CHAPTER II

REVIEW OF RELATED LITERATURE

2.1 Definition of Reading

Many definitions have been proposed by some author about what reading is. Mikulecky (2008) defines reading as a cognitive processess in making a sense of text. He also adds that reading is the vital of instruction in all facet of language learning. Also, Gilakjani & Ahmadi (2011) conclude that reading is a process in which readers construct a clear representation of text by mixing their own background knowledge with the information in a text to apprehend that text. In addition, Chastain (1988) states that reading is a process concerning with the activation of related knowledge and relevant language skills to perform an exchange of information from one person to another. Still, the definition of reading related to Fleming et al (2002), reading is a complex literary process that affect reading development and proficiency.

There are many kinds of reading, such as reading speed, reading accurancy, reading comprehension, etc. Reading comprehension is one of capability that include the extraction process and constructs meaning through interaction with written language (RAND Reading Study Group, 2002). Reading comprehension refers more to the understanding of the students in finding the idea of reading text and combining their understanding related to reading text itself. Iwai (2010) in his research concludes that reading comprehension is the process of constructing the meaning of written words with various approaches, such as using background knowledge, analyzing words, inferring the text, and identifying key vocabulary or information. Beside that, Pang (2003) defines reading

7

comprehension is an ability through the active process of deriving meaning from the text. He adds that it process includes using knowledge in drawing inferences from the words and expressions which the writer uses in communicate information and ideas to construct the meaning.

From these definition, it can be concluded that reading comprehension is one of the ability of reading as a proccess of constructing the meaning of what has been read through combining our basic knowledge with the information put in text in order to comprehend the text.

2.2 Definition of Test

According to Bertrand & Cebula (1980) which cited by Maciel et al (1983), tests are tool used to check and measure students' growth and development. Testing is the practice of making objective judgments regarding the extent to which the system meets (FHA, 2007). Related to the definition above, it can be concluded that test is a tool to find the extent of testee' understanding development.

A product, in this case iss a test, certainly has a purpose. As reported by the Federal Highway Administration (2007) that there are several fundamental purposes of testing. One of them is verifying that the product meets the functional, performance, design, and requirements.

Testing is highly important. Testing is very useful and necessary at the time. It is needed in order to provide the information about the achievement of learners. Without it, we would not be able to determine the extent of their understanding or ability (Hughes, 2003). In addition, the preparation of the test

cannot be done arbitrarily. There are many important things that must be considered. This is in line with the achievement of the purpose of the test itself. As stated by Clay (2001) that well-constructed test enables teachers to evaluate students mastery of course objectives. More than that, test motivates students and reinforces learning.

Reading test is known as a series of exercise in the form of reading material and items of question designed to find out the extent of students' understanding development. Furthermore, it mostly involves the ability in understanding what they read based on the following instruction through the text. Based on what discussed above, this type of reading test commonly tend to reading comprehension test.

Reading comprehension test is a test that measure students understanding towards the text they read. It includes the ability to answer the question about the meaning, find the idea based on passages that they read, find the specific information contained in reading passages, and find the specific vocabulary related to the reading text.

2.3 Form of Test

Form of test consists of two kinds, they are subjective and objective. The difference in between is about evaluating the test. As stated by Heaton (1975), test usually uses the term subjective and objective to refer to the scoring of tests. Brief explanation of each, subjective test and objective test, are as follows:

2.3.1 Subjective Test

Subjective test is a form of test which trains the students to express their ideas toward question possed to them. The test called subjective because it need decision as the part of scoring process. Heaton (1975) mentions that in doing subjective test, students must think of what to say then express it as well as possible. He adds that in subjective test, it requires good writing performance. In this item, the students' answer tend to be subjective. It means the answers depend on what they want to say. Eventually, there will be found some answers which are better than the others, and therefore it causes a problem in scoring the items.

There are several kinds of test items belong to the subjective category test. Here are several kinds of test item which are included in subjective test form:

- Short Answer

  The short answer question is one of test format which is similar to an essay question. This type usually consists of questions which have limited the length of the answer.

- Essay Type Question

  The essay question type is designed to short answer of a paragraph. The most common types of essay question are about the issue. Students give their argument or ideas then explain and exemplify their answer to make a conclusion based on questions posed to them.

2.3.2 Objective Test

According to Heaton (1975), objective test is a test that usually has only one correct answer and can be scored mechanically. He explains that objective test need more prepearations than subjective test, including the necessity of constructing each test item as carefully as possible (Heaton, 1975).

Objective test is often considered as one of the tests which are very simple. Heaton (1975) declares that although the objective test criticized that it is easier to answer than subjective test, but efforts to develop each test item is required to achieve the standard. Preparation in building the test items of the objective test is to go through the process of pre-tested before being administered. In addition, it also allows the test maker to calculate the approximate degree of difficulty of the test. Standards may be compared not only between students from different areas or schools but also between students taking the test in different years.

- Multiple-choice

    Multiple choice is the most widely used formats in the reading test. This type is useful to measure students' comprehension. In this format, there are four alternative answers for each question. They are one as the correct answer and others as detractors.

- True/false

    True/false format test allows students to choose whether the question statement that they read are true or false. The alternative answers of this type are limited, they just have a true and false

options. Students have to choose either true or false in the answer field.

2.4 Characteristic of Test

A good test mostly defines as a test which could give valid information close to the real situation. As a tool to measure students' comprehension, it is hoped that the information gotten from this reading test could relate to students and teaching learning in classroom practice. That is why a test should fulfill the criteria of a good test such as validity, reliability, item facility, and item discrimination.

2.4.1 Reliability

Reliability is one of criteria reading test must have. Many definitions have been proposed by some author about what reliability is. Noda (2003) defines reliability, as a crucial component of standardized testing and points out that test-takers receive almost the same score when they are presented a true test for multiple times. Razi (2012) also adds that reliability implies that if a reading test is true then the tester is sure that the exam is consistent and test-takers perform almost the same at all times the test is presented. Meanwhile, Carmines & Zeller (1979) state that reliability concerns to test which measures the same effect on repeated tests. This tendency toward consistency found in repeated measurement of the same phenomenon is mentioned to reliability.

Based on some definition above, reliability means a consistency of test in producing the same result in repeated measurement. The test is

considered reliable if the test produces the same result every time it is used. The reliability of a test is determined by the consistency of the scores. Hence, the more similar the scores will be, the more reliable the test is said to be.

There are many methods of evaluating the reliability of the language test. Korb (2009) explains four methods in investigating the reliability of the test:

2.4.1.1 Split-Half Reliability

This method is uses to determine how much error in a test score caused by poor test construction. The way to calculate is by administering one test and then calculating the reliability index by coefficient alpha, Kuder-Richardson formula 20 (KR-20) or the Spearman-Brown formula.

2.4.1.2 Test-Retest Reliability

This method is used to determine how much error in a test score is due to problems with test administration or example too much noise distract the participant. The way to calculate is by administering the same test to the same participants on two different occasions and then correlating the score of test of the two administrations of the same test.

2.4.1.3 Parallel Forms Reliability

This method is used to determine how comparable are two different versions of the same measure. The way to calculate is by administering the two tests to the same participants within a short period of time and then correlating the test scores of the two tests.

2.4.1.4 Inter-Rater Reliability

This method is used to determine how consistent are two separate raters of the instrument. The way to calculate is by giving the result from one test administration to two evaluators and correlating the two markings from the different raters.

Kuder-Richardson formula 20 (KR-20) is used to investigate the reliability of the test. Kuder-Richardson used if we have dichotomous items as multiple choice exams. Here is such a formula for measuring reliability using the Kuder-Richardson formula 20 (KR-20) :

$$r_{KR20} = \left(\frac{k}{k-1}\right)\left(1 - \frac{\sum pq}{\sigma^2}\right)$$

r_{KR20} = Kuder-Richardson formula 20

k = the total number of test items

$\sum$ = sum

p = the proportion of the test takers who pass an item

q = the proportion of the test takers who fail an item

σ^2 = the variation of the entire test

To calculate the KR-20, we should enter the data in an Excel Spreadsheet. Here is the example of the table in an excel spreadsheet:

Student Name	Question						Total Exam Score
	1	2	3	4	5	...	
A							
B							
C							
D							
E							
...							
Number of 1's							

Portion Passed (p)							
Portion Failed (q)							
p x q							

▢ = this column lists each student

▢ = mark 1 if the students answer the item correctly and 0 if the students answer incorrectly

▢ = to calculate the proportion of the sample who answer the item correctly, first count the number of 1's for each item. This gives the total number of students who answer the item correctly

▢ = divide the number of students who answer the item correctly by the number of students who take the test

▢ = use formula 1 – p (1 minus the proportion who passed the item) or count up the number of 0's or each item and then divided by the number of students who take the test

▢ = take p times q

▢ = count the total exam score

Beside of using KR-20, investigating reliability o the test can also be done by using SPSS. The data will be presented automatically by following the steps of calculating reliability.

2.4.2 Validity

In writing the test item, it does not produce an item which is ready to be tested until it is validated. Validity itself according to Carmines & Zeller (1979) concern the essential relationship between concept and indicator. An indicator of some abstract concepts are valid to the extent that it measures

what it aims to evaluate. Validity as the extent to which any measuring instrument measures what it is intended to assess. Also, the most traditional definition of validity is the extent to which a test really measures what it is supposed to be measure. If it does not fit that purpose then testing could be useless or misleading (Lemmeti, 2014)

Validity concerns to a compatibility test with the main targets that need to be measured. Evidences for validity:

2.4.2.1 Content Validity

Content validity addresses the match between test questions and the content or subject area they are designated to measure. This element of validity is based on a comparison between the test items and the test specification. Content validity includes any validity strategies that focus on the content of the test (Brown, 2000).

Investigating the content validity can be done in a way of comparing the test that has been developed by the curriculum or syllabus.

2.4.2.2 Criterion Validity

Criterion-related validity looks at the relationship between a test score and an outcome. It is done by analyzing test scores correlate with similar test scores who have a good characteristic. Criterion validity commonly divided into concurrent and predictive validity. Criterion-related validity usually includes any validity strategies that focuses on the correlation of the test being validated with some well-respected outside measure of the same objectives or specifications (Brown, 2000).

Investigating the criterion validity can be done in a way of analyzing test scores correlate with similar test scores that have a good characteristic. Holcomb Hathaway (2008) shows that the first method that we will consider in evaluating the criterion validity of norm-referenced measurement: a criterion measurement and a new measurement technique or measuring the same underlying constructs in a group of people.

2.4.2.3 Construct Validity

Construct validity refers to the degree to which a test or other measure assesses the underlying theoretical construct it is supposed to measure. Construct validation requires the compilation of multiple sources of evidence. Construct validity has traditionally been defined as the experimental demonstration that a test is measuring the construct it claims to be measure (Brown, 2000).

In investigating the construct validity, it should be established by an accumulation of evidence. For example, taking the unified definition of construct validity, we could demonstrate it using content analysis, correlation coefficients, factor analysis, ANOVA studies demonstrating differences between different groups or pretest-posttest intervention studies, factor analysis, multi-trait/multi-method studies, etc (Brown,2000).

2.4.3 Item facility

Item facility is about how easy or difficult an item of test or the subject of the test. If the item of test can be answered by many students, it belongs to easy item. On the other hand, if the item of test just can be

answered by a few students, it belongs to difficult item. Item facility index is expressed with a range between 0.00-1.00. Djiwandono (1996) in his study states that the ideal index of item facility is between 0.20-0.80.

The formula used for investigating item facility of multiple choice items is as follows:

$$IF = N\ correct : N\ total$$

IF : Item Facility

N correct : Number of students who answer correctly

N total : Number of students who take the test

2.4.4 Item Discrimination

Item discrimination is one of the criteria in the analysis of a test item. A test is said to be good if it has passed the item discrimination analysis. This refers to the extent to which the success or failure of the item in indicating students performance who perform well and who perform bad (Boopathiraj & Chellamani, 2013). The value range of Item Discrimination between 0.0 and 1.00. The higher the value, the higher the discrimination of the item. A highly discriminating item indicates that the students who had high tests scores got the item correct whereas students who had low test scores got the item incorrect (Boopathiraj & Chellamani, 2013).

The formula used for investigating item discrimination of multiple choice items is as follows:

$$\text{Item Discrimination} : \frac{High\ group\ \#\ correct - low\ group\ \#\ correct}{1/2\ X\ total\ of\ two\ comparison\ group}$$

The formula used for investigating item discrimination of essay test items is as follows:

$$\text{Item Discrimination} = \frac{Sh - Sl}{N(Scmax - Scmin)}$$

Sh = Correct answers for upper group

Sl = Correct answers for lower group

N = Number of subject taking the test

Scmax = Maximum score given to an item

Scmin = Minimum score given to an item

2.5 Developing Test

In developing a test, there is a lot of process that should be done. Bachman and Palmer (1996) assert that test development is the entire process of creating and using a test, begins with its initial conceptualization and design, and ends up with one or more archived test and the results of their use. They propose three stages of test development: design, operationalization, and administration.

2.5.1 Steps of Developing Test

When we want to create a product, of course, there are many steps that must be passed. Step of this development is also known as design development test. There are many kinds of design development test. Among them is the design of developing test by Harris (1969).

Harris (1969) figures out the design of developing the test: (1) Planning a test (2) Preparing the test items and directions (3) Reviewing the Test (4) Pre-testing the test (5) Analyzing the result of the test. These steps are important that should be followed if we want to develop a test.

2.5.1.1 Planning the test

Planning is the most important step when dealing with test development. In this step, the researcher determines the general objective of the course, proposes the objective of the test, establishes the general design of the test, specifies the format and kind of test, and gives the description of text used.

2.5.1.2 Preparing the test items and directions

Preparing the test items and directions include the process of developing each item of the test, the direction, provide the answer key, describe the content and specification, and describe how is the scoring and grading. In preparing the items of the test, it's necessary to create more items than the total items that will be needed. It is as an anticipation of item test reduction that will be necessary as long as reviewing process of the entire test.

2.5.1.3 Reviewing the test

Reviewing the test means the test item which has been written by the researcher are submitted to the one who experience in this subject matter, in this case, the English teacher.

2.5.1.4 Pre-testing the test

After reviewing the entire test, firstly, the test has been tried out to the subject whom the test is being designed. This step gives the chance for the developer to check the estimated time required or doing the test.

2.5.1.5 Analyzing the result of the test

After getting the result of the test, the researcher can do the data analysis. Data analysis includes calculating the mean score, investigating item facility, reliability, validity and item discrimination. At the end of these steps, the researcher should give the general description or summarized of the entire process of developing reading test and review some important aspect while developing this reading test.

2.5.2 Developing Test for Seventh Grade

In developing test or seventh grade student of junior high school first we have to know the general objective of the course. The general objectives of teaching and learning of seventh grade in second semester based on the syllabus is as follows:

3. Understanding the knowledge (factual, conceptual and procedural) based on curiosity about science, technology, art, culture and events related to the visible phenomenon.

3.6   Understanding the social function, the structure of the text, and linguistic elements of text label name and a list of items, according to the context of use.

3.7   Understanding the social function, the structure of the text, and linguistic elements in texts to express and ask behavior/action /function of people, animals, objects, according to the context of use.

3.8   Understanding the social function, the structure of the text, and linguistic elements from the descriptive text by stating and asking about

descriptions of people, animals, and objects, very short and simple, according to the context of use.

3.9 Understanding the social function and linguistic elements in the song.

4. Trying, processing, and serving in the realm of the concrete (using, break down, compose, modify, and make) and the realm of the abstract (writing, reading, counting, drawing, and writing) in accordance with what have been learned by the students in school and the other in the same source of viewpoints/theories.

4.7 Getting the meaning of short and simple text instruction, marks or signs, warning signs, verbal and write very short and simple.

4.8 Getting the meaning in oral and written descriptive texts, very short and simple.

4.9 Getting the meaning of the song.

2.6 Scoring and Grading

After the test has already been tried out, there are steps that should be done by the test related to developing test step. Those are scoring and grading.

2.6.1 Scoring

There are many definitions have been purposed by several researcher about what scoring is. Rivera (2007) mentions that test scores were used to help clasify students. Moreover, Fulcher & Davidson (2007) state that scoring focused on how good of language testing.

Scoring in its application depends on the type of questions that will be assessed. If the type of question is different, the way used in doing scoring will be different too. In this case, the test consists of two types, namely an objective test and subjective test. Each way of scoring on both types of text, of course, have their own way.

Hanafi (2007) in his study states that in assessing subjective test, subjective judgment is the key in doing scoring for its type. He also adds that test score given must uphold the truth answers based on subjective judgments in accordance with predetermined criteria of scoring. Meanwhile, to assess the objective test, the answers of the testee are determined by the criteria that have been made.

In this cases, the type of test scoring is limited multiple choice and essay test scoring. For any kind of question is certainly has a different way of scoring. For multiple choice scoring, there is only one correct answer to each question. Correct answer will get one point, while wrong answer will automatically get zero point. To determine the overall score, it can be obtained by dividing the total number of correct answers to the total questions in the test then modified in terms of percent. For example, when the student answers 23 questions correctly and total questions in the test are 25, then the results obtain 23 / 25x100% = 80%. If the score is converted in the range of 1-100, then the result 92.

As for the scoring type of essay test, there is a scale used to determine the correctness of students answers. Brown (2005) in his study suggests that in doing the scoring on the type of essay test, it can use two ways. Those are

interrater and intrarater. This is intended to minimize the subjectivity in determining the correctness of students answer. In interrater, scoring is done by two assessors. Both of score then summed up and divided by two. The result of that division would be the final score in the essay test. In the second case, namely intrarater, scoring is done only by the test maker. Scoring performed twice on the same student's work. Both scores are then added together and then divided by two to get the final score of answers.

2.6.2 Grading

The whole process of test design, development, implementation, validation and valuation which have been described so far lead up to the point where we can evaluate the performance of each students and report that in some ways. In this contexts, test indeed ranks them from highest to lowest. This step is called grading.

As mentioned by Walvoord & Anderson (1998), grading is a process which a teacher evaluates student learning through classroom test and assignment. Frisbie & Waltman (2001) add that grade provides the information of students' performance level of last achievement. It also provides the information to the students for self-evaluation, and creates assumption about students achievement.

There are two ways in determining test takers's grade. Those are Criterion Reference (CR) and Norm Reference (NR). Criterion Reference (CR) means the result of test will be compared with the criteria which have been determined before. This criterion is generated by determining a score as

a minimum score that considered whether they pass or fail. If the scores gotten by the students exceed the minimum criteria of score, they certainly pass the test. But if student score is less than the minimum criteria which have been determined, then the student is considered fail the test (Hanafi, 2007).

In norm-referenced, students' score is interpreted based on the normal scores. Normal score gets from the scores obtained by most of the students. A score will be given to 2 % of students lying on 2 standard deviation (2 S) and above. B score will be given to 14% of the students lying between the first and second standard deviation (+1 S up to +2 S) above the mean. C score will be given to 68% of the students (34%+34) which lying from -1 Strandard deviation (S) to +1 students. D score will be given to 14% of the students lying between one standard deviation below the mean (-1 S) and two standard deviation below the mean (-2 S). E score will be given to the other 2% of the students lying on -2 Standard Deviation (-2 S) below (Hanafi, 2007). The following explanation below can be interpreted as follows:
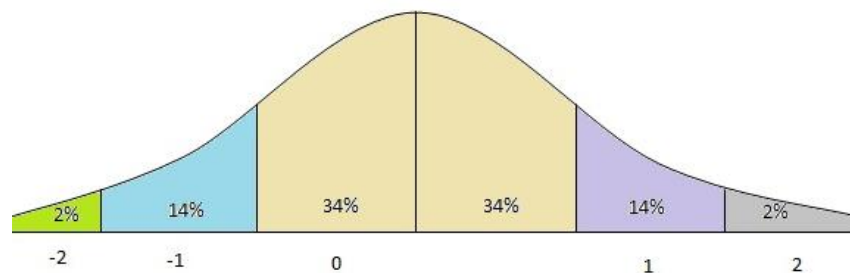

Figure 2.1 Distribution of Norm-reference Grading

The scores lies below (X<-2 S) is	E
The scores lies between (X-1 S) up to (X-2 S) is	D
The scores lies between (X-1 S) up to (X+1 S) is	C
The scores lies between (X+1 S) up to (X+2 S) is	B
The scores lies above (X>+2 S) is	A

Note: X= The mean Score
       S= Standard Deviation

2.7 Previous Study

To know the successful of the teaching learning, so the test instrument is needed to measure it. The previous study used to review some other previous researches related, to find the gap of the study.

According to Zainiyah (2014) with the tittle *"Developing a Reading Test Using Hot Potatoes Software for Junior High School in Mamba'ul Ulum Giri Kebomas Gresik"*, the developing test focused on 7th grade junior high school. The test in accordance with the 2006 curriculum, KTSP. Besides, the test is the result of the students need analysis in that school related to the test of interest using attractive software-based.

Another study is from Hanafi. In 2007, Hanafi conducted a study of developing reading test with the title *"Developing Reading Comprehension Test for The First Semester Students of English Program of Muhammadiyah University"*. This study focuses on developing reading test for reading comprehension I course in semester 1. This research based on the need of improvement related to the internal condition at English program of Muhammadiyah University that is not ideal enough. So, it need improvement, one of the way is by doing standardized test development.

This research tries to develop a test to see how far the achievement of all materials in semester 2 of 7th grade based on the description of the basic competence which presented on the indicators. The development of the course of objective in this test has been correlated to the 2013 curriculum. The result of this research hopefully can be used to measure the students understanding in one semester including to train the students reading comprehension. The ability

measured in this test more focused on the lower order thinking ability, which includes the ability to know, understand, and explore.