

### **III. RESEARCH METHOD**

This chapter discusses setting of the research, research design, population and sample, data collecting technique, try-out of the research instrument, research procedures, scoring system, data analysis and hypothesis testing.

#### **3.1 Setting of the Research**

The setting of this research was in MA Tribhakti at-Taqwa Rama Puja Lampung Timur. It was the second grade students at the early of second semester. Eight meetings had been conducted in both of experimental class. One meeting was administered to try out the instrument, two meetings were administered for pre-test for both classes and six meetings were administered treatments for both experimental class. Then one meeting for post test for both classes, after that the researcher administered interview.

#### **3.2 Research Design**

In this research, the researcher intended to find out whether there is any significant difference between the students' reading comprehension achievement who read literary text and those who read non-literary text through critical reading approach and find out students's responses after being taught by using literary text and non-literary text through critical reading approach. And to gain the answer to

the research questions in this research, the researcher carried out qualitative and quantitative study. In qualitative study, the researcher administered interview to find out the students' responses after being taught using literary text and non-literary text through critical reading approach. Meanwhile, in the quantitative study, the researcher carried out *pre-test-post test design*. The researcher selected two classes, one as the experimental class one and another as the experimental class two. According to Hatch and Farhady (1982:22) the design of the research was described as follows:

G1 : T1 X<sub>1</sub> T2

G2 : T1 X<sub>2</sub> T2

*Notes:*

G1 : experimental class one

G2 : experimental class two

T1 : pre test

T2 : post test

X<sub>1</sub> : treatment for experimental class one (using literary text through critical reading approach)

X<sub>2</sub> : treatment for experimental class two (using non-literary text through critical reading approach)

Referring to the design above, it stated that there were two independent variables: that is literary text and non-literary text. Meanwhile, the dependent variable was students' reading comprehension achievement.

This research answered the first research question by comparing the mean score of post test in experimental class one and the mean score of post test in experimental class two. Then the researcher make conclusions based on the results of that comparison.

### **3.3 Population and Sample**

The population of this research was the second grade students of MA Tribhakti At-Taqwa Lampung Timur in the second semester. There were four classes in XI Science and each class consisted of 30 students. Two classes have been taken as the sample of this research; the first class as the experimental class one, and the second class as the experimental class two that is class XI Science 1 which consisted of 30 students and XI Science 2 that consisted of 30 students. The two sample classes were selected randomly.

### **3.4 Research Question 1**

This part is describing the procedure of answering the first research question: Is there any significant difference between the students' reading comprehension achievement who were taught by using literary text and those who were taught by using non-literary text through critical reading approach.

### **3.4.1. Data Collecting Techniques**

To collect the data, the researcher administered reading comprehension test (pre-test and post test) and interview. Then, she analyzed the result of those activities which can be clarified as follows:

#### **1. Reading Comprehension test**

##### **1. Pre-test**

The pre-test was administered in order to investigate the students' reading entry point before the treatments. The type of the test was multiple choice in which the students was asked to choose one correct answer from the option a, b, c, or d. In this pre-test the students was given 20 items of reading comprehension and it was conducted within 60 minutes.

##### **2. Post test**

The aim of this test was to determine the effect of the treatments towards the students' reading comprehension after being give the treatment. This test consisted of 20 items of multiple choices for 60 minutes.

### **3.4.2 Try-out of the Research Instrument**

The try-out test was administered to determine the quality of the test that was used in taking the data. Try-out was the test that was given before pre-test and post-test. It used an objective test in form of 50 items and has four options a, b, c and d in 90 minutes. This test was conducted in order to determine the level of difficulty, discrimination power, reliability and validity of the test items before giving pre-

test and post test to the class. A good test should meet four criteria: a good level of difficulty, discrimination power, validity and reliability.

Practically, to determine the reliability, discrimination of power and level of difficulty in order to determine the quality of the reading test, the researcher used ITEMAN. According to Assessment Systems Corporation (ASC) (1989-2006), ITEMAN can be defined as one of the analysis program that compares Assessment Systems Corporation's item and Test Analysis package. Basically, ITEMAN can be used to analyze test and survey item response data and provide conventional item analysis (e.g., proportion/percentage endorsing and item-total correlation) for each item. The input data in order to be analyzable by ITEMAN should be formatted in ASCII (text-only) files. This can be completed successfully through the use of the ITEMAN for windows text editor, Notepad, a word-processing editor that produces true ASCII output, or a program written specially to format the data. The ITEMAN program can work only with multiple choice items.

### **1) Validity**

A test is considered valid if the test measures the object to be measured and suitable with the criteria (Hatch and Farhady, 1982).

There were four types of validity that were:

1. Face validity, concerns with the layout of the test;
2. Content validity, depends on a careful analysis of the language being stated;
3. Construct validity: measures certain specific characteristic in accordance with a theory of language learning;

4. Criterion-related validity, concerns with measuring the success in the future, as in replacement test;

Based on the types of validities above, the researcher used content and construct validity. It was to measure whether the test had a good quality or not because the other two were considered to be less needed. Both of them can be explain as follows:

**a. Content validity**

Content validity refers to the extent to which a test measures a representative sample the subject matter contents. The focus of the content validity is adequate of the sample and simply on the appearance of the best (Hatch and Farhady, 1982). To get the content validity of reading comprehension, the researcher tried to find the materials based on the standard competence in syllabus for second grade of se high school students in second semester that are students are able to construct meaning of functional text and simple monolog of narrative text as literary text and analytical exposition as non literary text to communicate with surroundings and the objectives of teaching by using critical reading approach were able to identify the main ideas, identify the specific details or information, making prediction, making evaluation and determine the inference. The researcher made a table of specification in order to judge whether the content validity already good or not.

**Table 3.1. Table of Specification**

Aspects of Reading	Number of Items
Identifying main idea	6, 12, 18, 23, 27, 33, 36, 40, 43, 48
Identifying details	1, 2, 3, 13, 20, 21, 35, 42, 50
Making predictions	10, 15, 25, 29, 31, 44, 45, 46
evaluation	4, 5, 9, 11, 14, 16, 17, 22, 26
Determining inference/conclusion	7, 8, 19, 24, 28, 30, 32, 34, 37, 38, 39, 41, 47, 49

**b. Construct validity**

Construct validity was concerned with whether the test was actually in line with the theory of what it means to determine the language that will be measured. If a test had construct validity, it would have capable of measuring certain specific characteristics in accordance with a theory of language behaviour and learning. This type of validity assumed the existence of certain learning theories or constructs underlying the acquisition of abilities and skills (Heaton, 1988). This research used five aspects of reading: identifying the main ideas, identifying the specific details or information, making prediction, making evaluation and determine the inference.

**2) Reliability**

Reliability can be defined as the extent to which a test produces consistent result when administered under similar condition (Hatch and Farhady, 1982). There were three major important functions of reliability and its analysis. First, they may be used to construct reliable measurement scales, secondly, to improve existing scale, and finally to evaluate the reliability of scales already in use. Specifically, Reliability and item Analysis was aid in the design and evaluation of sum scales

that was scales that are made up of multiple individual measurements (e.g., different items, repeated measurements, different measurement devices, etc.) Numerous statistics can be computed to allow building and evaluating scales following the so-called *classical testing theory* model.

Reliability can be defined as the extent to which a test produces consistent result when administered under similar condition (Hatch and Farhady, 1982). The test can be determined by using Pearson Product Moment which measures the correlation coefficient of the reliability between odd and even number (reliability of half test) in the following formula:

$$r_{xy} = \frac{\sum xy}{\sqrt{(\sum x^2)(\sum y^2)}}$$

where:

$r_{xy}$  : coefficient of reliability between odd and even numbers items

$x$  : odd number

$y$  : even number

$\sum x^2$  : total score of odd number items

$\sum y^2$  : total score of even number items

$\sum xy$  : total score of odd and even number

After getting the reliability of half test, the researcher will use “Spearman Brown’s prophecy formula” (Hatch and Farhady, 1982:247) to determine the reliability of the whole tests, as follows:

$$r_k = \frac{2 r_{xy}}{1 + r_{xy}}$$



where:

$r_k$  : the reliability of the whole tests

$r_{xy}$  : the reliability of half test

(Hatch and Farhady, 1982:247)

The criteria of reliability are as follows:

0.701– 1.000 = high/good

0.401– 0.700 = Average/sufficient

0.000– 0.400 = low/not sufficient

(Suparman, 2011)

Practically, the reliability of the test in this research was analyzed by using ITEMAN. The result of the reliability can be seen in Appendix 9. It was 0.781. The criteria was 0.701– 1.000 = high/good, 0.401– 0.700 = Average/sufficient, 0.000– 0.400 = low/not sufficient. So that, it can be concluded that the reliability of the test was high/good.

### **3) Level of Difficulty**

Level difficulty (LD) level of an item shows how easy or difficult that particular item will be done by the participants (Heaton, 1991). Level of item difficulty also can be classified as the percentage of the students taking of the test who answered the item correctly. In short, it can be stated that the larger the percentage getting

an item right, the easier the item. The higher the difficult index, the easier the item is understood to be. Matlock-Hetzel (1997) states that to compute the item difficulty. The examiner can divide the number of students answering the item correctly by the total number of students answering item. The proportion for the item is usually denoted as  $p$  and is called difficult. An item answered correctly by 85% of the examinees would have an item difficulty, or  $p$  value, of .85, whereas an item answered correctly by 50% of the examinees would have a lower item difficulty, or  $p$  value, of .50%.

The easiest to measure the level of difficulty of an item is by using proportional scale or proportion correct ( $p$ ), that is, the number of test takers answering correctly on the items under analysis is compared with the total number of test takers. The equation was as follows:

$$P = \frac{\sum B}{N}$$

Where:

$P$  : the proportion of test takers who answer correctly a certain item under analysis

$\sum B$  : the numbers of test takers who answer correctly

$N$  : the total number of test takers

The criteria were as follows:

0.000 – 0.099 very difficult/need total revising

0.099 – 0.299 difficult/need revising

0.300 – 0.700 average/good

0.700 – 1.000 easy/needs dropping or total revising

(Suparman, 2011)

The results of level of difficulty by analyzing the data using ITEMAN can be described on Table 2:

**Table 3.2. Level of Difficulty of Try out Test**

Number Items	Total	Remark
2, 4, 8, 11, 14, 20, 21, 23, 24, 27, 33	11	Easy items (0.700-0.1000: need revised)
1, 3, 5, 7, 10, 12, 13, 15, 16, 17, 18, 25, 26, 29, 31, 32, 34, 36, 39, 40, 41, 42, 44, 47, 48, 49, 50	27	Average (0.300-0.700: good)
46, 45, 43, 38, 37, 35, 30, 28, 22, 19, 9, 6	12	Difficult items (0.099-0.299: need revising)

Based on Table 3.2, the difficult items and easy items were revised, while the average items were administered in pre-test and post test without any revision.

Here are the examples of easy, average and difficult items.

1. Why did Malin's mother curse him?
  - a. Because she did not realize that Malin is her son.
  - b. Because Malin was shy to recognize that the poor woman was his mother.
  - c. Because Malin became a rich man and the mother hated it.
  - d. Because she hated Malin for getting married without her permission.

The item above was the item number 2 in try out test. It was an example of easy item. The level of difficulty was 0.73. Then, the criteria is  $0.700 - 1.000 =$  easy/needs dropping or total revising. It means that the item was total revised.

2. Why did the girl want to leave the guy?
  - a. Because her parents did not like the guy.
  - b. Because the girl did not like a poor man.
  - c. Because the girl did not want to see the guy sad because for her illness.
  - d. Because the girl wanted to make the guy become a rich man after he leaves her.

The item above was the item number 7 in try out test. It was an example of average item in the level of difficulty. Its level of difficulty was 0.47. The criteria is  $0.300 - 0.700 =$  average/good. It means that the item was good and it was directly administered.

3. Below are some factors that the landslide often destroyed many houses and buildings and sometimes there was many victims in this accident, except?
- There is no warning before that there will be a landslide.
  - The landslide happened in sudden and unpredictable.
  - The landslide is natural disaster.
  - The landslide often happened in short time.

The item above was the item number 46 in the try out test. It was an example of difficult item. The level of difficulty was 0.23. The criteria is  $0.099 - 0.299 =$  difficult/need revising. It means that the item was difficult and it needed to revise.

#### 4) Discrimination Power

Discrimination power referred to the extent in which the items were able to differentiate between high and low level students on the test. Discrimination power was used to differentiate between the students who had high ability and those who had low ability. A good item according to the criteria was one which good students done well and bad students were fail. To determine the discrimination power, the researcher used the following formula:

$$DP = \frac{U-L}{\frac{1}{2}N}$$

Notes:

DP : discrimination power

U : the number of upper class who answer correctly

L : the number of lower class who answer correctly

N : the total number of the students in upper and lower classes

The criteria are:

0.199 – = very low/need dropping total revising

0.200 – 0.299 = low/need revising

0.300 – 0.399 = quite average/without revision

0.400 = high  $\geq$  D/very good

(Suparman, 2011)

The results of level of difficulty of try out were analyzed by using ITEMAN. It was described on Table 4.3:

**Table 3.3. Discrimination Power of Try out Test**

Number of Items	Total	Remark
38, 30, 27, 28, 26, 25, 22, 24, 19, 15, 11, 8, 2	13	Very low
49, 34, 33, 32, 21, 14, 18, 20, 13	9	Low
48, 46, 4, 42, 41, 40, 36, 35, 31, 23, 12, 10, 5	13	Quite Average
3, 1, 6, 7, 9, 16, 17, 29, 37, 39, 43, 44, 45, 47, 50	15	High

Based on table 3.3, in discrimination power, there were 13 items which was very low items. It means that these items should be dropped or total revised. Then, there were 9 items which was low items were revised. It meant that these items should be revised. 13 quite average items and 15 high items. The items had negative discrimination power were revised, in which the average items were administered in pre-test and post test. Here are the examples of very low, low, quite average and high items.

1. Which statement is not true about the article?
  - a. Dieng Plateau lies in Banjarnegara regency, Central Java.
  - b. Dieng is popular for the traditional tangled hair-cutting ritual.
  - c. The ritual in Dieng this year was held two months ago, attracting at least 25,000 foreign and domestic tourists.
  - d. Dieng Plateau in Banjarnegara regency is crowned as the province's best tourist site for 2012.

The item above was the item number 38 in the try out test. It was an example of very low item. The discrimination of power was 0.12. The criteria is  $0.199 - =$  very low/need dropping or total revising. It means that the item was very low and it was total revised.

2. Why does the government make a health insurance card?
  - a. Because Indonesia can allocate more financial resources to health.
  - b. Because many Indonesian people get sick.
  - c. Because JKN can be good alternative to the citizen when they are sick.
  - d. Because in the other world there has been a health insurance card too.

The item above was the item number 49 in the try out test. It was an example of very low item. The discrimination of power was 0.26. The criteria is  $0.200 - 0.299 =$  low/need revising. It means that the item was low and it needed revised.

3. What is the meaning of the title "Let Me Love You" based on the story.
  - a. That is the guy's heart to the girl that he will always love the girl with her every condition.
  - b. The guy's love was refused by the girl.
  - c. The guy who never got the girl love.
  - d. The girl's heart who always loved her parents.

The item above was the item number 10 in the try out test. It was an example of very low item. The discrimination of power was 0.30. The criteria is  $0.300 - 0.399 =$  quite average/without revision. It means that the item was quite average and it was administered without any revision.

4. How can Ciliwung River cause flood?
  - a. Because the rain intensity is high.
  - b. Because there are many buildings near it
  - c. Because of the trash and residents near it.
  - d. Because the volume of water is high.

The item above was the item number 29 in the try out test. It was an example of very low item. The discrimination of power was 0.40. The criteria is  $0.400 = \text{high} \geq D/\text{very good}$ . It means that the item was very good and it was directly administered.

After analyzing the level of difficulty and discrimination power, it was found that 31 items were good and administered for pre-test and post test without revision, while 9 items were administered after revising the items. On the other side, 10 items, i.e. 8, 19, 20, 22, 25, 26, 27, 28, 30, and 38, were bad and dropped because they not fulfil the criteria of level of difficulty and discrimination power. The level of difficulty was difficult and easy, while discrimination power was very low. For more detail see Appendix 10. Here is the example of the item that was dropped.

1. What is the main problem of the article?
  - a. Many residents built on the bank Ciliwung River has to be cleaned-up.
  - b. Ciliwung River has made flood annually.
  - c. Ciliwung River must be cleaned-up from the building built near it and trash.
  - d. Many residents built on the bank Ciliwung River has to be relocated.

The item above was the item number 28 in the try out test. It was an example of the item that was dropped. The level of difficulty was 0.23 (difficult) and the discrimination of power was 0.12 (very low). In line with this, the item was dropped.

### **3.4.3 Research Procedures**

In constructing the research, the research procedures used these following steps:

### **1. Determining the problems**

The researcher had done pre-observation in MA Tri Bhakti at-Taqwa. She found some problems that had been discussed in chapter 1. From those problems, this research concerned about the material that was used by the teacher and the effective strategy that was used by the students in comprehending an English text.

### **2. Determining the population and sample of the research**

The population of this research was in the second grade students of MA Tri Bhakti at-Taqwa in the second semester. There were three classes and each class consisted of 30. Two classes were taken of this research, the first class as the experimental class one, and the second class as the experimental class two, that was class XI-A that consisted of 30 students and XI-B that consisted of 30 students. The two sample classes were selected randomly by using lottery.

### **3. Preparing the material**

The researcher used two kinds of reading text. The materials were three short stories as literary text and three articles as non-literary text.

### **4. Administering the try-out to determine the quality of the test**

Try-out test used an objective test in form of 50 items and have four options a, b, c and d in 90 minutes. This test was conducted in order to determine the level of difficulty, discrimination power, reliability and validity of the test items before giving pre-test and post test to the class.



## **5. Administering the pre-test**

The pre-test was administered in order to investigate the students' reading entry point before the treatments. The type of the test was multiple choice in which the students was asked to choose one correct answer from the options a, b, c, or d. In this pre-test the students was given 20 items of reading comprehension and it was conducted within 60 minutes.

## **6. Giving treatment**

After administering the pre-test, the students in both class: experimental class one and experimental class two were given three treatments. The experimental class one was treated by using literary text through critical reading approach then in the experimental class two was treated by using non-literary text through critical reading approach.

## **7. Administering the post-test**

The post-test was administered to both classes after giving the treatment to statistically determine whether there was any significant improvement in the reading comprehension ability of the participants in the experimental class one and experimental class two. This test consisted of 20 items of multiple choices within 60 minutes.

## **8. Administering Interview**

Interview was administered to identify the student's responses after being taught by using literary text and non-literary text through critical reading approach. There were 5 students of each class that were interviewed. The interview

consisted of fifteen questions about the opinion of students of the use of short story as literary texts and article as non-literary text and the students' opinion about the method that was applied in the treatment class, which was through critical reading approach in teaching reading comprehension. In the experimental class one who use literary text, the students was interviewed about their feeling when they were reading literary texts and their opinion when they used critical reading approach in reading literary texts. Meanwhile, in experimental class two who use non-literary text, the students was interviewed about their feeling when they were reading non-literary texts and their opinion when they used critical reading approach in reading literary texts.

## **9. Analyzing data**

After the data were processed, the mean of the two classes were computed and compared by Independent group t-test. It was used to statistically determine whether there was any significant improvement on the reading comprehension ability of the participants in the experimental class one and experimental class two.

### **3.4.4 Scoring System**

The researcher used multiple choices in order to gain the objectivity of the result. There were 50 items in try out test, 20 items in pre-test and 20 items in post-test. In scoring the students' work, this research employed Henning's formula.

The ideal highest score was 100. The score of pre-test and post test was calculated by using the following formula:

$$PS = \frac{R}{N} \times 100$$

Where:

PS : percentage of score

R : number of the right answer

N : total number of items on test

(Henning: 1987)

### 3.4.5 Data Analysis

In order to determine the students' progress in comprehending the text, the students' score computed by doing three activities:

1. Scoring the pre-test and post test
2. Tabulating the result of the test and finding the mean of the pre-test and the post test. The mean was calculated by applying the following formula:

$$M = \frac{\sum x}{N}$$

Notes:

M = mean (average score)

$\sum x$  = the total students' score

N = total number of students

(Hatch and Farhady, 1982)

### 3. Administering the Normality Test

Normality test is an idealized model which can be used to dealing with natural behaviour (Hatch and Farhady, 1982). Normality test has been used to identify whether the data is normally distributed or not. The data were tested by One-sample Kolmogorov-Smirnov Formula (SPSS 16.0).

The criteria are:

H0 = L-ratio is lower than L-label (the distribution is normal)

H1 = L-ratio is higher than L-label (the distribution is not normal)

4. Drawing conclusion from the tabulated results of pre-test and post-test administered, that was by statistically analyzing the data using statistical computerization, i.e., Match t-test of Statistical Package for social Science (SPSS) for Windows to test whether the improvement gained by the students were increase or not, in which the significance was determine by  $p > 0.05$ .

#### 3.4.6 Hypothesis Testing

The post test in experimental class one and experimental class two was compared in order to determine the gain. The researcher used *T-test* and *Pair t-test* towards the mean score of the post test in experimental class one and the mean score of the post test in experimental class two. The aim of *T-test* and *Pair t-test* was to compare two kinds of data or mean from the different sample. Moreover, the result of *T-test* and *Pair t-test* was used to determine whether there was any significant difference between the students' reading achievement who read literary text and those who read non-literary text through critical reading and to prove whether the proposed hypothesis was accepted or rejected. In this case, the

researcher used significant level of 0.05 in which that the probability of error in the hypothesis was only about 5%. The hypotheses were drawn as follows:

1.  $H_0$ : There is no significant difference between the students' reading achievement who read literary text and those who read non-literary text through critical reading approach.

$H_1$ : There is a significant difference between the students' reading achievement who read literary text and those who read non-literary text through critical reading approach.

The criteria for accepting the hypotheses are as follows:

1.  $H_0$  is accepted if the t-value is lower than T-table.

It means that there is no significant difference between the students' reading achievement who read literary text and those who read non-literary text through critical reading approach.

2.  $H_0$  is rejected if the t-value is higher than T-table.

It means that there is a significant difference between the students' reading achievement who read literary text and those who read non-literary text through critical reading approach.

### **3.5 Research Question 2**

This part is describing the procedure of answering the second research question: What are the student's responses after being taught by using literary text and non-literary text through critical reading approach?

To identify the student's responses after being taught by using literary text and non-literary text through critical reading approach, the researcher administered interview. There were 5 students of each class that were interviewed. The interview consisted of fifteen questions about the opinion of students of the use of short story as literary texts and article as non-literary text and the students' opinion about the method that was applied in the treatment class, which was through critical reading approach in teaching reading comprehension. In the experimental class one who use literary text, the students was interviewed about their feeling when they were reading literary texts and their opinion when they used critical reading approach in reading literary texts. Meanwhile, in experimental class two who use non-literary text, the students was interviewed about their feeling when they were reading non-literary texts and their opinion when they used critical reading approach in reading literary texts.

In analyzing interview, the researcher used descriptive analysis. This method was used to describe the data, such as interview. The researcher described the result of interview of the interviewees. Then she made a conclusion after analyzing the data of interview.

### **3.6 Research Schedule**

This research was conducted based on sequenced schedule in order to make this research runs well. Before the research was carried out, pre observation was done on Saturday, November 22<sup>th</sup>, 2014. This pre observation was conducted in order to investigate the students' problems in reading comprehension at MA Tri Bhakti

at-Taqwa. Then, try out test about reading comprehension of literary and non-literary text was administered in class XI Science 3 as try out class to determine the content and construct validity of the text, also the level difficulty and the discrimination power of its. The pre test was carried out in class XI Science 1 as experimental class one and XI Science as experimental class two in order to know the students' achievement of reading comprehension before being given treatments.

The pre-test was administered on January 26<sup>th</sup>, 2015. After that, three times treatments were administered in both classes on January 27<sup>th</sup>, 2015 until February 3<sup>th</sup>, 2015, then, the post test was given in order to know the difference of the students' reading comprehension achievement on Tuesday, February 9<sup>th</sup>, 2015. After that the interview was administered on February 10<sup>th</sup>, 2015. The schedule of the research can be seen on Appendix 1.