

CHAPTER II

REVIEW OF RELATED LITERATURE

This chapter discusses some literature which is closely related to the questions of this research. This literature will focus on: testing and teaching, the purpose of the test, types of a test, characteristic of a good test, and review of the previous study. Thus, each of these main parts is divided into several sub part.

A. Testing and Teaching

In teaching learning activities, Both testing and teaching are so closely interrelated (or the relationship between them is as a partner).The results of teaching without evaluating or testing will be useless, because testing help to show the achievement of the objective of education. From the result of the test can be seen the teaching learning process is successful or not. Both testing and teaching are relationship with each other that it is impossible to work in either field without being constantly concerned with the other¹³. It is clear that relation between testing and teaching can't be ignored. Teachers, students, and school want to know their effort to achieve the educational objectives are

¹³J. B. Heaton, *Writing English Language Test.* (New York: Logman, 1988), p. 5

successful or not. They will be satisfied if their effort are successful. But if their effort unsuccessful so they will change their ways.

B. The purpose of test

Test is used to measure students' mastering with the subject given¹⁴. Some experts mention the other purpose of test. According to Nurkancana and Sumartana, a test has many purposes. First, is to know how far the result of a program applied whether it has reached its goal or not. Second, is to see whether the materials should be re-taught or not. Third, is to get some information about the students' weakness and difficulties in learning about the given materials. Fourth, is to determine the students' achievement and to allow them going through to the grade. Fifth, is to select and group students based on their achievement.

C. Based on the test maker

There are two kinds of a test based on the test maker. They are standardized test and teacher made test.

a. Standardized test

Standardized test is a test made by professional test makers, administered and scored in a consistent manner. Harris stated that standardized test is a test designed to be used with thousands and sometimes hundred thousands of subjects throughout the nation or the

¹⁴Nurkancana, WayandanSumartana, *Evaluasipendidikan*, (Surabaya:usahanasional, 1986) p.1

world, and prepared by team of testing specialists¹⁵. There are some examples of standardized test such as: Final Examination, TOEFL, IELTS, try-out test, etc. Tryout test is a test hold before UN. Try-out test is consists of two words those are “try” and “out” which each of them has the main purposes. “try” has two purposes. First, is to measure how far the students’ readiness about the matter that will be tasted in UN and second is to measure the test itself (it means that the test is not just anything taken but it must be constructed based on curriculum). “out” also has two purposes first is to encourage the students in facing UN and second is to make the students get used to face a situation like in UN.

b. Teacher made test

In teacher-made test, the test is made by the teacher himself of group of teachers without being tried on first, analysed, and revised. Classroom test are generally prepared, administered, and scored by one teacher. The teacher-made test is a test made by the teacher himself or group of teachers is using untried out, unanalysed, and unrevised test items¹⁶. The teacher-made test was used to measure his student’s achievement on the objectives given after finishing the teaching learning progress. Therefore, he also states that the teacher-made test has average or lower reliability than standardized test. *UTS (Ujian Tengah Semester)*

¹⁵David P Harris, *Testing English as a Second Language*, (New York: McGraw-Hill.Inc, 1969) p. 2

¹⁶Ibid, p. 13

or mid form test and *UAS (Ujian Akhir Semester)* or final form test are the examples of teacher-made test.

D. Kinds of tests

According to Arthur Hughes, there are four types of test. Such as : proficiency test, achievement test, diagnostic test, and placement tests¹⁷.

a. Proficiency tests

Proficiency tests are designed to measure people's ability in language, regardless any training they may have in that language¹⁸. The content of a proficiency test was not based on the content of the objectives of language courses. It's based on a specification of what candidates have to be able to do in the language in order to be considered proficient.

b. Achievement test

The content of the course in achievement tests are directly related to language course. The purpose is being able to establish how successful individual students, group of students and the course in achieving the objectives.

There are two kinds of achievement tests¹⁹:

a. Final achievement test

b. Progress achievement test

¹⁷ Arthur Hughes, *Testing for Language teachers*, p. 11

¹⁸ Arthur Hughes, *Testing for Language Teachers*, p.9

¹⁹ Arthur Hughes, *testing for language teachers*, p 12

Final achievement tests are those administered at the end of a course of study. They may be written and administered by ministries of education, official examining boards, or by members of teaching institutions. Clearly the content of these tests must be related to the courses with which they are concerned, but the nature of this relationship is a matter of disagreement among language tester.

Progress achievement tests, as their name suggests, are intended to measure the progress that students are making. They contribute to formative assessment. Since 'progress' is towards the achievement of course objectives, these tests, too, should relate to objectives. But how? One way of measuring progress would be repeatedly to administer final achievement tests, the (hopefully) increasing scores indicating the progress made. This is not really feasible, particularly in the early stages of a course. The low scores obtained would be discouraging to students and quite possibly to their teachers. The alternative is to establish a series of well defined short-term objectives. These should make a clear progression towards the final achievement test based on course objectives. Then if the syllabus and teaching are appropriate to these objectives, progress tests based on short-term objectives will fit well with what has been taught.

The content of a final achievement test should be based directly on a detailed course syllabus or on the books and other materials used. This has been referred to as the *syllabus-content approach*. If *Progress achievement tests*, as their name suggests, are intended to measure the progress that students are making. They contribute to formative assessment. Since progress is towards the achievement of course objectives, these tests, too, should relate to objectives.

c. Diagnostic test

Diagnostic tests are used to identify learners' strengths and weaknesses. They are intended primarily to ascertain what learning still needs to take place. At the level of broad language skills this is reasonably straightforward. We can be fairly confident of our ability to create tests that will tell us that someone is particularly weak in, say, speaking as opposed to reading in a language. Indeed existing proficiency tests may often prove adequate for this purpose.

Diagnostic test concerns with the student's persistent learning difficulties that are left unsolved by the standard corrective prescriptions of Formative test. In other word we can say that diagnostic test was a test of student learning difficulties during instruction. The primary aim of Diagnostic test was to determine the

causes of learning problems and to formulate a plan for remedial action²⁰.

d. Placement test

It's intended to provide information that will help to place students at the stage of the teaching programme most appropriate to their abilities. Typically it's used to assign students to classes at different level. One possible exception is placement tests designed for use by language schools, where the similarity of popular text books used in them means that the schools' teaching programmers also tend to resemble each other²¹.

The placement tests are most successful are those constructed for particular situations. They depend on the identification of the key features at different levels of teaching in the institution. They are tailor made rather than bought off the peg. This usually means that they have been produced 'in house'. The work that goes into their construction is rewarded by the saving in time and effort through accurate placement²².

E. Characteristic of a Good Test

A test is an important instrument in teaching learning process to measure students' mastery on the materials. To know the affectivities of a

²⁰ibid, page 16

²¹ibid, page 17

²² ibid

test, it has criteria for testing a test. According to Heaton, there are some criteria of good test; validity, reliability, administration and item analysis includes index of difficulty, index of discrimination, and the Distractors²³.

a. Validity

A test was classified to be valid if it measures accurately what it is intended to measure. According to Heaton, validity of a test is the extent to which it measures what it is supposed to measure and nothing else. There are four types of validity ; face validity, content validity, construct validity, and empirical validity²⁴.

1) Face validity

Hughes states: a test was said to have face validity if it looks as if it measures what it is supposed to measure. Face validity is not scientific notion and is not seen as providing evidence for construct validity, yet it can be very important.²⁵

The test has face validity if the test looks right to other tester, teachers, and moderator and test-takers. It means that face validity measured by subjective judgment.

Face validity will be high if the students or test takers encounter some or the entire characteristic of good face validity, as follow:

²³ J. B. Heaton, *Writing English Language Test*. (New York: Logman, 1975, p.159

²⁴ J. B. Heaton. *Writing English Language Test*. (New York: Logman, 1975), p. 159

²⁵ Arthur Hughes, *testing for language teacher*, (Cambridge: cambridge university press, 1989), p.33

- a. The test well-constructed and familiar format task,
- b. The test is doable within the allotted time limit,
- c. The items are clear
- d. The test have clear directions,
- e. The test related to the course work,
- f. A difficulty level that presents a reasonable challenge²⁶.

2) Content validity

Content validity is defined as any attempt to show that the content of the test is a representative sample from the domain that is to be tested²⁷.

Hughes states that a test is said to have content validity if its content constitutes a representative sample of the language skills, structures, etc²⁸.

b. Reliability

One of the necessary characteristic of good test is reliability. The test was said to be reliable if it is consistent in the measurements. It means that the students must have same mark if the test marked by two or more examiners. Moreover, the reliability of the test was considered a number of factors that may contribute to the unreliability of the test. According to Heaton, the factors affecting the reliability are:

²⁶H. Douglas Brown, *Language Assessment*..... p. 27

²⁷Glenn Fulcher, *Language testing and assesment*.(New York: Routledge, 2007)p.6

²⁸Arthur Hughes, *Testing for*..... p.26

- 1) The extent of the material selected for testing. Reliability is concerned with the size of the test; it is not too long and not too short.
- 2) The administration of the test²⁹.
The students or test-takers must have same condition and time limit.
- 3) The instruction. The clarity of the instruction will affect the students' comprehension to answer the test.
- 4) Personal factors, such as motivation and illness.
- 5) Scoring the test. It means that the objective test is more reliable than the subjective test.

There are some methods to estimate reliability. such as test – retest method, split half, equivalent method, and internal consistency method. Here, the reseacher uses split half method to get reliability because the test did only one times.

This formula is

$$r_{\frac{11}{12}} = \frac{N \sum Y_1 X_1 - (\sum X_1) (\sum Y_1)}{\sqrt{\{(N \sum X_1^2) - (\sum X_1)^2\} \{(N \sum Y_1^2) - (\sum Y_1)^2\}}}$$

After that the result above to corelation with sperman Brown pattern,
this formula is :

²⁹J.B. heaton. *Writing English*..... p.162

$$r_{11} = \frac{2 X \left(r_{\frac{11}{12}} \right)}{1 + \left(r_{\frac{11}{12}} \right)}$$

This is Criteria of reliability

0.00-0.20	Not reliable
0.20-0.40	Less Reliable
0.40-0.60	Reliable enough
0.60-0.80	Reliable
0.80-1.00	Very Reliable

c. Administration

A test is said to have administration if the test practicable. The test has appropriate time, the test has clear directions, the various steps in the administering of each sub- test being in simple language and clearly numbered, easy to score³⁰. It means that the test does not spend more time and money, the test does not spend require many types of equipment. It also has effeciant process of scoring.

³⁰ J. B. Heaton, *writing English langugae test*, p. 167

d. Item Analysis

The purpose of items analysis was to identify the test items whether it is good or not. To know the answer, all items should be identified from the index of difficulty, index discrimination and Distractors.

1. Index of difficulty

The good test items are not too easy and not too difficult. According to Heaton, index of difficulty was used to know how easy or difficult particular items in the test are. It is generally expressed as fraction or percentage of the students who answered the item correctly. To calculate the index of difficulty, Heaton uses the following formula³¹.

$$FV = \frac{R}{N}$$

FV = index of difficulty

R = number of students whose correct answer

N = number of students

It means that a good test to be given the students is the test with the criterion index of difficulty between 0,30 – 0,70. Meanwhile, the index of difficulty which shows 0,00 – 0,30 and 0,70 – 1,00 was not good to be given to the students because the test is either too difficult or too easy for them.

³¹J. B. Heaton. *Writing English language test*, p. 178

2. Index of discrimination

Index of discrimination indicates the extent to which the items discriminate between the students. It is to discriminate the students who have high ability on the test and the students who have low ability on the test³². Heaton's formula to calculate index of discrimination is:

$$D = \frac{\text{Correct}U - \text{Correct}L}{n}$$

D = index of discrimination

Correct U = the number of students in upper group who answer the item correctly

Correct L = the number of students in lower group who answer the item correctly

N = number of candidate of one group

Arikunto classifies the criteria of index of discrimination as follows.³³

D: 0,00-0,20 = poor

D: 0,20- 0,40 = satisfactory

D: 0,40- 0,70 = good

D: 0,70- 1,00 = excellent

The range index of discrimination according to heaton as follows.

+1 = an item which discrimination perfectly

³²Ibid, p.180

³³Suharsini Arikunto. *Dsar- dasar Evaluasi*p. 223

0 = an item which does not discriminate in any way at all

-1 = an item which discriminates in entirely the wrong way.

3. The distractors

Analysing the distractors aimed not only to know which items that cannot work properly, but also to check why particular test taker failed to answer certain items correctly. Distractors can function well if these are chosen by students from the lower level. Arikunto states the distractor is chosen at least by 5% of students who taking the test is called good test.

No item	Option	Upper	Lower	comment
1	A*			
	B			
	C			
	D			
	0			

The analysis of distractor is done by comparing the number of students in upper group with students in lower group based on their answer. In addition, the good distractor will manipulate more students in lower group than students in upper group. Moreover, if there are more able students chosen the distractors, it means that the item does not function as expected, and it must be revised.

F. Review of previous study

Some reseaches with similar topic has analized the quality of the test. First is the reseacrh conducted by abidatul khoiro. This research analyzed teacher made English try out test for national examination 2010-2011 for the third graders of MAN Sidoarjo. The research analyzed content validity, the index of difficulty, and the index of discrimination of the teacher made English try-out test in national examination 2010 – 2011 for the third graders of MAN Sidoarjo. The result shows that the content validity of the teacher-made English try-out test of MAN Sidoarjo has good content validity since 52% items test covered the indicators of Standard of Graduates Competencies. The test has acceptable index of difficulty because the Science class have 60% items which are adequate items and Social class have 68% items which are adequate items. And the index of discrimination of the test was different between both of class. The result of Science class shows that 44% items can be used. It means that the test is unacceptable for the Science class. And the Social class has acceptable index of discrimination since 60% items has satisfactory and good criteria³⁴.

The thesis was done by Moch. Solihan (UNESA, 2000), in his thesis, he analysed the teacher made English test items in formative semester I

³⁴Abidatul khoiro, *an analyzed teacher made English try out fr national exam for the third graders of MAN Sidoarjo*, Thesis S1, (Surabaya: perpustakaan IAIN,2012)

2008/2009 of the first year students of SMAN 22 Surabaya. In his thesis, he found that the test had high content validity and adequate reliability but it did not acceptable index of difficulty, and the index of discrimination in multiple choice test was not adequate³⁵.

Iffah Mursyidah Mayangsari conducted the research in 2009. The research analysed teacher-made formative English test in SMA 2 Muhammadiyah Sidoarjo. The research focused on the content validity, reliability, item difficulty and item discrimination. This research used descriptive research as design in the study. The result of the analysis concluded that the test has high content validity, adequate reliability, acceptable item difficulty and acceptable item discrimination³⁶.

The thesis was done by Anita Nur Rahma. This research analyzed the English summative test items for semester 1 2002-2003 of the first year students of SMU in Surabaya. The research analyzed objective test. It focused on the reliability, index of difficulty, index of discrimination and item distracter of the test. The result of the research concluded that the test have

³⁵ Moch. Solihan, *analyzed the teacher made English test items in formative semester I 2008/2009 of the first year students of SMAN 22 Surabaya*, Thesis S1, (Surabaya: perpustakaan UNESA, 2000).

³⁶ Iffah mursyidah mayangsari, *an analyzed teacher made formative English in SMA 2 muhammadiyah sidoarjo*, Thesis S1, (surabaya : perpustakaan IAIN, 2009).

adequate reliability, poor index of discrimination and bad distracters. This test also categorized as easy test³⁷.

The thesis was done by Sigit Eko Saputro who analyzed the teacher-made English final test items for semester 1 2005-2006 of the first year students of SMA 2 Magetan. The research analyzed both multiple choice and completion test. It focused on the content validity, reliability, index of difficulty, index of discrimination and item distracter of the test. The result of the research concluded that the test have high content validity, adequate reliability, acceptable index of difficulty, did not have adequate index of discrimination and effective distracters³⁸.

³⁷Anita Nur Rahma, *'an analyzed the English summative test items for semester 1 2002-2003 of the first year students of SMU in Surabaya'*, Thesis S1, (Surabaya: perpustakaan UNESA, 2004).

³⁸Sigit Eko Saputro, *'an analyzed the teacher-made test items in UAS semester 1 2005-2006 of the first year students of SMA 2 Magetan'*, Thesis S1, (Surabaya: perpustakaan UNESA, 2006).