#### **CHAPTER II**

#### REVIEW OF RELATED LITERATURE

This chapter explains some literature which is related to the questions of this research. This literature will focus on: theoretical foundation and previous and review study.

# A. Testing

#### 1. Definition of Test

Test is set techniques, procedures, and items that constitute an instrument of some sort that require performance or activity on the part of the test taker (and sometimes on the part of the tester as well). Test is procedures designed to elicit certain behavior from which one can make inferences about certain characteristics of an individual.

In line of that, test as quoted from Webster's Collegiate by Daryanto, is any series of questions or exercise or other means of measuring the skill, knowledge, intelligence, capacities of aptitudes or an individual or group.<sup>3</sup>

<sup>&</sup>lt;sup>1</sup> Brown, H. Douglas, *Teaching by Principles: An Interactive Approach to Language Pedagogy, Second Edition.*(San Francisco: Longman Inc.) 2001. Pg 334

<sup>&</sup>lt;sup>2</sup> Bachman, Lyle F. Fundamental Considerations in Language Testing. (USA: Oxford University Press)

<sup>&</sup>lt;sup>3</sup> Daryanto, Drs. H. Evaluasi Pendidikan, (Jakarta, Rineka Cipta, 1999). Pg 35

In the other word, Kubizyn and Borich stated in their book, that test is just as tools that can contribute importantly to the process of evaluating pupils, the curriculum, and the teaching method.<sup>4</sup>

### 2. Testing and Teaching

The effect of testing on teaching learning is known as backwash, and can be harmful or beneficial. If a test is regarded as important, if the stakes are high, preparation for it can come to dominate all teaching and learning activities. And if the test content and testing techniques are at variance with the objective of the course, there is likely to be harmful backwash. An instance of this would be where students are following an English course that is meant to train them in language skills (including writing) necessary for university study in an English speaking country, but where the language test that they have to take in order to be admitted to a university does not test those skills directly. If the skill of the writing, for the example, is tested by multiple choice items, then there is great pressure to practice such items rather that practice the skills of writing itself. This is clearly undesirable.<sup>5</sup>

\_

<sup>&</sup>lt;sup>4</sup> Tom Kubiszyn and Gary Borich, *Educational Testing and Measurement* (Singapore, John Wiley & Sons, INC, 2003), Pg 1

<sup>&</sup>lt;sup>5</sup> Arthur Hughes. Testing for Language Teachers second edition. pg 1

### 3. The Effects of Testing on Teaching and Learning

Testing has assumed a prominent role in recent efforts to improve the quality of education. Viewing standardized tests as a significant, positive and cost-effective reform tool, educational policymaker has been using them at an increasing rate. The testing process now costs hundreds of millions of dollars and thousands of hours of administrative, teacher and student time.

The reasons for the increased use of testing are many. Following advice from testing advocates, policymakers believe that testing sets meaningful standards to which school systems, schools, teachers, and students can aspire; that test data can help shape instruction; that is serves important accountability purposes; and that coupled with effective incentives or sanctions, testing is powerful engine of change. As an evidence of the matter, proponents point with pride to rising test scores.<sup>6</sup>

# 4. Focus on assessment, not on tests

But deciding which of these test types is better or more appropriate is not easy. Knowing that each uses a unique format to provide different kinds of information does not bring us much closer to selecting one or the other alternative. Indeed, attempting to select the most appropriate among available testing alternatives on the basis of their characteristics alone would be like

<sup>6</sup> Joan Herman, Jean Dreyfus and Sharin Golan, The Effect of Testing on Teaching Learning, Pg 1-2

trying to choose between a hammer, a shovel, or a screwdriver based entirely on what these tools look like. Obviously, to select the appropriate tool, we first need to have an idea about the job to be accomplished. Alone, language tests tell us little about the jobs to be accomplished in language programs and classrooms. We cannot distinguish between good or bad, appropriate or inappropriate, reliable or not reliable, valid or not valid tests based solely on characteristics of the test instruments and procedures. Rather, we must focus instead on *language assessment*.

Language assessment is the process of *using* language tests to accomplish particular jobs in language classrooms and programs. In language assessment, we first gather information in a systematic way with the help of language testing tools. For example, we may use an oral interview to gather information about students' speaking abilities, then make interpretations based on that information. Or, we may make interpretations about students' abilities to perform a range of real-world speaking tasks based on how well students perform in the oral interview. Finally, based on these interpretations, we make a decision or take action within the classroom or program. We may decide that our students need more work on oral fluency and that we should therefore devote more class time to fluency-oriented activities.<sup>7</sup>

-

<sup>&</sup>lt;sup>7</sup> John M. Norris, *Purposeful Language Assessment: Selecting the Right Alternative Test*, pg 2

#### 5. Standards in testing

One area of increasing concern in language testing has been that of standards. The word 'standards' has various meanings in the literature, as the Task Force on Language Testing Standards set up by ILTA discovered One common meaning used by respondents to the ILTA survey was that of procedures for ensuring quality, standards to be upheld or adhered to, as in codes of practice. A second meaning was that of levels of proficiency - what standard have you reached? A related, third meaning was that contained in the phrase 'standardized test', which typically means a test whose difficulty level is known, which has been adequately piloted and analyzed, the results of which can be compared with those of a worming population: standardized tests are typically norm referenced tests. In the latter context 'standards' is equivalent to 'norms'.

In recent years, language testing has sought to establish standards in the first sense (codes of practice) and to investigate whether tests are developed following appropriate professional procedures. Groot argues that the standardization of procedures for test construction and validation is crucial to the comparability and exchangeability of test results across different education settings. Alderson and Buck and Alderson *et al.* describe widely accepted procedures for test development and report on a survey of the practice of British EFL examining boards. The results showed that current (in

the early 1990s) practice was wanting. Practice and procedures among boards varied greatly, yet (unpublished) information was available which could have attested to the quality of examinations. Exam boards appeared not to feel obliged to follow or indeed to understand accepted procedures, nor did they appear to be accountable to the public for the quality of the tests they produced. Fulcher and Bamford (1996) argue that testing bodies in the USA conduct and report reliability and validity studies partly because of a legal requirement to ensure that all tests meet technical standards. They conclude that British examination boards should be subject to similar pressures of litigation on the grounds that their tests are unreliable, invalid or biased. In the German context, Kieweg (1999) makes a plea for common standards in examining EFL, claiming that within schools there is little or no discussion of appropriate methods of testing or of procedures for ensuring the quality of language tests.<sup>8</sup>

### 6. The purpose of test

Test is used to measure students' mastering with the subject given<sup>9</sup>. Some experts mention the other purpose of test. According to Nurkanca and Sumartana, a test has many purposes. First, is to know how far the result of a programmer applied whether it has reached its goal or not. Second, is to see

<sup>8</sup> J Charles Alderson and Jayanti Banerjee, *Language testing and assessment (Part I)*, (United Kingdom: Cambridge University), 2001. Pg 218-219

<sup>9</sup> Nurkancana, Wayan dan Sumartana, Evaluasi pendidikan, (Surabaya:usaha nasional, 1986) pg 1

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.

David conducted six objectives of language testing: 10

- 1. To determine readiness for instructional programs.
- 2. To classify or place individuals in appropriate language classes.
- 3. To diagnose the individual's specific strengths and weaknesses.
- 4. To measure aptitude for learning.
- 5. To measure the extent of student achievement of the instructional goals.
- 6. To evaluate the effectiveness of instruction.

-

 $<sup>^{\</sup>rm 10}$  David P Harris, Testing English as a second Language, (New York: Mc Craw-ill, Inc, 1959), Pg 2

# B. Type & Kind of Test

# 1. Types of The Test

### a. Based on the Number of test-taker

Based on the number of the test taker, test is divided into:<sup>11</sup>

- Individual test: refers to a test where the tester tests only one testee,
   while
- Group test: refers to a test where the taster faces more than one testee.

### b. Based on the test maker

Test could be determined into teacher-made test and standard test.

Although both are have the same purpose to measure the progress of teaching learning process, however they differ each other.

### • Teacher-made Test

Teacher made test is the test that is made the teacher of that classroom/or course itself. This test purposes to know measure how far the students achieve the instructional aim of particular lesson or course that are taught in the classroom.

<sup>&</sup>lt;sup>11</sup> Prof. Drs. Anas Sudijono, Pengantar Evaluasi Pendidikan. (Jakarta: PT. Raja Grafindo Persada, 1996), Pg. 74

Teacher made tests are designed based on the particular aim and description of the lesson that are taught in that class. Generally, this kind of test is not tried out before and even revised after. Thus, the validity of teacher made test often considered poor. <sup>12</sup>

### Standardize Test

Standard test is a test constructed by test construction specialist, usually with the assistance of curriculum experts, teachers, and school administrators.<sup>13</sup>

Unlike teacher made test, standard tests are tried out before used. The tried out is not aimed to measure student's ability but to know whether the items test is adequate. The result of the tried out then analyzed through item analysis to get the coefficient of index difficulty and also index of discrimination. After that, the item that too difficult or too easy is should be revised as well as the item that has poor index of discrimination should be replace.<sup>14</sup>

Sons, INC, 2003), Pg 343

Burhan Nurgiyantoro, *Penilaian dalam Pengajaran Bahasa dan Sastra*, (Yogyakarta, BPFE-Yogyakarta, 2001), Pg 60
 Tom Kubiszyn and Gary Borich, Educational Testing and Measurement (Singapore, John Wiley &

<sup>&</sup>lt;sup>14</sup> Millatul Islamiyah, Content Validity and Item Analysis of Semester II English Final Test for Tenth Gr ade Students of SMAN 3 Sidoarjo, Thesis (Surabaya: Perpustakaan IAIN Sunan Ampel Surabaya, 20 10).

In addition, standard test is administered and scored according to specific and uniform. Thus it can be used in all and different school even it can used many times. 15 In the other word, a standard test administered and scored in Surabaya, would be administered and score in exactly the same manner in Malang, Sidoarjo, or anywhere in Indonesia.

#### 2. Kinds of tests

Acording Arthur hughes, there are four types of test. Such as : profeciency test, achievment test, diagnostic test, and placement tests<sup>16</sup>.

### a. Proficiency tests

Proficiency tests are designed to measure people's ability in language, regardless any training they may have in that language 17. 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.

#### **Achievment test**

The content of the course in achievement tests are directly related to language course. The purpose is being able to establish how

 $<sup>^{15}</sup>$  Tom Kubiszyn and Gary Borich, *Educational Testing and Measurement*, pg 343  $^{16}$  Arthur hughes, *Testing for Language teachers*, pg. 11

<sup>&</sup>lt;sup>17</sup>Athur Hughes, *Testing for Language Teachers*, pg.9

successful individual students, group of students and the course in achieving the objectives.

There are two kinds of achievement tests<sup>18</sup>:

#### a. Final achievement test

# b. Progress achievement test

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. Diagnotice test

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

 $^{18}$  Athur Hughes, Testing for Language Teachers , pg  $\,12$ 

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

causes of learning problems and to formulate a plan for remedial action<sup>19</sup>.

#### 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 other<sup>20</sup>.

#### C. Forms of Test

There are two kinds of form of test: objective and subjective test. The distinction between both tests is concern on method of scoring, and nothing else.<sup>21</sup> The following explanation will clarify enough about them.

### 1. Objective Test

Sudijono claimed that objective test is one type of test that is created using items tests, then what the entire test taker has to do is just answering the question by choosing one among several probably answers

 $<sup>^{19}</sup>$  Athur Hughes, Testing for Language Teachers, pg 16  $^{20}$  Athur Hughes, Testing for Language Teachers , pg 17

<sup>&</sup>lt;sup>21</sup> Athur Hughes, Testing for Language Teachers, Pg 22

available in each items or writing sentences or particular symbols in place provided in each item test.<sup>22</sup>

In line of that, objectives test as cited from Lado is:

"Objectives test are those that are scored rather than mechanically without need to evaluate complex performance on scale" 23

# a. Types of an objectives test

Sudijono also added that there are five types of objectives test including: true or false test, matching test, completion test, fill in test and also multiple choices. However in this thesis only will clarify the last one.

Multiple choices as stated by Sudijono are a test which is created likely incomplete sentences and the testee should complete the sentence in order to answer the question. <sup>24</sup>Before going to design multiple choice test, the test maker or in this case is teacher should know primarily several terms used in multiple choices. First is *stem* which refers to initial part of each multiple choice items. Second is option/responses/alternatives, refers to the options which are available

<sup>&</sup>lt;sup>22</sup> Prof. Drs. Anas Sudijono, *Pengantar Evaluasi Pendidikan*, (Jakarta: PT. Rayagrafindo Pustaka, 1990), pg 106

<sup>&</sup>lt;sup>23</sup> Robert Lado, *Language Testing*, (London: Longman Group, 1961), Pg 28

<sup>&</sup>lt;sup>24</sup> Prof. Drs. Anas Sudijono, *Pengantar evaluasi Pendidikan*, Pg 106

for student to select their answer. One option among them is called the correct answer and the other is distractors<sup>25</sup>

The illustration from the explanation above as follows:<sup>26</sup> Stay here until Mr. Short.....you to come. = *stem* 

D. tells = correct answer

- b. The Benefits and Weakness of Objective Tests
  - 1. The Benefits of Objective Test

When objective test of language are properly made, they have important values. Arikunto mentioned several goodness of objective test:<sup>27</sup>

- Represent more all objective materials that are being tested
- They can test in short time
- They can be scored with speed and ease
- They use careful objective score in evaluating the test.

 <sup>&</sup>lt;sup>25</sup>, J. B. Heaton, Writing English Language Tests (New York: Longman Group, 1988). Page 28
 <sup>26</sup> J. B. Heaton, Writing English Language Tests, pg 28

<sup>&</sup>lt;sup>27</sup> Suharsimi Arikunto, *Dasar-dasar Evaluasi Pendidikan*, (Jakarta, Bumi Aksara, 1993), Page 164

They able to be scored not only by teacher or test maker.

### 2. The weakness of Objective Test

The usual objections to objective test are mentioned by Lado as follows:<sup>28</sup>

- They are too simple
- They do not require real thinking but simply memory
- They do not test the ability of the students to organize his thoughts.

Beside the three previous objective test's weakness, Arikunto also added the rest objections<sup>29</sup>:

- Objective test enable student being speculative in responding the question in a test
- Open widely possibilities for students to cheating each other in doing a test
- It is more difficult to construct the objectives test than subjective test because it contains a lot of item tests.

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

 $<sup>^{28}</sup>$ Robert Lado, *Language Testing*, (London: Longman Group, 1961), Page 35  $^{29}$  Suharsimi Arikunto, *Dasar-dasar Evaluasi Pendidikan*, Pg 164

# 2. Subjective Test

As quoted from Lado, subjective test is:

"Tests that require an opinion and a judgment on the part of the examiner".

In the other word, Nurgiyantoro have said that subjective test is a test that require student to answer in essay using their word.<sup>31</sup>

# a. Scoring an essay test

Scoring an essay test generally based on the weight of each item test, the level of difficulty, and the amount of the element contained by the answer which is considered as the rightes answer.

For example, there are 5 items test in essay test. The tester had determined that all items have the same level of difficulty, and the elements in each item had made in the same amount. Based on that, tester decided that testee who could answer with the rightest answer or which the answer provides the entire element that required by the tester within the item test, will get 10 marks. When the testee answer almost perfectly or the answer provide mostly the element that required by the test taker, will get 9 mark, and so on.<sup>32</sup>

<sup>&</sup>lt;sup>30</sup> Robert Lado, *Language Testing*, (London: Longman Group, 1961), Pg 28

<sup>&</sup>lt;sup>31</sup> Burhan Nurgiyantoro, *Penilaian dalam Pengajaran Bahasa dan Sastra*, (Yogyakarta, BPFE Yogyakarta, 2001), Pg 71

<sup>&</sup>lt;sup>32</sup> Prof. Drs. Anas Sudijono, *Pengantar Evaluasi Pendidikan*, (Jakarta, PT. Raja Grafindo persada, 1996), Pg 301

### b. The Benefit and the weakness of subjective test

The characteristics of subjective can be seen from its benefits and weakness as follow

- a) The Benefit
  - Subjective test can create easily and fast
  - Avoid students being speculative in answering the items test
  - The test taker is able to know how far students understand the material
  - Motivate student to organize their thoughts

# b) The weakness

- Less able to represent all materials
- It is difficult to score the subjective test. It because the answer of each item might be varieties and wide. Thus, it needs a lot of time, and thoughts to score it.
- Enable test taker to score subjectively
- Validity and reliability of subjective test is poor.

#### D. 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 test,

it has criteria for testing a test. According to Arikunto, there are some criteria of good test; validity, reliability, objectivity, practicality, economy<sup>33</sup>.

### a. Validity

A test was classified to be valid if it measures accuracy what it is intended to measure. According to Heaton, validity of a test is the extent to which it measure what it is supposed to measure and nothing else. There are four types of validity; face validity, content validity, contruct validity, and emperical validity<sup>34</sup>.

#### 1) Face validity

Face validity refers to researchers' subjective assessments of the presentation and relevance of the measuring instrument as to whether the items in the instrument appear to be relevant, reasonable, unambiguous and clear.<sup>35</sup> 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.<sup>36</sup>

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.

.

<sup>&</sup>lt;sup>33</sup>H. Douglas Brown., *language Assessment: principles and classroom practices*, (New York: Pearson Education, 2004), pg.3

<sup>&</sup>lt;sup>34</sup>J. B. Heaton. Writing English Language Test. (New York: Logman, 1975). pg. 159

<sup>&</sup>lt;sup>35</sup> Ayodele James, Oluwatayo, *Validity and Reliability Issues in Educational Research (Vol 2)*, (Nigeria: Institute of Education, Ekiti State University, 2012) pg. 392

<sup>&</sup>lt;sup>36</sup>Athur Hughes, testing for language teacher, pg.33.

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

- 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 challenge<sup>37</sup>.

### 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 tested<sup>38</sup>.

Hughes states that a test is said to have content validity if its content constitutes a representative sample of the language skills, structures, etc<sup>39</sup>.

### 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

<sup>39</sup>Athur Hughes, Testing for Language Teachers, pg. 26

<sup>&</sup>lt;sup>37</sup>H. Douglas Brown, language Assessment: principles and classroom practices, pg. 27

<sup>&</sup>lt;sup>38</sup>Glenn Fulcher, *Language testing and assesment*. (New York: Routledge, 2007), pg. 6

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:

- 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.
- The administration of the test<sup>40</sup>.
   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 researcher uses split half method to get reliability because the test did only one times.

This formula is

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

<sup>&</sup>lt;sup>40</sup>J.B. heaton. Writing English Language Test pg.162

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

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

This is Criteri reliable

0.00-0.20	Not reliable
0.20-0.40	Less Reliable
0.40-0.60	Reliable enough
0.60-0.80	R <mark>eli</mark> able
0.80-1.00	Very Reliable

# c. Objectivitas

According to arikunto the test is called objective if it is free from subjective factors which influence the test<sup>41</sup>. Objectivity of a test can be increased by using more objective types test items and the answers are scored according to model answers provided. Arikunto adds that there are

<sup>&</sup>lt;sup>41</sup>Suharsimi arikunto, dasar – dasar evaluasi pendidikan, pg. 59

two factors that influence the objectivity of a test they are the form of a test and the test scorer.

### d. Practicality

A test is called as practical test if it is easy to do and does not require many equipments and give freedom to the students to do the easier part, easy to score, is completed with clear instructions<sup>42</sup>. Arikunto stated that practicality of a test deals with a level of difficulties in admintering the test it self.

### e. Economy

According to Arikunto the economices in a test related with the amount money, time and energy that a test taker spends to take a test<sup>43</sup>. it means that the test doesn't need expensive fee, a long time and extra energy to finish the test.

# f. Item Analysis

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

### a) 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

<sup>&</sup>lt;sup>42</sup>Suharsimi arikunto, dasar – dasar evaluasi pendidikan, pg. 61

<sup>&</sup>lt;sup>43</sup>Suharsimi arikunto, *dasar – dasar evaluasi pendidikan*, pg. 61

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 formula<sup>44</sup>.

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

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.

### b) 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 test<sup>45</sup>. Heaton's formula to calculate inex of discrimination is:

<sup>44</sup>J. B. Heaton. Writing English Language Test p. 178

<sup>&</sup>lt;sup>45</sup>J. B. Heaton. Writing English Language Test, p.180

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

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. 46

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

0 = an ite which does not discrimination in any way at all

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

#### c) The distractors

Analyzing the distractors aimed not only to know which items that cannot work properly, but also to check why particular test taker

<sup>&</sup>lt;sup>46</sup>Suharsini Arikunto. *Dsar- dasar Evaluasi Pendidikan*, pg. 223

failed to answer certain items correctly. Distractors can function well if these are chosen by students from the lower level. Arikunto state the distractor is chosen at least by 5% students who taking the test is called good test.

No item	Option	Upper	lower	comment
1	A*			
	В			
	C			
	D			
	0			

When we want to analyze a test, we should know about the criteria of a good test itself. Based on arikunto, the criteria a good test, there are five criteria good a test: validity, reliability, objectivity, practically, and economy. Because My research concern on the test of multiple choice, I only use two criteria. Those are validity and items analysis include index difficulty, index discrimination, and distractors.

# E. Previous Study

Some research with similar topic has analized the quality of the test. First is the research 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 teachermade 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 criteria<sup>47</sup>.

Second research was conducted by Milatul Islamiyah. This research analyzed the content validity and item analysis of English final test at last semester for tenth grade students of SMAN 3 Sidoarjo by. She finds that SMAN 3 Sidoarjo has good content validity of the test and acceptable index

<sup>&</sup>lt;sup>47</sup>Abidatul khoiro, an analyzed teacher made English try out fr national exam for the third graders of MAN Sidoarjo, Thesis S1, (Surabaya: perpustakaan IAIN,2012)

difficulty of the test. Her research design is descriptive research and quantitative approach to collect numerical calculation data.<sup>48</sup>

Iffah Mursyidah Mayangsari conducted the research in 2009. The research analyzed 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 discrimination<sup>49</sup>.

From those previous studies above, the researcher prove that this research was different with previous study that showed above, the researcher do this research in KBRI school that located in Kuala Lumpur Malaysia, and the researcher focus on what the language testing technique that used in KBRI school and how the quality of the English testing that school. The researcher still do not know what exactly technique of testing that used by KBRI school Malaysia in this case Sekolah Indonesia Kuala Lumpur

\_

<sup>&</sup>lt;sup>48</sup> Millatul Islamiyah, Content Validity and Item Analysis of Semester II English Final Test for Tenth Gr ade Students of SMAN 3 Sidoarjo, Thesis (Surabaya: Perpustakaan IAIN Sunan Ampel Surabaya, 20 10)

<sup>&</sup>lt;sup>49</sup> Iffah Mursyidah Mayangsari, analyzed teacher-made formative English test in SMA 2 Muhammadiyah Sidoarjo, Thesis S1, (Surabaya: perpustakaan IAIN, 2009)