CHAPTER III

RESEARCH METHODOLOGY

A. The Research Design

The research design refers to the conceptual structure within which research has been conducted. This is important because it facilitates research to be as efficient as possible yielding maximal information. The type of this research was an experimental research. Creswell stated that there are three kinds of experimental design; true, quasi, and pre-experimental research. According to Creswell (2008:298) an experimental research is:"the traditional approach to conduct quantitative research, whether it influences an outcome or dependent variable ". Cresswell (2008:299) also stated that experiment is testing an idea to determine whether it influences an outcome or dependent variable.

This research used a quasi-experimental design: the pretest-post-test, non-equivalent group design. The researcher uses the quasi-experimental design because in this research the researcher uses all of the subject in group to get a treatment. In this research, there were two groups; experimental and control group. The researcher gave different treatments to experimental and control groups. In experimental group, the researcher gave treatment using STAD, but in control group conventional teaching was used. There were two variables in this research, the first variable was Student Team Achievement

Division (STAD) method as independent variable (x variable), speaking skill as dependent variable (y variable).

Therefore the experimental class was provided with pre-test, treatment, post-test. The quasi-experimental design: the pretest-post-test, non-equivalent group design can be presented as follows (Cohen, 2007:283):

Table III.1
The Research Design

Group	Pre-Test	Treatment	Post-Test
Experimental	O1	X	O2
Control	O3	-	O4

Where:

O1	= Pre-test of Experimental Group
X	= Treatment of Experimental Group
O2	= Post-test of Experimental Group
O3	= Pre-test of Control Group
O4	= Post-test of Experimental Group

B. The Time and Location of the Research

This research was conducted from April to May 2016. This research was located at Islamic Senior High School Pesantren Teknologi Riau Pekanbaru.

C. Subject and Object of the Research

The subject of this research was the eleventh grade students of Islamic Senior High School Pesantren Teknologi Riau Pekanbaru in the academic year 2015/2016. Besides, the object of this research was the effect of using Student Team Achievement Division (STAD) Method on students' speaking skill of

the eleventh grade at Islamic Senior High School Pesantren Teknologi Riau Pekanbaru.

D. Population and Sample of the Research

Population refers to the members of any well-defined class of people, events or objects (Ary et al, 1985:148). The population of the research was the eleventh grade students of Islamic Senior High School Pesantren Teknologi Riau Pekanbaru. They consisted of two classes, XI Science and Xi Social. The total number of population was 28 students.

Table III.2
The Population of the Eleventh Grade Students of State Islamic Senior High School Pesantren Teknologi Riau Pekanbaru

No	Class	Total
1	XI Science	14 Students
2	XI Social	14 Students
	Total	28 Students

In order to select a sample for the research, the researcher decided the way of selecting a sample. In this case, the researcher used total sampling as the method sampling of the research. Acoording to Sugiyono 2007 (in Suryani and Hendryadi 2015:203) The researcher took the whole of the available population as the sample of the research because the number of the population was not so large.

Table III.3
The Sample of the Second Grade Students of Islamic Senior High
School Pesantren Teknologi Riau Pekanbaru

No.	Class	Number of Students	Group
1.	XI Science	14	Control
2.	XI Social	14	Experiment
Total 28 Students			

E. The Technique of Collecting Data

Collecting data is the most crucial thing in conducting a research. In this research, the researcher used test and observation in collecting the data. The test used was oral presentation. In this research, taking the score of students in that way was the most common format for the testing of oral presentation. The teacher divided the students into four groups. Each group consisted of three to four students. Each group was given different spoof texts. Teacher asked the students to discuss about the text with their friends in group for 40 minutes. The activities such as; understanding the spoof text, finding the generic structure of spoof text, and take turn to explain the spoof text in their group. After 40 minutes, the teacher asked the students to present the spoof text in front of the class orally one by one. The teacher used tape recorder to record students' speaking skill. The researcher also invited two raters to score and analyzed the students' speaking skill. The following table describes the final score of the students from the two raters:

Table III.4
The Final Score Of Students' Speaking Skill
In Pre-Test Of Experimental Class

NO	LIST OF STUDENTS	PRE-TEST SCORE		FINAL SCORE
		RATER 1	RATER 2	
1	STUDENT 1	75	80	77.5
2	STUDENT 2	75	80	77.5
3	STUDENT 3	95	85	90
4	STUDENT 4	65	85	75
5	STUDENT 5	55	75	65
6	STUDENT 6	80	85	82.5
7	STUDENT 7	85	85	85
8	STUDENT 8	65	70	67.5
9	STUDENT 9	95	85	90
10	STUDENT 10	80	85	82.5
11	STUDENT 11	75	70	72.5
12	STUDENT 12	70	80	75
13	STUDENT 13	70	80	75
14	STUDENT 14	75	75	75
TOTAL		1060	1120	1090

Table III.5
The Final Score Of Students' Speaking Skill
In Post-Test Of Experimental Class

NO	LIST OF STUDENTS	POST-TEST SCORE		FINAL SCORE
		RATER 1	RATER 2	
1	STUDENT 1	95	90	92.5
2	STUDENT 2	95	90	92.5
3	STUDENT 3	100	90	95
4	STUDENT 4	85	90	87.5
5	STUDENT 5	80	90	85
6	STUDENT 6	90	100	95
7	STUDENT 7	100	90	95
8	STUDENT 8	95	90	92.5
9	STUDENT 9	100	100	100
10	STUDENT 10	100	95	97.5
11	STUDENT 11	95	90	92.5
12	STUDENT 12	95	100	97.5
13	STUDENT 13	90	85	87.5
14	STUDENT 14	95	85	90
	TOTAL	1315	1285	1300

Table III.6
The Final Score Of Students' Speaking Skill
In Pre-Test Of Control Class

NO	LIST OF STUDENTS	PRE-TEST SCORE		FINAL SCORE
		RATER 1	RATER 2	
1	STUDENT 1	75	60	67.5
2	STUDENT 2	55	65	60
3	STUDENT 3	85	85	85
4	STUDENT 4	75	90	82.5
5	STUDENT 5	65	80	72.5
6	STUDENT 6	75	80	77.5
7	STUDENT 7	70	70	70
8	STUDENT 8	50	70	60
9	STUDENT 9	85	80	82.5
10	STUDENT 10	95	85	90
11	STUDENT 11	70	65	67.5
12	STUDENT 12	70	90	80
13	STUDENT 13	70	60	65
14	STUDENT 14	65	80	72.5
TOTAL		1005	1060	1032,5

Table III.7
The Final Score Of Students' Speaking Skill
In Post-Test Of Control Class

NO	LIST OF STUDENTS	PRE-TEST SCORE		FINAL SCORE
		RATER 1	RATER 2	
1	STUDENT 1	80	60	70
2	STUDENT 2	55	70	62.5
3	STUDENT 3	95	90	92.5
4	STUDENT 4	75	90	82.5
5	STUDENT 5	75	80	77.5
6	STUDENT 6	75	80	77.5
7	STUDENT 7	70	75	72.5
8	STUDENT 8	55	80	67.5
9	STUDENT 9	90	80	85
10	STUDENT 10	100	90	95
11	STUDENT 11	75	75	75
12	STUDENT 12	75	90	82.5
13	STUDENT 13	80	90	85
14	STUDENT 14	65	85	75
TOTAL		1065	1135	1100

Pre Test and Post Test contained the same test items. They were just different in time allocation. These pretest and posttest were taken by giving oral presentation. The researcher used Pre Test, treatment, and Post Test.

1. Pretest

The pre Test was administered before using STAD in teaching speaking. It aimed at knowing students' skill in speaking.

Pre-test was used to collect the data about students' speaking skill in spoof text before they were taught by using Student Team Achievement Division (STAD). It was given to both experimental and control. In this test, the researcher used oral presentation.

2. Post test

Post-test was used to collect the data about students' speaking skill in spoof text after they were taught by using Student Team Achievement Division (STAD). It was given to both experimental and control classes. In this test, the researcher also used oral presentation.

After the students did the test, then the researcher took the total score from the result of the speaking skill test. According to Hughes (2003:132) there are some components that should be considered in giving students' score; they are accent, grammatical, vocabulary, fluency, and comprehension. They have typical scale where each component has asset of qualities/level to be rated and a series of possible ranting. He describes the ranting as follow:

Table III.8
Accent

Score	Requirement
1	Pronunciation frequently unintelligible
2	Frequent gross error and a very heavy accent make understanding difficult, require frequently repetition
3	"foreign second" requires concentrated listening, and mispronunciations lead to occasional misunderstanding and apparent errors in grammar of vocabulary
4	Marked "foreign accent" and occasional mispronunciation which do not interfere with understanding

Table III.9
Grammar

Score	Requirement
1	Grammar almost entirely inaccurate except in stock phrases
2	Constant errors showing control of very view major patterns and frequently preventing communication
3	Frequent errors showing some major pattern uncontrolled and causing occasional irritation and misunderstanding
4	Occasional errors showing imperfect control of some pattern but no weaknesses that causes misunderstanding

Table III.10
Vocabulary

Score	Requirement
1	Vocabulary inadequate for even the simplest conversation
2	Vocabulary limited to basic personal and survival areas (time, food, transportation, family, etc.)
3	Choice of words sometimes inaccurate, limitations of vocabulary prevent discussion of some common professional and social topics.
4	Professional vocabulary adequate to discuss special interest, general vocabulary permits discussion of any non-technical subject with some circumlocutions.

Table III.11
Fluency

Score	Requirement
1	Speech is no halting and fragmentary that conversation is virtually impossible
2	Speech is very slow and uneven except for short or routine sentences
3	Speech is frequently hesitant and jerky; sentences may be left uncompleted
4	Speech is occasionally hesitant, with some unevenness caused by rephrasing and grouping for words

Table III.12
Comprehension

Score	Requirement
1	Understand too little for the simplest type of conversation
2	Understands only slow, very simple speech on common social and touristic topics; require constant repetition and rephrasing
3	Understand careful, somewhat simplified speech when engaged in a dialogue, but may require considerable repetition and rephrasing
4	Understand quite well normal educated speech when engaged in a dialogue, but occasional repetition or rephrasing

The result of speaking was scored by using five components and each component had score or level. Each component had 20 as the highest score. The total of all components is 100. The specification of the test is as follows

Table III.13
The Specification of the Test

Criteria	Range of the Score
1	0-5
2	6-10
3	11-15
4	16-20
Total Highest Score	20

Table III.14
The Conversion of the Test

No	Speaking Skill	The Highest Score
1	Accent	20
2	Grammatical	20
3	Vocabulary	20
4	Fluency	20
5	Comprehension	20
	Total	100

There were six meetings based on the syllabus, and the result of speaking was scored by using the scoring rubric, and categorize of the test as follow:

Table III.15
The Students' Score
Category of students' Speaking Skill

No	Score	Category
1	80-100	Very Good
2	66-79	Good
3	56-65	Enough
4	40-55	Less
5	30-39	Fail

(Arikunto, 2009:245)

F. The Validity and Reliability of Test

The test used for testing students' speaking skill had to have reliability and validity. The test is valid if it measures accurately what it is intended to measure. According to Gay (2000:163-167) states that there are three types of validity. They are content validity, criterion-related validity, and construct validity. In this research, the writer used content validity to know the validity of speaking skill test. According to Brown (2003:22)states that content validity is partly a matter of determining if the content that the instruments contains is an adequate sample of the domain of content it is supposed to represent. Thus, the test was given based on the material studied by the students. The material of the test was taken from English Alive text book from Yudhisira, the text book used by the eleventh grade of Islamic Senior High School Pesantren Teknologi Riau Pekanbaru.

The validity and reliability is related. It is possible for a test to be reliable without being valid for a specific purpose, but it is impossible a test to be valid without first being reliable. According to Hughes (1989: 20), a reliable test is consistent and dependable. If the same test is given to the same student or matched students on two different occasions, the test should yield similar results. There are five types of reliability: stability, equivalence, equivalence and reliability, internal consistency, and rater agreement. In this research, to know the reliability of the test the writer used the rater agreement type concerned with inter-rater reliability, because the writer has two raters to score the students' speaking skill.

The following table describes the correlation between the first raters' scores and the second raters' scores by using SPSS 16:

Table III.16

Correlations

		Rater1	Rater2
Rater1	Pearson Correlation	1	.324
	Sig. (2-tailed)		.092
	N	14	14
Rater2	Pearson Correlation	.324	1
	Sig. (2-tailed)	.092	
	N	14	14

**. Correlation is significant at the 0.01 level (2-tailed).

It was necessary to find out the df (degree of freedom) as follow:

df = N – Nr

Where:

df = degree of freedom

N = Number of freedom

Nr = number of correlated variable

$$df = 14 - 2 = 12$$

Df would be correlated at level 5% and 1%. At level 5%, t_{table} was 0.374. While, at level 1% t_{table} was 0.478. Thus, the $r_{obtained}$ (0.324) was higher than r_{table} at level 5% and 1%. It can be read 0.374<0.644> 0.578. So the writer concluded that there was a significant correlation between score given by rater 1 and score given by rater 2. In other words, the written test was reliable. Then, the writer calculated by using Spearman-Brown prophecy formula in order to know the classification of reliability. The formula could be seen as follow:

$$r_{tt} = \frac{nr_{A,B}}{1 + (n-1)r_{A,B}}$$

Where:

r_{tt} = Inter-rater reliability

n   = The number of raters whose combined estimate the final mark

    for the examines

r_{AB}= The correlation between raters, or the average correlation among all raters if there are more than two

While, the calculation of it was as follow:

$$r_{tt} = \frac{n r_{A,B}}{1 + (n-1)r_{A,B}}$$

$$r_{tt} = \frac{2 \ x \ (0.324)}{1+(2-1)x \ 0.324}$$

$$r_{tt} = \frac{0.648}{1.324}$$

$$r_{tt} = 0.48$$

The writer used the categories of reliability that could be seen in the following table, (Zelly, 2011: 35).

Table III. 17
The Categories of Reliability

No	Reliability	Level of reliability
1	0.0 - 0.20	Low
2	0.21 - 0.40	Sufficient
3	0.41 - 0.70	High
4	0.71 - 1.0	Very High

From From result of calculation above, the inter-rater reliability in this research was 0.48. Thus, it was categorized into high level.

G. Technique of Data Analysis

1. Normality and Homogenity

Before analyzing the data by using t-test formula, the researcher had to find out the normality test of the data. The normality test of the data was analyzed by using Kolmogorov-Smirnove technique with SPSS 16 version.

Analysis:

H_o: population with normal distribution

H_a : population with not normal distribution

If the probability > 0.05 H_o was accepted

If the probability < 0.05 H_o was rejected

Then, the writer also had to find out the homogeneity of the test. To analyze the homogeneity was by comparing sig. in Based on trimmed mean with 0.05.

Analysis:

Sig. > 0.05 the data is homogenous

Sig. < 0.05 the data is not homogenous

2. Analysis Data of Independent Sample T- test

In analyzing the data, the writer used the score both pre-test and post-test of the students from experimental and control classes. Those scores were analyzed by using statistical analysis. In order to find out whether there is a significant difference of using Student Team Achievement Division (STAD) Method on students' speaking skill or not using it, the data were analyzed by using Independent sample t-test. Pallant

(2010:239) stated that independent samples t-test is used to compare the mean score of two different groups of people or conditions.In taking the conclusion, the writer concluded by comparing t-observed with significant value (0.05).

Analysis:

H_o: sig. (2 tailed) > significant value

H_a: sig. (2 tailed) < significant value

a. H_o is accepted if sig. (2 tailed) > significant value or there is no significant different of using Student Team Achievement Division (STAD) Method on students' speaking skill at the eleventh grade of Islamic Senior High School Pesantren Teknologi Riau Pekanbaru

b. H_a is accepted if sig. (2 tailed) < significant value or there is a significant different of using Student Team Achievement Division (STAD) Method on students' speaking skill at the eleventh grade of Islamic Senior High School Pesantren Teknologi Riau Pekanbaru

To identify the level effect of using Student Team Achievement Division (STAD) Method on students' speaking skill of the eleventh grade of Islamic Senior High School Pesantren Teknologi Riau Pekanbaru, it was calculated by using eta squared formula. Based on the results presented above shows about the difference obtained in two sets of scores, it was not show about the magnitude of the effect. One way to do this is

that calculating the effect size statistic. According to Pallant (2010:210) effect size is a set of statistics indicating the magnitude of the differences between means in the dependent variable that is predictable from knowledge of the levels of the independent variable. and the guidelines from Cohen as follows:

Eta Square (η^2)

$$\eta^2 = \frac{t^2}{t^2 + (n_1 + n_2 - 2)}$$

Where:
T : value of t test
N : number of students of first group
N2 : number of students of second group

Then, according to Cohen (1988 in pallant, 2005:209). The guidlines for interpreting the value of eta sequare are as follow:

Table III. 18
Effect Size Guidelines

0.01= Small Effect
0.06= Moderate Effect
0.14 = Large Effect