

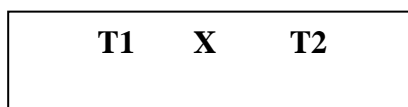
III. RESEARCH METHOD

This chapter describes the design of the research, subject of the research, and how the data collecting technique of the research. This chapter also describes research procedure, scoring system, analysis research instrument, how to analyze data, and hypothesis testing.

3.1 Research Design

This research was investigated in order to find out a difference of students' reading comprehension achievement before and after being taught through STAD technique and to investigate the students' activities in teaching learning process using STAD technique. In conducting this research, the writer used *one group pretest posttest design* (Hatch and Farhady, 1982:20). Pretest and posttest were administered to see whether STAD technique can be used to improve students' reading comprehension achievement.

This research used one class. The class has both pretest and posttest and three times treatment. The design can be illustrated as follows:



Notes:

T1 : Pretest

X : Treatment (using STAD technique)

T2 : Posttest

(Hatch and Farhady, 1982:20)

The pre-test was administrated before the treatment implemented, to see the students' basic reading comprehension. Then, the class was given the treatment of teaching reading comprehension through STAD technique. The post-test was administrated afterward, to analyze the difference of students' reading comprehension achievement before and after being taught through STAD technique.

3.2 Subject of the Research

The research was conducted at the eleventh grade of senior high school at SMAN 1 Gunung Pelindung in the academic year 2013. There were three classes of the eleventh grade students and each class consisted of 28-31 students. One class (XI IPA) was taken as sample that was given treatment (teaching reading using STAD technique), and one class (XI IPS 2) was taken as control class to administer try out. The classes were chosen randomly by using lottery since there is no stratified and priority class. So that, all the second grade got the same chance to be the sample in order to avoid subjectivity and to guarantee that every class had the same opportunity.

3.3 Data Collecting Technique

In collecting the data the writer used the following technique:

1. Pre Test

The pretest was administered in order to find out the students' reading comprehension achievement before giving treatment.

2. Post Test

Posttest was administered at the end of treatments in order to find out the results of students' reading comprehension achievement after the three-time treatments.

3. Observation

The observation is conducted to investigate the students' activities in teaching learning process using STAD technique. The observation sheet was used to find out the students' attention to the teacher's, the students' explanation responding to the teacher instruction and question, and also students' group activity. Raters were used in collecting the data to ensure the reliability of the observation and to avoid the subjective of the research. In case that the raters were two English teachers at SMAN 1 Gunung Pelindung, who observed the class during teaching learning process.

3.4 Research Procedures

In collecting the data, the writer carried out the following procedures which can be described as follows:

1. Determining the subject

The subject of this research was the eleventh grade students of SMA Negeri 1 Gunung Pelindung. The writer chosen one class from three classes in the eleventh year that took randomly by using lottery.

2. Selecting and determining the materials

The materials were based on the School Based Curriculum (KTSP) 2006 for the eleventh grade students. They were taken from the students' text books and internet. As it was discussed in Chapter 1, this research focussed on narrative text.

3. Determining the Research Instrument

The instrument of this research was objective text of multiple choices test. This was supported by Henning (1975), who stated that to measure reading comprehension, requesting students to write short sentence answers in written questions were less valid procedure than multiple choice selection (as cited in Henning, 1987:48). The instruments were used in order to support this research. The test consisted of two sessions, the first was pre test and the second one was post test. Each test consisted of 40 items of multiple choices of comprehension questions and some reading texts. The question has four alternative answers for each (A, B, C, and D), one was correct answer and rests were distractors.

4. Administering Try-out Test

The try-out test was administered for 50 items in 90 minutes. The writer administered the try-out using reading texts with 50 items of multiple choices with four option (A, B, C, and D).

5. Administering Pre-Test

Pre-test was administered to reveal the students' basic reading comprehension before treatments. The test was administered with 40 items of multiple choices reading test.

6. Conducting the Treatments

The treatments was classroom activities which applied STAD technique. The students were taught three times. During the treatment, the observers helped in observing the students' activities in the teaching learning process.

7. Administering Post-Test

Post-test was given at the end of treatments in order to find out the significant in improve students' reading comprehension achievement. The test was administered with 40 items of multiple choices reading test.

8. Analyzing the result of the Test

The data were analyzed by comparing the average score (mean) of the pre test and post test to know whether there was a difference of students' reading comprehension achievement before and after being taught through STAD technique.

3.5 Scoring System

In scoring system students' result of the test, the writer used percentage correct of Lyman's formula. The score of pre test and post test were calculated by using formula as follow:

$$X\%c = 100 \cdot \frac{R}{T}$$

Notes:

X%c : percentage of correct score

R : number of the right answer

T : total number of items on test

(Lyman, 1971: 95)

After administering the test and giving an individual score, researcher gave a point to each group which can be carried out the following procedures:

1. Giving an individual point

According to Slavin (Trianto, 2007: 55), to account the improvement of individual score can be done as follow:

Table 2.Improvement of individual score calculation

No.	Score	The Improvement Score
1.	More than 10 points < standard score	0 point
2.	10 - 1 point < standard score	10 points
3.	0 – 10 points > standard score	20 points
4.	More than 10 points > standard score	30 points
5.	Maximal score	30 points

2. Accumulating the group score

The group's score could be accumulated with counting all improvement individual score and divide it with total members of the group. Based on those average score, we could get the group score as follow:

Table 3.Improvement of group score calculation

No.	Average Score	Qualification
1.	$0 \leq N \leq 5$	-
2.	$6 \leq N \leq 15$	Good Team
3.	$16 \leq N \leq 20$	Great Team
4.	$21 \leq N \leq 30$	Super Team

3.6 Criteria of Good Test

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

1. Validity

A test is considered valid if the test measures the object to be measured and suitable with the criteria (Hatch and Farhady, 1982:250). Heaton (1988:159) also states that validity of the test is the extent to which it measures what it is supposed to measure. A test must aim to provide true measure of the particular skill which it is intended to measure.

There are four types of validity that are: (1) face validity, concerns with the lay out 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 validity above, the writer used content and construct validity because the other two were considered to be less needed. Both of them were explained as follows:

1. Content Validity

Content validity was concern with whether the test was sufficiently representative and comprehensive for the test. The focus of the content validity was adequacy of the sample and not simply on the appearance of the test (Hatch and Farhady, 1982:251). Content validity could be found by relating the material of the test with the curriculum for Senior High School. It meant that the test was designed based on the curriculum in the school. In this case, to know whether the instrument had fulfilled the criteria of content validity the writer had checked in Competence-based English Developing competencies in English for second grade Senior High School. The writer had also consulted the instrument test to the English teacher at the school that had been chosen as a place for research.

2. Construct Validity

Construct validity was concerned with whether the test was actually in line with the theory of what it means to know the language (Shohamy, 1985: 74).

The writer formulated table of specification, so every test items can be matched with the goal and the materials have been taught. The table of specification was an instrument that helps the test constructor plans the test. The test was based on 2006 English curriculum, and the syllabus of second Grade of students Senior High School and represent of the materials that will be taught by the teacher. The content of the test was presented in the table of specification below:

Table 4. Table of specification of Data Collecting Instrument

No.	Skills of Reading	Item Number	Percentage of Item
1	Identifying the main idea	2, 10, 12, 21, 26, 32, 33, 38	16 %
2	Specific information	1, 6, 8, 9, 11, 15, 17, 31, 40, 41, 44, 46, 47.	26 %
3	Reference	3, 20, 27, 37, 39, 42, 49.	14 %
4	Inference	4, 7, 13, 14, 16, 18, 19, 23, 24, 29, 30,34, 45, 50.	28 %
5	Vocabulary	5, 22, 25, 28, 35, 36, 43, 48.	16 %
Total		50	100%

2. Reliability

Shohamy (1985:70) states that reliability refers to the extent to which the test was consistent in its score, and it gave an indication of how accurate the test score. The writer used split-half method to estimate the reliability of the test, since the formula was simple. It was because (1) it avoids troublesome correlation and (2) in addition to the number of item in the test, it involves only the test, mean and standard deviation, both of which are normally calculated anyhow as a matter of routine. To obtain the reliability of the instrument, this study used SPSS to analyze the reliability of the test. The steps to run the reliability in SPSS were as follows:

Entering the data → Analyze → Scale → Reliability Analysis. To measure the coefficient of the reliability the first and second half group, the writer used the following formula:

$$r_l = \frac{\sum XY}{\sqrt{[\sum X^2][\sum Y^2]}}$$

Notes:

r_l : coefficient of reliability between the first half and the second half items

X : total numbers of odd numbers items

Y : total numbers of even numbers items

X^2 : square of X

Y^2 : square of Y (Lado in Hughes, 1991: 3)

The writer uses *Spearman Brown's Prophecy Formula* (Hatch and Farhady, 1982: 247). The formula is as follows:

$$rk = \frac{2rl}{1 + rl}$$

Notes:

rk : the reliability of the test

rl : coefficient of reliability between the first half and the second half items

(Hatch and Farhady, 1982: 247)

The criterion of reliability is:

0.90 – 1.00 : high

0.50 – 0.89 : moderate

0.0 – 0.49 : low

To know the result of reliability of try-out test, the writer used Pearson Product Moment. The result showed that the reliability of the test was 0.99 (**see appendix 7, p.80**). According to the criteria of the reliability proposed by Hatch and Farhady (1982:268), the test has high reliability in the range 0.90-1.00. It indicated that the instrument of this research was reliable and good.

3. Level of Difficulty

Difficulty level related to how easy or difficult the item was from the point of view of the students who take the test. It was important since the items, which are too easy (that students get right) told us nothing about differences within the test population. To see the level difficulty, the writer used the formula as follow:

$$LD = \frac{U + L}{N}$$

Notes:

LD : level of difficulty

U : the proportion of upper group students who answer correctly

L : the proportion of lower group students who answer correctly

N : total number of students

The criteria are:

< 0.30 = difficult

0.30-0.70 = average

>0.70 = easy

(Shohamy, 1985:79).

From the computation of level difficulty (**see appendix 6.p.79**), it was found out that 5 items which were less than 0.30 (items number 20, 26, 28, 35, 43). It means that the items were difficult. There were 5 items which were higher than 0.70 (items number 2, 3, 4, 16, 24). It means that the items were easy and 40 items were average (0.30-0.70) (items number 1, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 17, 18, 19, 21, 22, 23, 25, 27, 29, 30, 31, 32, 33, 34, 36, 37, 38, 39, 40, 41, 42, 44, 45, 46, 47, 48, 49, 50).

4. Discrimination Power

This index referred to the extent to which the item differentiates between high and low levels students on the test. A good item according to this criterion was one that good students do well on and bad students fail. To see the discrimination index, the writer used the following formula:

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

(Shohamy, 1985:81)

Notes:

DP : discrimination power

U : the proportion of upper group students who answer correctly

L : the proportion of lower group students who answer correctly

N : total number of students

The criteria are:

1. If the value is positive discrimination – a large number or more knowledgeable students than poor students get the item correct. If the value is zero, it means that there is no discrimination.
2. If the value is negative, it means that more low students than high level students get the item correct.
3. In general, the higher, the discrimination index, the better. In classroom situation most items should be higher than 0.20 indexes. (Shohamy, 1985:82).

From the computation of discrimination power (see **appendix 6.p.79**) the writer got that there were 5 bad items (has negative value in discrimination in number 4, 12, 16, 35, 43), 14 items were poor (has less than 2.00 index in number 8, 10, 15, 17, 19, 25, 26, 29, 30, 36, 40, 46, 48, 50), 3 item had no discrimination (the value is zero) and 28 items were satisfactory (has higher than 2.00 index in number 1, 2, 3, 6, 7, 9, 11, 14, 18, 20, 21, 22, 23, 27, 28, 31, 32, 33, 34, 37, 38, 39, 41, 42, 44, 45, 47, 49)

3.7 Data Analysis

The data were analyzed in order to determine whether the students' reading comprehension achievement was improved or not. The writer examined the students' score by doing the following steps:

1. Scoring the pretest and posttest.
2. Tabulating the results of the tests and calculating the scores of the pretest and posttest.
3. Drawing conclusion from the tabulated results of the pretest and posttest which statistically analyzed using *Repeated Measure T-Test* computed through SPSS version 17.0.

Furthermore, in analyzing the data from the observation of students' learning activities, the writer counted the number of activities done by the students and then calculated the percentage of the students' activities.

The formula is as follows:

$$\% A = \frac{A \times 100\%}{n}$$

Where :

%A : percentage of students 'activities

A : number of students' activities observed

n : number of students in the class

Arikunto (2006:7)

The indicator of the students' activities was more than 75%. If more than 75% students were actively involved in teaching learning activities, it mean that the teaching learning process can be categorized as a good level (Arikunto, 2006:7).

3.8 Normality Test

The normality test was counted by using SPSS. This test was used to measure whether data in experiment class are normal distributed or not. To run normality test, the steps were as follows:

Entering the data → analyze → Descriptive Statistics → Explore.

To see the normality, it could be seen in three ways; variants coefficient, Skewness ratio, and Kurtosis Ratio.

3.8.1 Variants Coefficient

The data was called as normal distribution data if the variants coefficient < 30 %.

This value should be counted first with this formula:

$$\text{Variants Coefficient} = \frac{\text{Std Deviation}}{\text{Mean}} \times 100 \%$$

3.8.2 Skewness Ratio

The data was called as normal distribution data if the Skewness ratio was in the range of -2 until 2. The formula to count this was as follows :

$$\text{Skewness ratio} = \frac{\text{Skewness}}{\text{Skewness Std.error}}$$

3.8.3 Kurtosis Ratio

It was almost the same as Skewness Ratio. The data called as normal distribution data if the range was between -2 and 2. The formula was:

$$\text{Skewness ratio} = \frac{\text{Kurtosis}}{\text{Kurtosis Std.error}}$$

3.9 Hypothesis Testing

The hypothesis was analyzed by using *Repeated Measure t-test* in order to know the level of significance of the treatment effect. By seeing the probability level (p) which was shown by two tail significance as the value of significance, we can draw the conclusion (Setiyadi, 2006: 172). H_1 was approved if $p < \alpha$. The research used level of significance i.e. 0.05, and the probability of error in the hypothesis was 5%.

Therefore, the hypothesis which would be cited was as follows:

H_1 : There is a difference of students' reading comprehension achievement before and after being taught through STAD technique.

H_0 : There is no difference of students' reading comprehension achievement before and after being taught through STAD technique.

Besides that, to investigate the students' activities in teaching learning process using STAD technique, the writer analyzed the observation data and concluded the result after the observation sheets completed by the observers.