 14: 179.

<sup>&</sup>lt;sup>27</sup>Yusrizal, (2008). *Pengujian Validitas Kontruk dengan menggunakan analisis factor.* Medan: Jurnal Tabularasa PPS UNIMED

concept of validity in testing the lecturer performance and this concept also can be use in testing the test or evaluation for the students.

Those are some results from research which are conducted by some researchers, and most of them are talking about the relationship between the validity, especially the content validity and the performance of the testee. Thus, the researcher of this research analyzes the content validity of the placement test conducted in SMP N 5 Surabaya. And researcher focused to measure the validity of the test, whether this test agree with the curriculum from Educational Ministry which is applied in Indonesia or not.