# CHAPTER II

# REVIEW OF RELATED LITERATURE

## 2.1 Testing and Teaching

Testing is a crucial part of every teaching and learning experience. Heaton (1988:5) in his book stated that "Both testing and teaching are so closely interrelated that it is virtually impossible to work in other field without being constantly concerned with the other." It can be informed that testing and teaching can be not separated in achieving course.

In teaching process, testing used to measure how far students' achievement and learning purposes have been done. Sometimes, there may be a situation which teaching process possibly good, but students' result of doing a test give output badly and may be vice versa. For this reason, as a measurement tool, a test must be designed in an appropriate test that fulfills its criteria that expected able to give information that can be accounted for its truth. It means the measuring tool is able to provide information about the students in accordance based on their actual situation. It is very important because the information will be used to consider then decide various cases both with regard to the students and general teaching activities.

## 2.2 Definition of Test

As widely known, test in general is a device used for measure students' abilities in particular course. According to Arikunto (2013:67) test is tool or procedure used to measure something in prescribed manner and rules. In addition,

according to Wahyuni and Ibrahim (2012:11) a test can be informed that it may be both of technique and tool in evaluation, hence to conduct a test needs a set of task or question; thus it known by test instrument.

Moreover, the term as like as Brown's argued "A test, in simple terms, is a *method of measuring a person's ability, knowledge or performance in a given domain.*" Brown (2013:3). It can be said that test is a way of measuring students' achievement.

Besides test, the term of assessment is widely and sometimes known as a test in current education. It is thought that assessment have same meaning with test. In fact, they are different each other. Brown (2003:4) defined "Assessment… is an ongoing process that encompasses a much wider domain." In contrast, "Tests are prepared administrative procedures that occur at identifiable times in a curriculum when learners muster all their faculties to offer peak performance, knowing that their responses are being measured and evaluated." It can be inferred that assessment has wider scope than test. Additionally, a test belongs to assessment set.

Moreover, there are another terms related to the test; they are measurement and evaluation. Measurement is a process to giving value for students based on their achievement in a learning experience. Measurement will answer toward question: *how much*. Then, result of measurement is in quantitative form that is numerical. It can be informed that measurement is a scoring process to the test-takers in numerical form.

Furthermore, evaluation is an activity when you identify what the planned program has been achieved or no. It conducted to describe some information how far result of the teaching and learning process achieved; what the purpose is achieved or no. Then you decide what will you do based on the identifying result given (Wahyuni & Ibrahim, 2012:3). In addition, evaluation is a way to get feedback that can be obtained from students about a course had been taught by their teacher (Foyewa, 2015: 32)

In summary, the four terms can be not separated each other. To do evaluation in teaching and learning process, it must be based on the assessment first; assessment itself can use measurement; and it is test in this discussion. So test just a little element of measurement, assessment, and evaluation.

## 2.3 Test Types

### 2.3.1 Based on the Purpose

The first way to design a test, you must determine the aim for the test. Based on the purpose, Heaton (1988:171) divided test into four types, they are achievement tests, proficiency tests, aptitude tests, and diagnostic tests. Moreover, Brown (2013:43) added it one; placement tests. So, there are five types of tests in this discussion.

### a. Achievement Test

An achievement test is related directly to language course in classroom (Brown, 2013:47). It designed to describe a progress that had been done by students; whether they had learned (income) with regard to stated course outcome. However, the test is usually conducted at the end of a unit or term of study; mid-

and end- point of the semester or academic year. It means that the test is usually developed by schools in particular period.

**b. Proficiency Test**

A proficiency test; in contrast, is not based on the particular curriculum or language course. This test aims to test global competence to measure all of the language skills. Coombe & Hubley (2007:11) said "They are designed to assess the overall language ability of students at varying levels." They are also may tell us how accomplish the students is in a particular skill language.

Conversely, a teacher perhaps not has many opportunities to creates the tests in classroom like the achievement tests, because the proficiency tests are usually developed by external bodies such as quoted on Coombe & Hubley (2007:11) "…examination boards like Educational Testing Services (ETS) or Cambridge ESOL." For the reason a typical example of the test is Test of English as a Foreign Language (TOEFL). The TOEFL contains of sections on listening comprehension, structure (grammatical accuracy), reading comprehension, and written expression (Brown, 2013:45).

**c. Language Aptitude Test**

Brown informed that an aptitude test can predict a person's success prior to exposure to the second language. Brown (2003:43) said "**A Language Aptitude Test** is designed to measures capacity or general ability to learn foreign language and ultimate success in that undertaking." Moreover, Heaton (1988:173) said in his book "Language learning aptitude is complex matter, consisting of such factors as intelligence, age, motivation, memory, phonological sensitivity

and sensitivity to grammatical patterning." It can be concluded that the tests is designed to measure an individual's aptitude for learning a foreign language.

Additionally, types of the test are similar to proficiency test; that is not administered by any school. Modern Language Aptitude Test (MLAT) and Pimsleur Language Aptitude Battery (PLAB) are two standardized aptitude tests that have been used in the United States (Brown, 2013:43). Both of them require test-takers to do a set task related to language.

**d.  Diagnostic Test**

As the previous tests have each purpose; a diagnostic test also, but is different with other types, where other tests used to find out students' proficiency. These types of test used to identify weaknesses of students (Rani, Priyadarsaini, & Rao, 2007:68). It means a diagnostic test is designed to diagnose students' difficulties of language learning. For the reason, students need further help to master the failure course. For example, a student may be diagnosed grammatical features of English writing that is so hard may be for the student and should become part of remedial action then. In addition, the information is crucial for next materials.

**e.  Placement Test**

A placement test is created to measure students' abilities of a language curriculum or school in order to place them in appropriate grade or class (Brown, 2013:45). The test will answer toward a question: what student's level is? Typically, the tests are most widely used when students will get in school. The

main purpose of the test is to create group which have similar level. In addition, the tests are frequently used to diagnose test-takers.

**2.3.2 Based on Scoring Manner**

Considering that types of test are wide discussion, in this topic the researcher defines some forms of test. There are two types of test based on scoring manner; they are subjective and objective test. Those two types of test can be used to measure almost any educational achievement. Heaton (1998:25) informed that those subjective and objective tests are terms used to score test. The following explanation will clarify about them:

a.   **Subjective Tests**

A subjective test is one that requires scoring by students' description (Coombe & Hubley, 2007:11). The test needs students' opinion. The students have to explain their idea by their own words; they may come in form of essay letter writing. However, students need more time to do it. In addition, in this type of test, person's thinking is crucial element.

Talking about subjective test, there are various kinds of questions in this types, they are, (1) comparing, (2) defensing idea, (3) causality, (4) explaining, (5) abbreviating, (6) analyzing, (7) describing, (8) giving assessment of an idea, (9) formulating a problem, (10) summarizing, etc. (Wahyuni & Ibrahim, 2012:13).

Furthermore, according to Arifin (2012:137-139) the test divides into two forms as follow:

(i)      *Restricted Response Items:* To answer this question, students need express particular case as boundaries. Although students' responses are

different each other, there should be important point in the systematic answer that refer to the limits that had been determined and desired in the question. So that why, this form of description has a set of certain answer relatively, so that scoring can be objectively. In addition, value of a correct answer is 1 and a wrong answer is 0.

(ii)    *Extended Response Item:* In this question, students need to answer the question freely based on their abilities. There is no limit of method and systematic to answer. That is why students have several systematics to respond the question. So, scoring in this tests are spelled out in ranges; highest score is determined by the complexity of students' responses, such as 0-2, 0-4, 0-6, 0-8, 0-10 etc. Minimum score should be 0; it is for students who do not give their responses. While maximum score is determined by particular appropriateness of their answer required in the test.

b.  **Objective Test**

Generally, an objective test is a test requires scored by comparing students' answer with the correct option (answer key) (Coombe & Hubley, 2007:11). An objective test has several alternatives; one of them is the correct answer that should be chosen by the test-takers. In addition, this test called as objective because its assessment is objective; the result will be same since the tests have definite response whoever its checker (Arifin, 2012:153).

Since objective test look like a simple test in answering, the kinds of this test has several strength; the test can be used to assess lessons which are wide

scope, the test can be answered freely and guided (because there are options), the test can be assessed objectively. Then, the test forces students to have serious learning because it is hard to speculate on which part of the lesson to be learned (Purwanto, 2013:39).

On the contrary, the test has number of weaknesses; students don't create their own answer, they just will guess the answer if they don't know the true one, than the test is incapable to provide information to be used as a feedback to diagnose learning experience (Wahyuni & Ibrahim, 2012:11). In addition, creating the test need carefulness, spend long time and less economy because the test needs a lot of paper (Purwanto, 2013:39).

According to Wahyuni and Ibrahim (2012:12) there are several kinds of objective tests, namely:

(i)     *True-False:* consists of statements related into two possible responses; true or false; students have to determine their decision.

(ii)    *Matching:* there are set of questions and answers placed into two different columns, students have to match it.

(iii)   *Multiple Choices:* consists of a stem and responses, test-takers have to choose one correct response. In brief, the test has equally definite answer. Since this research concerns on multiple choices test, both of the tests will not be discussed specifically in this discussion.

**2.4 Multiple Choice Questions (MCQs)**

MCQs can be used to measure learning result more complex. Popham (2011:148) said that multiple choice items can be used to measure students' abilities to engage in higher level of thinking. Moreover, according to Arikunto (2013:183) multiple choices test consists of stem and options. Among of the options consist of correct response and others are distractors. Similar concern also stated by Arifin (2013:138) that MCQs consists of main matter and response choices. The main matter can be both of *question* or *statement* form that don't perfect yet, generally called as *stem.* While the response choices are called as *options;* that may be word, number and sentence.

Moreover, Arifin (2013:138) said that there is no standard for amount of the options actually. It may consist of 3, 4, or 5 alternatives. More alternatives are better. It aims to deduct chance of guessing. The more alternatives mean less possibility for students to guess the correct answer. Therefore, the situation forces students to have critical thinking.

In other words, Coombey & Hubley (2017:17-18) have different idea; they said that the best number of options for F/SL testing is four; one should be an unambiguous correct or best answer. The three remaining options function as distractors. The case is talked in their book as guiding principles for teacher when writing MCQs. Besides that, there are other principles as stated by Coombey & Hubley (2017:17-18), as follow:

a) Distractors should attract students who are unsure of the answer and should be related in some way (e.g. same part of speech).

b) All options should be the same length and level of difficulty.

c) The question should be clear from the stem of the MCQ.

d) The language of the stem and responses should be as simple as possible to avoid skill contamination.

e) The selection of the correct answer should involve interpretation of the passage/stem, not merely the activation of background knowledge or "verbatim selection".

f) Avoid using "all of the above", "none of the above", or "a, b, and sometimes c, but never d" options.

g) All options should be grammatically correct unless error identification is part of your course outcomes.

h) Correct answers should appear equally in all positions.

i) Make sure there is an unambiguous correct answer for each item!

j) As much context as possible should be provided.

k) Recurring information in response options should be moved to the stem.

l) Avoid writing absurd distractors!

m) Avoid extraneous clues!

n) Avoid sequential items where the successful completion of one question presupposes a correct answer to the preceding question!

Many people believe that MCQs test is easy one to answer and score a wide element of knowledge. Nowadays, further, the test can be scored mechanically. In fact, they can be marked by computer is the result of MCQs can be used to examining among bodies for testing a large number of test takers.

MCQs test provides varying response in their relative correctness. Then, the test-takers require choosing one correct answer among provided alternative.

However, MCQs is difficult and hardest types of objective tests to construct for among classroom teachers (Coombe & Hubley, 2007:17). They require carefulness to pay attention in per item that have to have a same position each other. Clearly, the test should have same length, correct grammar, and correct spelling. In addition, designing MCQs is time consuming process for the test-makers.

### 2.5 Item Analysis

Having constructed and assessed a test, a teacher needs to know how good the test questions are, and whether the test items are able to reflect students' performances in a course related to the learning experience. It means that the teacher have to analyze the item tests. If this activity always be done, it will improve the teacher's creativities to write tests.

Talking about item analysis, moreover, need to know what the item analysis is? It is one of the methods may use to evaluate teaching and learning process. "Item analysis is used to assess the assessment tool and to ensure that the used tool is reliable and valid." (Khoshaim & Rashid, 2016:120). Moreover, Arikunto (2013:220) added that item analysis as systematic procedures that will give us specific information toward item tests have been created. It can be said that item analysis is a process to collect information from students' response to assess the whole test or its items.

Item analysis aims to describe characteristics, quality of test items, and also other things related to improve, arrange, and use test that already be appropriate. However, the test need to be held when it has weaknesses that should be fixed for conducted test future. In addition, purpose of item analysis is describing strength and weaknesses of test items, so they can be selected and revised (Wahyuni & Ibrahim, 2012:128).

However, performance of the test, itself, is very crucial in compiling future test. Since a great deal, that multiple choice questions need hard effort and spend a lot of time to completed, most teachers or test-constructors want to use the test again without further changes or adapting them for next use. Consequently, it is important for them to analyze the item which is the test answered correctly and badly by among students taking the tests.

For the reason above, commonly, there are two ways to analyze tests that need to know by students or test-makers, viz. theoretical or qualitative and empirical or also known by quantitative manner (Wahyuni & Ibrahim, 2012:56-57).

Qualitative manner is considering process to analyze item tests that have been constructed, checking appropriateness with basic competence and indicator measured and also requirements that may be about content, construct, and language. It can be assumed that qualitative manner is analyzing tests before the tests given to the students or test takers.

On the other hand, quantitative manner is organized when the tests have been conducted. Teacher will check the items to answer what they have stood in

appropriate level otherwise they don't fulfill standard of its. Quantitative manner, nonetheless, can be done by examining point of view of their difficulty level and their level of discrimination. In addition, Arikunto (2013:222) said that there are three points related to the item analysis, they are difficulty, discrimination power, and distractor items. The more explanation can be comprehended as follow:

### 2.5.1 Difficulty Index

Difficulty index of an item simply look how easy and hard of the particular item provided in the tests (Heaton, 1988:178). The too easy tests will not stimulate students to improve their efforts to solve it. On the contrary, the too difficult tests will caused students feel downhearted, then they don't enthusiasm to solve it because they think can't do it. Moreover, analyzing difficulty level of the test items show generally, what the tests include difficult, acceptable or easy (Wahyuni & Ibrahim, 2012:129). Furthermore, difficulty level, commonly, is a comparing between total of students' correct answers and total of students. If a number of students are able to answer the item correctly are great, it indicates that the tests are easier and vice versa.

Difficulty index can be stated as a percentage. For instance, an item that can be answered correctly by all students taking the test, then difficulty level indicated in the amount of 100%. For example, if the correct answer selected by 16 of 40 students, it means that value of difficulty level will be 40% or 0.40. (Wahyuni & Ibrahim, 2012:129)

A number that indicated easy or hard a test like the example above (0.40) called difficulty index. Moreover, numbers of difficulty level of a test range

Perpustakaan Universitas Islam Riau

Dokumen ini adalah Arsip Milik :

between 0.00- 1.00 difficulty indexes. Clearly, a test is in 0.00 difficulty level indicates that the test is too hard, and a test that is in 1.00 level, indicates that the test is too easy.

Furthermore, analyzing difficulty index can be measured by the following formula:

$$IF = \frac{FH+FL}{N}$$

which:

IF  : *Item Facility,* difficulty index of the test item

FH  : *Frequency High,* total of correct answer upper group

FL  : *Frequency Low*, total of correct answer lower group

N  : Total of students upper and lower groups

Moreover, based on difficulty index, Arikunto (2013:225) classified it into three categories, such as:

**Table 2.1 Difficulty Index**

| Difficulty Index | Interpretation |
|------------------|----------------|
| 0.00-0.30 | Too difficult |
| 0.31-0.70 | Acceptable |
| 0.71-1.00 | Too easy |

From the information above, appropriate difficulty index is in 0.30-0.70 range. Item test that is out of the numbers should be revised for future used. To make it clear, there are provided examples:

**Table 2.2 Examples of Analysis Difficulty Index**

**Upper Group**

| Name | Question Number | | | | | | | | | | | | | | | Score Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | |
| L | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 14 |
| B | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 12 |
| N | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 12 |
| E | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 11 |
| A | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 10 |
| K | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 10 |
| Total | 5 | 5 | 4 | 5 | 6 | 4 | 3 | 5 | 4 | 3 | 5 | 5 | 3 | 6 | 6 | |

**Lower Group**

| Name | Question Number | | | | | | | | | | | | | | | Score Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | |
| R | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 7 |
| T | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 7 |
| H | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 6 |
| M | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 6 |
| Q | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 6 |
| J | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 4 |
| Total | 2 | 2 | 4 | 3 | 1 | 2 | 3 | 2 | 3 | 0 | 4 | 0 | 4 | 4 | 2 | |

Item number $1 = \frac{5+2}{12} = 0.58 \rightarrow$ acceptable

Item number $2 = \frac{5+2}{12} = 0.58 \rightarrow$ acceptable

Therefore, to conducting a test, make sure that the test does not too easy or too hard, because it thought disable to provide whole and helpful information about students taking the test. It means that, the item tests is in about 1.00 or 0.00 range need to keep away or don't be used again. An appropriate difficulty index is about 0.50, it is appropriate for the whole test or for each item (Wahyuni & Ibrahim, 2012:130).

Moreover, opposite opinion stated by Daryanto that a test which is too easy and too hard does not mean that should not be used. It depends on the usage. Clearly, for example the difficult tests may be used when you want to select just

the top students of much numbers. Vice-versa, the easy tests may be used when you want to recruit most of participants (Daryanto, 2008:183).

### 2.5.2 Discrimination Index

Discrimination index of an item indicates difference between the more able group taking the test and the less able the group. Similarly, discrimination index provide answer of a question: How much a test item is able to separate students of upper and lower group? Furthermore, our logic as a based considering of it, to be exacts, the upper group should be able to provide more correct answers than the lower group. As higher discrimination index of a test item, indicates possibilities to differentiate between the upper group and the lower group (Wahyuni & Ibrahim, 2012:135).

According to Arifin (2012:350) calculation of discrimination index is a measurement to describe how far of an item able to distinguish between learners who have grasped competence and who have not under a certain criteria. The higher discrimination index coefficient of the item indicates the item is more able to distinguish between the students. It can be said that the higher discrimination item is effective.

Furthermore, number of discrimination power is similar to difficulty index. Lowest value (0.00) of a test item is poor one, because it disables to differentiate the good students and the bad one. For instance, a test item answered correctly by 15 students of upper group and 15 students of lower group is an item that has value amount of 0.00 discrimination power. That is to say, the test disables to show the difference.

The discrimination index will be higher one, otherwise, founding difference amount of correct answer showed by the two groups. As higher difference total of the correct answer between the two groups indicates discrimination index will be higher also. The highest discrimination index of an item is in 1.00, that is to say all of upper group tend to do well on the test and nothing of the lower group can do it.

Moreover, analyzing the discrimination index can be counted by formula as follow:

$$D = \frac{FH - FL}{n}$$

which:

D       : Discrimination item, discrimination index of the test item

FH      : Frequency High, total of correct answer upper group

FL      : Frequency Low, total of correct answer lower group

n       : Total of students upper and lower groups; 27,5% of subject

Furthermore, an appropriate test items should have discrimination index falling in at least 0.20 range. In addition, Arikunto (2013:232) provided interpretations of discrimination index as follow:

**Table 2.3 Discrimination Index**

| DISCRIMINATION INDEX | INTERPRETATION |
|---|---|
| 0.00 - 0.20 | Poor |
| 0.21 - 0.40 | Satisfactory |
| 0.41 - 0.70 | Good |
| 0.71 - 1.00 | Excellent |
| - (negative) | Rejected |

In addition, based on the data of difficulty index, let see examples of discrimination index of the test item to be clear, as follow:

Item number 1 = $\frac{5+2}{6}$ = 0.50 → excellent

Item number 2 = $\frac{5+2}{6}$ = 0.50 → excellent

### 2.5.3 Distractor Index

Seeing a good test item is not only from both of difficulty and discrimination index but distractor index also. Fulcher & Davidson (2007:326) said that improving multiple choice items based on the result seen not only in the key but in the distractors also. Distractors are incorrect choice, they must be reasonable to students taking the test that do not learn yet about the knowledge taught. In addition, distractors must look like correct options in grammatical form, style, and length (Haladyna & Rodriguez, 2013:67).

Moreover, Arikunto (2013:233-234) stated that in distractor analysis can be done through counting students who selected option A, B, C, D, or E or who didn't any selected known as Omit (O). Furthermore, in this analyzing reported whether a distractor is effective or isn't. A distractor that does not be selected indicates that it is poor/non-functional distractor; too striking and mislead. On the contrary, a distractor is effective when it has an appropriate pull for students who less abilities or do not comprehend a content in the test.

There are three categories for distractor analysis, such as:

**Table 2.4 Distractor Index**

| Distractor Categorizes | Students |
|---|---|
| 1. Effective | $\geq 5\%$ |
| 2. Less effective | $< 5\%$ |
| 3. Non-functional | No one selected |

(adapted from Arikunto, 2013:234)

From the table above can be said that an effective distractor is option selected by 5% or more students. Then, a non-functional distractor is indicated when none student chose the option. Next, if amount of students selected an option less than 5%, it means the option is less effective. It can be concluded that a useful distractor is chosen by at least 5% of students taking the test.

Furthermore, a distractor that has no fill the appropriate criteria is better to revised. The lack may be from the sentence formulations so need to be written through changing it as necessarily. Constructing the test is not easy to be done. For the reason, if the item still able to be revised, it's better to do; don't be thrown away.

To make it clear, there are provided example of distractor analysis here:

**Table 2.5 Example of Distractor Analysis**

| Group | A* | B | C | D | Om | Total |
|---|---|---|---|---|---|---|
| Upper Group | 2 | 1 | 9 | 2 | 1 | 15 |
| Lower Group | 1 | 4 | 5 | 4 | 1 | 15 |
| Total | 3 | 5 | 14 | 6 | 2 | 30 |

Item with * is correct answer

From the table above can be described that option A is the key; there are 3 of 30 students selected it; it means that the test item is too difficult. Option B is effective. Then, option C needs to be considering because amount of upper group more than lower group. Then, option D is effective. Last, there are 2 students who didn't select the option; it is thought as proper case.

28

## 2.6 Relevance Studies

There have been some previous studies that have similar concern about item analysis, such as Boopathiraj & Chellamani (2013), Mehta & Mokhasi (2014), and Colte (2015). The information about each of the study can be found in the following explanation:

The first previous study was done by Boopathiraj & Chellamani in 2013. They tried to analyzed test items of a researcher made test in the subject of Research in Education for the student-teachers of Master of Education (M.Ed). In their research, they have been analyzed about two principles of item analysis in MCQs test, entitled "Analysis of Test Items on Difficulty Level and Discrimination Index in the Test for Research in Education". The objective of their work is to find out the item difficulty and the power of discrimination of multiple choice test items. Then, method of their study is descriptive qualitative research. Then, result of their research showed that thirteen items out of 60 (21%) were rejected either due to difficulty level or discrimination index. Thirty five items (58%) were accepted without revision while 12 items were accepted provided that necessary revision. It means that most of the items were falling in acceptable range of difficulty and discrimination level however some items were rejected due to their to poor discrimination index.

The second previous research is about item analysis also which is concern on three principles of item analysis. It conducted by Mehta & Varsha in 2014, entitled "Item Analysis of Multiple Choice Questions- An Assessment of the Assessment Tool" that aims to assess the quality of MCQs for creating a variable

question bank for future and to identify the low achievers and their learning difficulties which can be corrected by counseling and modifying learning methods. Then, method of their research is descriptive qualitative research. Then, for the result, they found that Difficulty index of 31(62%) items was in the acceptable range (p value 30-70%), 16(32%) items were too easy (p value >70%) and 3(6%) items were too difficult (p value <30%). Discrimination index of 26 (52%) items was excellent (d value>0.35), 9(18%) items was good (d value 0.20-0.34) and 15(30%) items were poor (d value<0.2%). A total of fifty items had 150 distractors. Amongst these, 53(35.3%) were nonfunctional distractors, 38(18.6%) were functional distractors and 69(46.06%) had nil response i.e. not attempted by any student. On the basis of non-functional distractors, distractor effectiveness of each item was assessed. Inter- relationship between these indices was analyzed. It means that items having average difficulty and high discrimination power with functional distractors should be incorporated into future tests to improve the test development and review.

The third previous study conducted by Kolte in 2015 under the title "Item analysis of Multiple Choice Questions in Physiology Examination." Objective of the study is to develop a pool of valid items & to update question bank for designing question paper as per the need of assessment. Next, method of the study is total 40 items from psychology preliminary examination of 100 students of 1st year M.B.B.S. were analyzed. Each item was analyzed for Difficulty index, Discrimination index, and Distractor effectiveness. Then, result of the research showed that Difficulty index i.e. 'p' value of analyzed MCQs ranged from 6.25%

(lowest) to 90.6% (Highest) & Discriminative index i.e. 'd' value ranged from 0 (lowest) to 0.63 (Highest). Total 65% items were in acceptable range of difficulty level ('p' value 30 – 70%) & 10 % items were very difficult which later discussed with students. Discrimination index of 60% items was excellent (d value>0.35). No item had Negative discriminative power. About 47.5% items had100% Distracter Efficiency (DE) whereas 7.5% items had 0% DE.

Based on three previous studies above can be informed that they have similar concern with this research. For the result, they are so closely interrelated to this research that the writer will do. Those previous studies discussed about item analysis of MCQs test for bank and research. In this study, the researcher focused on English mid test.

## 2.7 Conceptual Framework

In this research, the writer presents the conceptual framework into the following figure bellow:

**Figure 2.7.1 Conceptual Framework**
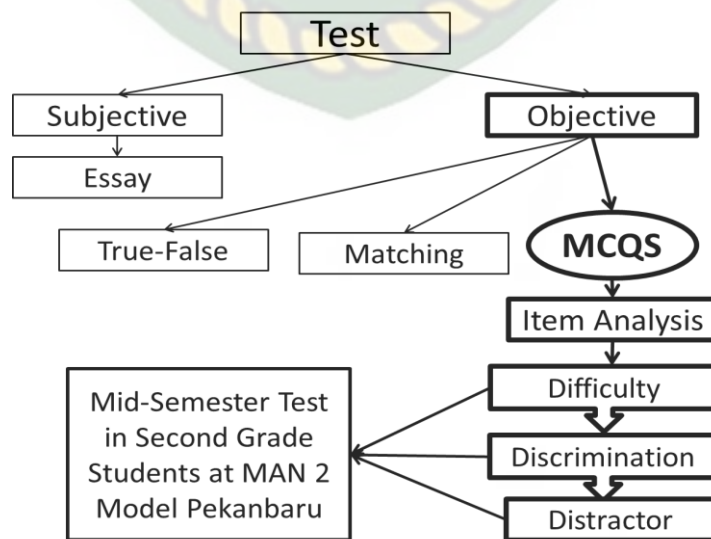
Testing is a tool to evaluate students' achievement and to determine the progress of objective of a course. There are two kinds of test; subjective and objective. Furthermore, subjective test consists of one form of essay test. On the other hand, objective test consists of three forms of test, they are true-false, matching, and multiple choice questions tests (MCQs). This study focuses on multiple choice test. There some ways to assess the test, one of them is item analysis. To conduct item analysis, there are three case to do it; difficulty, discrimination, and distractors indices analysis. In addition, this study analyzed those terms in form of MCQs in mid semester test of second grade students at MAN 2 Model Pekanbaru.

## 2.8 Assumption

In this study, the researcher assumed that most of English multiple choices questions in mid test used by second grade students at MAN 2 Model fulfilled the appropriate criteria of item analysis in terms of difficulty, discrimination and distractor indices; it is an appropriate test to evaluate students' achievement.