CHAPTER IV

THE FINDING AND DISCUSSION

This chapter provides short review of the finding consisting of data description, data analysis, and discussion relate to the result of research that was done in SMP Islam Raden Paku Surabaya.

A. Finding

After assigning post-test toward experimental and control class, then researcher calculated by using IMB SPSS 2.5 version. It was purposely done to find the difference of mean score of pre and post-test on experimental and control class.

1. The Data Description

The researcher conducted the research by using quasi-experimental research design. It was done from 15th may to 6th June 2018 in SMP Islam Raden Paku Surabaya on Jl. Klampis Ngasem No. 34. The population of this research was seventh grade that consisted of three classes namely, class of VII A, class of VII B, and class of VII B. In doing the research, the researcher chosen the sample of research that was taken from class of VII A and class of VII class B which each class consisted of 20 students and they were divided into two classes, namely class of VII A as experimental class and class of VII B as control class. The experimental class was class that was taught speaking skills on describing the characteristics of animals by using hot seat game method while control class used usual method.

In this research, there were several steps that were done by researcher in doing research. The first step, researcher held try out to another class. The second, the researcher conducted pre-test to both classes, namely experimental and control class after assigning the pre-test to both classes, the researcher and teacher scored the pre-test of experimental and control class based on the rubric assessment that was adopted from *Buku Guru Bahasa Inggris* 2017. The third, the researcher gave the treatment to experimental class. The fourth, the researcher conducted post-test to both classes. The

fifth, the researcher analyzed the data of try out and pre and post-test of experimental and control class.

After collecting the scores of pre and post-test from experimental and control class, then the researcher tabulated and analyzed all scores of both classes by using IBM SPSS 2.5 version. The data of pre and post-test were described in the tables as follows:

Note: The technique of taking score in these tables were the higher score of both two raters' scores are chosen.

Table 4.1 Students' Pre-Test Score of Experimental Class

Students' Code	Experimental Class		Score
	Rater-1	Rater-2	
1	65	70	70
2	60	65	65
3	55	60	60
4	70	65	70
5	60	65	65
6	70	70	70
7	50	60	60
8	60	55	60
9	60	60	60
10	60	65	65
11	55	60	60
12	55	60	60
13	60	55	60
14	65	65	65
15	70	70	70
16	70	75	75
17	60	75	75
18	70	60	70
19	70	70	70
20	75	70	75
Total Score			1325
Higher			75
Lower			60
Average			66.25

Table 4.1 above is the students' pre-test score of experimental class (VII A) that was scored by two raters, namely rater-1 as teacher and rater-2 as researcher. It describes that the total score is 1325, the higher score is 75, the lower score is 60, and the average score is 66.25.

Table 4.2 Students' Pre-Test Score of Control Class

Students' Score	Control class		Score
	Rater-1	Rater-2	
1	70	60	70
2	65	60	65
3	55	60	60
4	70	70	70
5	65	60	65
6	65	70	70
7	60	60	60
8	60	55	60
9	60	60	60
10	60	65	65
11	60	55	60
12	60	55	60
13	60	60	60
14	65	60	65
15	70	70	70
16	75	70	75
17	70	75	75
18	65	70	70
19	70	70	70
20	75	70	75
Total Score			1325
Higher			70
Lower			55
Average			66.25

Table 4.2 describes the students' pre-test score of control class (class VII B). Based on data of score that was scored by two raters, on the pre-test, the control class obtained the total score is 1325, the higher score is 70, the lower score is 55, and the average is 66.25.

Inter-rater reliability was also presented by researcher in scoring students' performance during oral speaking test. It meant consisting of two raters to score namely, the first rater was English teacher of SMP Islam Raden Paku Surabaya and the second rater was researcher. The reliability of post-test was calculated on the correlation by using IBM SPSS 2.5 version. The score results of post-test as follow.

Table 4.3 Students' Post-Test Score of Experiment Class

Student's Code	Experiment Class		Score
	Rater-1	Rater-2	
1	80	80	80
2	80	80	80
3	85	85	85

4	90	85	90
5	85	90	90
6	85	85	85
7	85	80	85
8	80	80	80
9	90	90	90
10	95	95	95
11	90	90	90
12	80	80	80
13	80	80	80
14	90	95	95
15	90	90	90
16	95	95	95
17	95	90	95
18	90	90	90
19	90	90	90
20	90	85	90
Total Score			1755
Higher			95
Lower			80
Average			87.75

Table 4.3 describes the result of post-test score of experimental class that was scored after researcher gave treatment by using hot seat game, then both two raters scored after conducting oral speaking test as instrument of test. It shows that total score was gained by experimental class in the post-test is 1755, the higher score is 95, the lower score is 80, and the average is 87.75

Table 4.4 Students' Post-Test Score of Control Class

Student's Code	Control Class		Score
	Rater-1	Rater-2	
1	60	70	70
2	70	70	70
3	70	75	75
4	75	80	80
5	80	80	80
6	75	80	80
7	75	75	75
8	80	80	80
9	80	80	80
10	70	85	85
11	75	85	85
12	80	80	80
13	80	80	80
14	80	80	80
15	75	75	75
16	80	80	80
17	90	90	90

18	80	80	80
19	80	80	80
20	80	80	80
Total Score			1585
Higher			95
Lower			75
Average			79.25

Table 4.4 elaborates the result of score in post-test. In this table shows that total score was obtained by control class is 1585, the higher score is 95, the lower score is 75, and the average is 79.25. To understand more detail relates to tables above and the finding of this research, researcher explained in the analysis of data that calculated the data into IBM SPSS Statistics 25 version.

2. The Result of Data Analysis

To facilitate in deciding the analysis result of data that were gained from pre and post-test of experimental and class control and answering the research questions of this research, the researcher collected and analyzed the data by using IBM SPSS Statistics 2.5 version. In this sub-chapter, the observer analyzed pre and post-test scores by using the result of reliability of try out, the result of reliability of pre and post-test, the test of normal distribution, the test of homogeneity, and t-test calculation that counted the scores based on tables of the data description above.

a) The Result of Reliability of Try Out

The researcher served the try out to another class before assigning the experimental and control class with pre-test. It was aimed to test and to know the consistency of instrument that was given to both classes. Brown (2003:20) points out that reliability is the consistence and dependable of a test that is assigned to same student on two different occasion obtains similar results. In measuring the degree of consistency and dependable of test, the researcher used inter-rater where teacher as the first rater (rater-1) and researcher as second rater (rater-2). In the same way, Ravid (2011:197) says that inter-rater reliability consists of two or more raters who score the same performance or behavior and have the same grade of consistency and

agreement among scores being assigned. To find and decide the reliability of try out scores, the researcher used to IBM SPSS 2.5 version that analyzed into correlation.

Table 4.5 Result of Reliability of Try Out

		Rater-1	Rater-2
Rater-1	Pearson correlation	1	.663**
	Sig. (2-tailed)		.001
	n	20	20
Rater-2	Pearson correlation	.663**	1
	Sig. (2-tailed)	.001	
	n	20	20

**Reliability is significant at level 0.01 level (2-tiled)

Based on table 4.5, it shows that the instrument test of try out in another class is reliable. It can be seen from two raters' scores on the pearson correlation which gained .663**. Refers to the variance level of reliability which is interpreted by Ravid, it proves that the reliability of try out instrument is very high. As Ravid (2011:120) interprets the level of reliability as follows.

Table 4.6 Reliability Interpretation

Reliability	Interpretation
.00 - .20	No Reliability
.20 - .40	Low
.40 - .60	Moderate
.60 - .80	High
.80 - .1.00	Very High

b) The Result of Reliability of Pre and Post-Test

To calculate the reliability of pre and post-test, the researcher also counted with correlation formula. Ravid (2011:192) states that reliability is the level of an instrument consistency that can obtain same results when the instrument is used in other times. In counting the reliability of pre and post-test, researcher used IBM SPSS 25 version.

(1) The Result of Reliability Test of Pre-Test in Experimental and Control Class

Table 4.7 Result of Reliability of Pre-Test of Experimental Class

		Rater-1	Rater-2
Rater-1	Pearson Correlation	1	.575**
	Sig. (2-tailed)		.008
	n	20	20
Rater-2	Pearson Correlation	.575**	1

	Sig. (2-tailed)	.008	
	n	20	20

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

Table 4.7 indicates that the score of pre-test that experimental class obtained is reliable. The rater-1 and rater-2 in pearson correlation gained .575**. Relates to the significance of reliability is at the 0.01 level (2-tailed). The result of both two raters obtained .575> 0.01. Refers to the variance level of reliability which is interpreted by Ravid, it indicates that level of reliability of data is very high. It can be concluded that the reliability of pre-test in experimental class is reliable.

Table 4.8 Result Reliability of Pre-Test of Control Class

		Rater-1	Rater-2
Rater-1	Pearson correlation	1	.814**
	Sig. (2-tailed)		.000
	n	20	20
Rater-2	Pearson correlation	.814**	1
	Sig. (2-tailed)	.000	
	n	20	20

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

Based on table 4.8, it shows that between the rater-1 and the rater-2 obtained the same score that appears the Pearson correlation is .814**. Refers to the significance of reliability is at the level 0.01 (2-tailed), it means that .814**> 0.01. Ravid (2011:120) categorizes that .80 to 1.00 is very high level of reliability. It proves that the level of reliability data is very high. It can be categorized that the result of pre-test on control class is reliable.

(2) The Result of Reliability of Post-Test in Experimental and Control Class

Table 4.9 Result of Reliability of Post-Test of Experimental Class

		Rater-1	Rater-2
Rater-1	Pearson correlation	1	.867**
	Sig. (2-tailed)		.000
	n	20	20
Rater-2	Pearson correlation	.867**	1
	Sig. (2-tailed)	.000	
	n	20	20

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

Table 4.9 shows that between pearson correlation of rater-1 and pearson of rater-2 reached the same score. Pearson correlation of both raters are 0.867**. It

indicates that level of reliability of data is very high. From the review above, it can be concluded that the result of reliability test is reliable.

Table 4.10 Result of Reliability of Post-test of Control Class

		Rater-1	Rater-2
Rater-1	Pearson correlation	1	.652**
	Sig. (2-tailed)		.002
	n	20	20
Rater-2	Pearson correlation	.652**	1
	Sig. (2-tailed)	.002	
	n	20	20

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

Based on table 4.10, it indicates that the result of pearson correlation between rater-1 and rater-2 have the same score. The score of both raters that appears in pearson correlation is .652**. So, it shows that the score of post-test in control class is reliable.

c) The Result of Test of Normal Distribution

In this case, to ensure the English teacher's statement of SMP Islam Raden Paku Surabaya and to know that students of experimental class and control class have normal distribution, the researcher did test of normal distribution. It was done after pre and post-test were assigned by researcher for experimental class and control class. The test of normal distribution was referred to the pre-test score that was gained by researcher from experimental and control class. To analyze the data, the researcher used IDM SPSS software 2.5 version and to test the normality of distribution, the researcher used one sample Kolmogorov Smirnov.

Table 4.11 Result of Test of Normal Distribution of Experimental and Control Class in Pre-Test

One-Sample Kolmogorov-Smirnov Test			
n		Experimental Class	Control Class
		20	20
Normal Parameters	Mean	63.00	64.75
	Std. Deviation	6.767	5.955
Most Extreme Differences	Absolute	.221	.187
	Positive	.221	.187
	Negative	-200	187
Test Statistics		.221	.187
Asymp Sig. (2-tailed)		.011[c]	.64[c]

a. Test distribution is Normal

b. Calculated from data

Table 4.11 describes the result test of normal distribution of experimental and control class in pre-test. From the description of pre-test on the table above shows that significant of experimental class .011> α (0.05) and the significant of control class is .64> α (0.05) which significant values are higher from α (0.05). It indicates that H_A is accepted and it can be decided that the test distributions of experimental and control class are normal.

After testing the normality of pre-test in experimental and control class, the researcher continued to test the normality of post-test. The data of normality result of both class in post-test in table 4.12 as follows.

Table 4.12 Result of Test of Normal Distribution of Experimental and Control Class in Post-Test

One-Sample Kolmogorov-Smirnov Test			
n		Experimental Class	Control Class
		20	20
Normal parameter	Mean	87.25	86.75
	Std. Deviation	5.250	5.447
Most Extreme Differences	Absolute	.250	.225
	Positive	.166	.192
	Negative	-250	-225
Test Statistics		.250	.225
Asymp Sig. (2-tailed)		.0031[c]	.009[c]

a. Test distribution is normal

b. Calculated from data

Table 4.12 reports that significant result of experimental class is .0031> α (0.05) and the significant result of control class is .009> α (0.05). It can be proved that the significant score of both classes are higher than α (0.05) which indicates that H_A is accepted and H_0 is rejected. Therefore, it can be concluded that the test distributions of both classes are normal.

d) The Result of Test of Homogeneity of Variance

The main purpose of applying the test of homogeneity of variance was to support and prove English teacher's statement of SMP Islam Raden Paku Surabaya that students of experimental class and control class have equal ability. And it also

was purposely tested to know the homogeneity whether the sample of research had the same characteristics and skills in speaking. The test of homogeneity of variance was done after researcher examined the normal distribution. The researcher examined the homogeneity test by using IDM SPSS software 2.5 version. The data of homogeneity test was obtained from experimental class and control class in pre-test scores. Table of homogeneity test as follows.

Table 4.13 Result Test of Homogeneity of Variances of Experimental Class and Control Class in Pre-Test

Test of Homogeneity of Variances			
Levene Statistic	df 1	df 2	Sig.
.159	1	38	.692

The criteria of testing homogeneity were hypothesized in chapter three as follow:

a) The significance value (2-tailed) >0.05, it means that the result of test is homogenous and the H_A is accepted.
b) The significance value (2-tailed) < 0.05, it means that the result of test is not homogenous and the H_0 is rejected.

Based on table 4.13 shows that significant value of homogeneity test is .692> α (0.05) it can be said that H_A is accepted and also shows that between experimental and control class have the same characteristics and abilities. But, if result of significant of homogeneity test is lower than α (0.05) or < α (0.05) it informs that both classes are not homogeneous and the sample cannot be used as sample of research. The result of homogeneity test above proves that the level of significant is higher than α (0.05), so it can be used as sample of research

The researcher also described the scores that were gained from pre and post-test that were presented by using descriptive statistic. Ravid (2011:29) defines that descriptive statistics where a particular group of observation is classified, organized, and summarized in numerical data. In addition, the descriptive statistics in the research particularly reports the result of pre and post-test that were organized in the table which facilitates the researcher to decide the significant difference of results between pre and post-test of experimental and control class.

Table 4.14 Descriptive Statistics of Pre and Post-Test Score in Experimental Class

n		Minimum	Maximum	Sum	Mean	Std. Deviation
Score of pre-test	20	60	75	1325	66.25	5.590
Score of post-test	20	80	95	1755	87.75	5.495
Valid n (listwise)	20					

Table 4.14 indicates that the descriptive statistics of pre and post-test of experimental class which consists of 20 participants obtained the minimum score of pre-test is 60, the maximum score of pre-test is 75, the sum of pre-test score is 1325, the mean of pre-test score is 66.25, and the std. deviation of pre-test is 5.590 while the descriptive statistics of post-test in experimental class reached the minimum score of post-test is 80, the maximum score of post-test is 95, the sum score of post-test is 1755, the mean score of post-test is 87.75, and the std. deviation of post-test is 5.495.

Table 4.15 Descriptive Statistics of Pre and Post-Test Score in Control Class

	n	Minimum	Maximum	Sum	Mean	Std. Deviation
Score of pre-test	20	55	80	1385	66.25	5.911
Score of post-test	20	70	90	1585	79.25	4.667
Valid n (listwise)	20					

Table 4.15 shows the descriptive statistics of pre and post-test in control class which consists of 20 participants reached the minimum score of pre-test is 55, the maximum score of pre-test is 80, the sum score of pre-test is 1385, the mean score of pre-test is 66.25, and the std. deviation of pre-test is 5.911 while in the post-test that also consists of 20 participants obtained the minimum score of post-test is 70, the maximum score of post-test is 90, the sum score of post-test is 1585, the mean score of post-test is 79.25, and the std. deviation of post-test is 4.667

e) T-Test Calculation

After the normality and homogeneity test of experimental class and control class were calculated, researcher calculated the result of pre and post-test in experimental class and control class using IBM SPSS 2.5 version. It meant to find the mean scores of both classes and know whether the hot seat game method was effective for teaching speaking skill. As Ravid (2011:144) points out that t-test is a statistic that is used to compare the mean of two group. In addition, t-test was presented to find and compare the mean score between experimental class and control class. In t-test calculation, the researcher calculated the result of two test scores namely, pre and post-test.

(1) The Result of T-Test Calculation in Pre-Test

To know the mean score between experimental class and control class on pre-test, the researcher analyzed the outcome of pre-test scores that were gained by researcher from both classes. In analyzing the data of pre-test, the researcher used independent sample test.

(a) Independent Sample T-Test

The researcher used independent sample test which was purposely to find the means score of experimental and control class. Ravid (2011:146) mention that independent sample t-test is a test which is used in experimental design to know and to compare the mean result of two group. In the same way, Machali (2015:66) defines that independent sample t-test is a test which is used to find the deference of mean scores of two groups or classes. As the basic of interpretation decision, the researcher made and decided as follow:

(a) H_A = The experimental and control class have the same mean scores.
(b) H_0 = The experimental and control class have different mean scores.
(c) If the result of sig. (2-tailed) is < 0.05 it shows that experimental and control class have significant difference. So, the H_A is accepted and the H_0 is rejected.
(d) If the result of sig. (2-tailed) is > 0.05 it shows that both of classes has not the significant difference. So, the H_A is refused and H_0 is accepted

Table 4.16 Descriptive Statistics of Mean Score in Pre-Test

Group Statistics					
Score of pre-test	Class	n	Mean	Std. Deviation	Std. Error Mean
	A	20	66.25	5.550	1.250
	B	20	66.25	5.590	1.250

From table 4.16, it can be known that the score of A class in pre-test is 66.25, standard deviation is 5.550, and standard error mean is 1.250 while B class gains mean score is 66.25, standard deviation is 5.590, and standard error mean is 1.250. From the mean score of both classes which shows the same mean, it indicates that both classes have the same ability.

Table 4.17 Result of Independent Sample T-Test of Pre-Test

Independent Sample T-Test										
		Levene's t-test for equality of variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval Difference	
									Lower	upper
Score of pre-test	Equal variances assumed	.000	.1.000	.000	38	.1.000	-000	1.768	-3.579	3.579
	Equal variances not assumed			.000	38.000	.1000	-000	1.768	-3.579	3.579

Table 4.17 reports that the result of counting the significance of levene's t-test for equality of variances is .1.000> α (0.05) meaning that the significant level is higher than 0.05. In finding the result of t-test for equality of means, the researcher chose table sig. (2-tailed) in the first line which refers to equal variances assumed as basic of decision in determining whether experimental and control class have the same significant or not in the pre-test. The sig. (2 tailed) of t-test for equality means is .1.000> α (0.05) which indicates level of significant is higher than 0.05.

Based on the criteria of interpretation that researcher made, if the result of sig. (2-tailed) is lower than 0.05 it shows that both classes have different significance, the H_0 is accepted and the H_A is rejected. But, if the result of sig. (2-tailed) is higher than 0.05 it indicates that both classes has not different significance, the H_A is accepted and the H_0 is rejected. The result of independent sample test in t-test for equality of means reports that the sig. (2-tailed) which refers to Equal variances assumed is .1.000. It means that the sig (2-tailed) is higher than 0.05. From the result of analysis in pre-test above proves that both classes have the same ability. So, both of classes has not different significance.

(2) The Result of T-Test Calculation in Post-Test

To obtain the mean score of post-test in experimental class and control class, the researcher analyzed the result of post-test scores that were gained by researcher from both classes after participants got post-test. In analyzing the data of post-test, the researcher also used independent sample test and paired sample test. The main aim of finding the mean score of both classes was to know whether experimental and control class have equal or unequal mean scores and to decide whether the hot seat game method was effective in teaching speaking skill.

(a) Independent Sample T-Test

To find the mean score of experimental and control class in post-test, the researcher also used independent sample test which was counted by IBM SPSS 2.5 version. The criteria and hypotheses that researcher made and decided as the basic of interpretation decision was the same as criteria and hypotheses in pre-test as follows:

(a) H_A = The experimental and control class have the same mean scores.
(b) H_0 = The experimental and control class have different mean scores.
(c) If the result of sig. (2-tailed) is < 0.05 it shows that the experimental and control class have significant difference. So, the H_A is accepted and the H_0 is rejected.
(d) If the result of sig. (2-tailed) is > 0.05 it shows that both of classes has not the significant difference. So, the H_A is refused and H_0 is accepted

Table 4.18 Descriptive Statistics of Mean Score in Post-Test

Group Statistics					
Score of post-test	Class	n	Mean	Std. Deviation	Std. Error Mean
	A	20	87.75	5.495	1.229
	B	20	79.25	4.667	1.043

Based on table 4.18 above, the mean score of A class and B class in the post-test have different mean. In the group statistics, the mean scores of A class and B class are different which is the mean score of A class is 87.75, standard deviation is 5.495, and standard error mean is 1.229 while the mean score of B class is 79.25, standard deviation is 4.667, and standard error mean is 1.043. It proves that the mean score of A class is higher than the mean score of B class.

Table 4.19 Result of Independent Sample T-Test of Post-Test

Independent Sample T-Test										
		Levene's test for equality of variances					t-test for equality of means		95% confidence interval of the difference	
		F	Sig.	T	Df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Score of post-test	Equal variances assumed	2.690	.109	5.273	38	.000	8.500	1.612	5.237	11.763
	Equal variances not assumed			5.273	37.028	.000	8.500	1.612	5.234	11.766

Based on table 4.19 the result of analyzing the significance of levene's t-test for equality of variances is .109> α (0.05). It means that the significant level is higher than 0.05. In finding the result of t-test for equality of means, the researcher chose table sig. (2-tailed) in the first line which refers to equal variances assumed as basic of decision in determining whether experimental and control class have the same significant or not in the pre-test. The sig. (2 tailed) of t-test for equality means is .000> α (0.05) which interprets the level of significant is smaller than 0.05.

From the criteria of interpretation that researcher made if the result of sig. (2-tailed) is lower than 0.05 it shows that both classes have different significant, the H_0 is accepted and the H_A is rejected. But, if the result of sig. (2-tailed) is higher than

0.05 it indicates that both classes have not different significant, the H_A is accepted and the H_0 is rejected. The result of independent sample t-test in t-test for equality of means reports that the sig. (2-tailed) which refers to Equal variances assumed is .000. It means that the sig (2-tailed) is smaller than 0.05. So, from the result of analysis in pre-test proves that experimental class and control class have different significance.

(b) Paired Sample T-Test

Researcher also used the paired sample t-test. It was used to analyze whether the mean score result of pre and post-test are different or not. Ravid (2011:151) argues that paired sample t-test is a test being used to compare the two mean scores from two sets of one group. In addition, to know whether independent variable influence dependent variable, it can be measured from the result of pre and post-test mean scores. If the mean score result of post-test is higher than mean score pre-test, it shows that independent variable effected dependent variable.

To facilitate the researcher in deciding and interpreting the result of calculating tests that were counted by IBM SPSS 2.5 version, the researcher used the criteria and hypotheses that elaborated as follows:

(a) H_A = The hot seat game method is effective in teaching speaking skills.
(b) H_0 = The hot seat game method is not effective in teaching speaking skills.
(c) If the result of sig. (2-tailed) is < 0.05 it can be stated that the hot seat game method is effective in teaching speaking skills. So, the H_A is accepted and the H_0 is rejected.
(d) If the result of sig. (2-tailed) is > 0.05 it shows that the hot seat game method is not effective in teaching speaking skills. So, the H_A is refused and H_0 is accepted

The results of pre and post-test mean score of experimental class will be listed as follows.

Table 4.20 Paired Samples of Statistics in Experimental Class

	Mean		n	Std. Deviation	Std. Error Mean
Pair 1	Pre-test	66.25	20	5.590	1.250
	Post-test	89.00	20	5.282	1.181

The mean score of pre and post-test in experimental class are not same and each of tests has 20 participants. The mean score of pre-test is 66.25, standard deviation is 5.590, and standard error mean is 1.250 while the mean score of post-test is 89.00, standard deviation is 5.282, and standard error mean is 1.181. From the outcome of analyzing the mean score of pre and post-test above, it can be said that experimental class has significant development after gaining the treatment. And the detailer description of mean score differences of pre and post-test can be seen in table 4.19 as follows.

Table 4.21 Result of Paired Samples T-Test of Experimental Class

Paired Sample T-Test									
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of Difference		T	Df	Sig. (2-taile)
					Lower	upper			
Pair 1	Pre-test – post-test	22.750	4.723	1.056	24.960	20.540	21.544	19	.000

The result of calculation test on paired samples t-test shows that t count is 21.544, degree of freedom is 19, the mean is 22.750, and significance (two-tailed) is .000. Referring to the decisions of criteria that were chosen as written in chapter III, if the significance value (two-tailed) < 0.05, it means that H_A is accepted and H_0 is rejected. But, if the level of significance (two-tailed) > 0.05, it indicates that H_A is refused and H_0 is accepted. The significance value (two-tailed) on the paired samples t-test is .000. It proves that .000 is smaller than 0.05. So, the H_A is accepted and H_0 is rejected.

B. DISCUSSION

The main point of this explanation relates to the results of the result of reliability of try out, the result of reliability of pre and post-test, the test of normal distribution, the test of homogeneity, and t-test calculation t-test. In this matter, the researcher elaborated the data of research result that were analyzed by using IBM SPSS 2.5 version.

The first point relates to the result of reliability of try out. The try out was presented to another class before observer assigned pre-test to the experimental and control class, it meant to know and measure the accuracy and consistency of two raters that consisted of an English teacher as the first rater (rater-1) and researcher as second rater (rater-2). Based on the analysis result of reliability of try out on table 4.5 above shows that result of reliability test is significant which both raters obtained .663**> 0.01. It was referred to Ravid's interpretation that 80 to 1.00 is very high reliability. So, the result of reliability test was very high and instrument of test could be used in pre and post-test.

The second point explains the result of reliability of experimental class in pre-test. The pre-test was assigned to the experimental class students before researcher gave the treatment. It was purposely provided to measure whether experimental class had increase after obtaining the treatment. Based on result of reliability test of pre-test on table 4.7 above, it indicates that both raters attained in high significance reliability which was described in pearson correlation .575**. Refers to the significance level of reliability at 0.01 (2-tailed), it can be said that 575> 0.01. It proves that the result of reliability test is high.

The third point elaborates the result of reliability test of control class in pre-test. The pre-test also was provided for control class, although it was not treated as experimental class gained. It was also done to measure the mean score of control class that did not gain the same treatment with experimental class. From table 4.8 reports that both raters namely, rater-1 and rater-2 reached .814**> 0.01. Ravid (2011:120) categorizes that .80 to 1.00 is very high reliability. It indicates that reliability is very high.

The fourth point explains the result of reliability test of post-test in experimental class. The reliability test was conducted by researcher to know and measure the accuracy and consistency of an instrument. After giving treatment to experimental class that was applied in hot seat game method, the researcher assigned post-test. The result of reliability test can be seen on the table 4.9. It shows that both raters obtained .867** as appears in pearson correlation that higher than 0.01. From this description, it can be decided that reliability of instrument in post-test is reliable.

The fifth point relates to the result of reliability test of post-test in control class. Post-test was also conducted in control class. It was purposely done to know and measure

the mean scores between experimental and control class. After assigning post-test to control class, researcher examined the reliability of scores which had been gained form control class. It was done to decide whether the scores of both raters are reliable. Refers to the result of reliability test on table 4.10 appeared that both raters had reached .652** as appears on pearson correlation. The significance of reliability is at 0.01 level (2-tailed). It proves that scores of both raters higher than 0.01and it is reliable.

The sixth point relates to the result of normal distribution test of experimental and control class in pre-test. The normality test was done to ensure whether the data distribution was normal or not. After pre-test was conducted in both classes, then the researcher analyzed the score of pre-test using IBM SPSS 2.5. But, before processing the data, researcher formulated the normality test criteria namely, H_A = the sample of data distribution is normal, H_0 = the sample of data distribution is not normal, the significance value (2-tailed) >0.05 indicates that the sample of data is normal distribution and the H_A is accepted, and the significance value (2-tailed) < 0.05 indicates that the sample of data is not normal distribution and the H_0 is refused. Based on table 4.11 that was analysis result by using IBM SPSS 2.5 version shows that level of significance of experimental class is .011> α (0.05), and the significant of control class is .064> α (0.05) which significant values are higher from α (0.05). It proves that H_A is accepted and it can be decided that the test distributions of experimental and control class are normal.

The seventh point explains about the result of normal distribution test of experimental and control class in post-test. To know whether the data of post-tests are normal or not, the researcher analyzed the data that was gotten from both classes. Before analyzing all data, the researcher also used the same criteria with pre-test. These are H_A = the sample of data distribution is normal, H_0 = the sample of data distribution is not normal, the significance value (2-tailed) >0.05 indicates that the sample of data is normal distribution and the H_A is accepted, and the significance value (2-tailed) < 0.05 indicates that the sample of data is not normal distribution and the H_0 is refused. Relates to the analysis result of normality distribution test was analyzed by using IBM SPSS version 2.5 which was illustrated in table 4.12 showed that significant result of experimental class is .0031> α (0.05) and the significant result of control class is .009> α (0.05). It meant that level of significance of both classes are higher than α (0.05). So, from this illustration proves that the data of post-tests are normal.

The eighth point elaborates the result of homogeneity test that was purposely conducted to test whether the students of both classes had the same abilities. It was important for researcher to ensure that students of both classes as sample of research had equal skills. Before analyzing the score results of pre-test of both classes, the researcher determined several criteria of testing homogeneity as categorized H_A = The sample of research is homogenous while H_0 = The sample of research is not homogenous, the significance value (2-tailed) > 0.05, it means that the result of test is homogenous and the H_A is accepted, and the significance value (2-tailed) <0.05 it means that the result of test is not homogenous and the H_0 is accepted. Based on the result of homogeneity test which was displayed on table 4.13 informed that the level of significance that was reached by both classes are .692> 0.05. Refers to the criteria of testing homogeneity that was made by researcher, if significance value (2-tailed) > 0.05, it means that the result of test is homogenous and the H_A is accepted. From the previous explanation, it proves that both class homogeneity.

The last point of this discussion relates to the result test of t-test calculation. In this matter, the researcher elaborates the result of independent samples t-test and paired samples t-test that were used to analyze the scores of both classes. The tests were conducted by researcher in different cases involving pre and post-test.

The independent samples test is one of test that involved two classes namely, experimental and control class. It was conducted to analyze the mean score of both classes. Before calculating the scores of both classes, researcher had estimated several criteria of assessment which were the basic of decision in interpreting the output result of IBM SPSS 2.5 version analysis as noted H_A = the experiment and control class have the same mean scores, H_0 = the experiment and control class have different mean scores, if the result of sig. (2-tailed) is < 0.05 it can be seen that the experimental and control class have significant difference. So, the H_A is accepted and the H_0 is rejected, and if the result of sig. (2-tailed) is > 0.05 it shows that both classes have not the significant difference. So, the H_A is refused and H_0 is accepted.

In this case, the independent samples t-test was used to count the mean scores of pre-tests of experimental class being coded with A and control class being coded with B. The first discussion relates to the result of independent samples t-test in pre-test. After assigning the pre-test toward A class and B class, the researcher counted the scores were

obtained from both classes used IBM SPSS 2.5 version. Based on analysis result on table 4.16 shows that A class reached the mean score 66.25 and also A class reached the same mean scores. It indicated that the mean score of both classes are equal.

Refers to the analysis result of independent samples t-test in table 4.17 indicates that the significance of levene's t-test for equality of variances is .1.000> α (0.05) showing that the significance level is higher than 0.05 which indicates that both classes have equal variances. But, to know whether A class and B class have significant difference of mean scores, the researcher chose table sig. (2-tailed) in the first line connecting to the equal variances assumed as basic of decision in determining whether both classes have the different significance or not in the pre-test. T-test for equality means shows that the sig. (2-tailed) .1.000> α (0.05) which indicates level of significance is higher than 0.05. Based on the decision criteria that were made if the result of sig. (2-tailed) is < 0.05 it can be seen that both classes have different significance. So, the H_A is accepted and the H_0 is rejected. But, if the result of sig. (2-tailed) is > 0.05 it shows that both classes have not the significant difference. So, the H_A is refused and H_0 is accepted. From explanation indicates that sig. (2-tailed) is higher than 0.05. It proves that have not significant different of both classes' mean scores. So, the H_A is accepted and H_0 is refused.

After applying the treatment to the A class, the researcher assigned post-test toward A and B class. The post-test was conducted to measure and know the effectiveness of teaching speaking skills through hot seat game that was treated for A class. The scores of post-tests of both classes were counted with IBM SPSS 2.5 version using independent samples t-test. To facilitate the process of analyzing the mean scores of both classes on IMB SPSS 2.5 version output, researcher used the criteria of decisions as used in independent samples t-test of pre-test.

Based on analysis outcome that was appeared in table 4.18 shows that each class consisted of 20 participants which mean score of A class is 87.75 while B class is 79.25, and standard deviation of A class is 5.495 and B class is 4.667. To know and to ensure that both classes had difference of significant mean scores, it can be analyzed on table 4.19 the independent samples t-test of post-test.

Before analyzing the independent samples t-test of post-test output, the researcher decided to use the criteria that provided for analyzing the independent samples t-test of

pre-test outcome. Refers to table 4.19 independent samples t-test of post-test output shows that the level of significance on levene's t-test for equality of variances is 0.109> α (0.05). It indicates that both classes have the same variance.

To determine that both classes have difference of significant mean score, it can be seen from the value of sig. (two-tailed) that refers to equal variances assumed on the t-test for equality of means. Before analyzing the output of t-test, researcher refers to the basic of decision interpretation namely, if the result of sig. (two-tailed) is < 0.05 it indicates that both classes have difference of significant mean score. So, the H_A is accepted and the H_0 is rejected. But, if the result of sig. (two-tailed) is > 0.05 it shows that both classes have not the difference of significant mean score. So, the H_A is refused and H_0 is accepted. Based on the output of independent samples t-test appears that the value of sig. (two-tailed) is .000< 0.05. It means that p value on sig. (two-tailed) is lower than 0.05. From this explanation proves that A class and B class have difference of significant mean score. So, H_A is accepted and H_0 is rejected.

To measure and ensure whether the treatment that was applied in A class was effective or not, researcher assigned post-test to A class after giving treatment. In this matter, researcher answered the research questions that were formulated on the statement of study in chapter one. It is the main point of this discussion. The score of post-test was analyzed with paired sample t-test by using IBM SPSS 2.5 version.

Paired samples statistics was used to compare the mean score of pre and post-test on one group namely A class. After handling the pre and post-test score, then researcher calculated with IBM SPSS 2.5 version. Based on output of paired samples statistics on table 4.20 shows that mean value of pre-test is 66.25 and mean value of post-test is 89.00.

To know the difference of significant mean value on pre and post-test of A class, it can be analyzed on table 4.21 paired samples t-test output. Before analyzing the output of paired samples t-test, researcher determined the criteria of the basic of decision interpretation namely, if the result of sig. (two-tailed) is < 0.05 it indicates that pre and post-test have difference of significant mean score. So, the H_A is accepted and the H_0 is rejected. But, if the result of sig. (two-tailed) is > 0.05 it shows that pre and post-test have not the difference of significant mean score. So, the H_A is refused and H_0 is accepted. Based on the analysis output of paired samples t-test indicates that sig. (two-tailed) is .000< 0.05. It proves that p value is smaller than 0.05. So, H_A is accepted and H_0 is

refused. From this explanation, it can be concluded that hot seat game method is effective for teaching speaking.