CHAPTER III

RESEARCH METHODOLOGY

A. The Research Design

The type of this research was an experimental research. This research used quasi-experimental research design with nonequivalent control group. John Creswell states that quasi-experiment is experimental situation in which the researcher assigns participants to groups, but not randomly. [1] Furthermore, Gay and Peter Airasian state that quasi-experimental design is used when the researcher keeps the students in existing classroom intact and the entire classrooms are assigned to treatments.[2] It is the appropriate one to use in this research.

This design used two classes. One class was as a control class taught without using just-a-minute (JAM) game and the other class was as an experimental class taught by using just-a-minute (JAM) game. In both of classes, the researcher gave pre-test first. It was to know the basic ability of students. After that, the researcher gave treatment by using just-a-minute (JAM) game to the experimental class and without using just-a-minute (JAM) game to the control class. The last one, the researcher gave post test to experimental and control class to know the significant difference between using just-a-minute (JAM) game and

[1] John W Cresswell, *Educational Research (Third Edition) (*New York: Pearson Prentice-Hall, 2008), p. 313.

[2] L.R. Gay, and Peter Airasian, *Educational Research: Competencies for Analysis and Application (Sixth Edition)* (New Jersey: Pearson Prentice-Hall, 2000), p. 394.

without using just-a-minute (JAM) game. To give the clear one, the researcher delivers the illustration as follows:

TABLE III.1
RESEARCH DESIGN

1	**Experimental Class**	Sample	Pre-test	Treatments	Post-test
2	**Control Class**	Sample	Pre-test	No treatment	Post-test

B. The Time and Location of Research

The research was carried out at SMAN 1 Lubuk dalam Siak Regency in 2013/2014 academic year. The research was done on April 2014.

C. The Subject and Object of Research

The subject of this research was the second year students at SMAN 1 Lubuk Dalam Siak Regency in 2013-2014 academic year. The objects of this research was the significant difference between the students' speaking ability taught by using just-a-minute (JAM) game and without using just-a-minute (JAM) game at SMAN 1 Lubuk Dalam Siak Regency.

D. The Population and Sample

1. Population

The population of this research was the second year students of SMAN 1 Lubuk Dalam in 2013-2014 academic years. It had four classes. The number of population was 121 students. The number of the population can be seen in the following table:

TABLE III.2
THE TOTAL POPULATION OF THE SECOND YEAR STUDENTS OF SMAN 1 LUBUK DALAM SIAK REGENCY

No	Class / Major	Number of Students
1	XI IPA 1	23
2	XI IPA 2	23
3	XI IPS 1	36
4	XI IPS 2	39
	Total	**121**

2. Sample

The total number of the second year students at SMAN 1 Lubuk Dalam Siak Regency in 2013-2014 academic year was large enough to be taken all as sample of the research. In this research, the researcher only needed two classes. One class was as the experimental class and the other was as the control class. So the researcher limited the population by using purposive sampling as the way to choose the sample of population. According to Gay, purposive sampling selects a sample based on experience or knowledge of the group to be sampled.[3] Next, Arikunto states that purposive sampling selects the sample based on a reason or purpose.[4] The writer used purposive sampling, because the students' speaking ability at XI IPA 1 consisted of 23 students and class XI IPA 2 consisted of 23 students that was homogenous and the students were at the same level, both had

[3] Ibid, p. 138.

[4] Suharsimi Arikunto, *Penelitian: Suatu Pendekatan Praktik,* (Jakarta: Rineka Cipta, 2010), p 183.

the similar material and teacher. Furthermore, the English Teacher of SMAN 1 Lubuk Dalam also suggested the writer took the IPA class as the sample of research.

E. The Technique of Collecting Data

To find out the significant difference of speaking ability between the students who are taught by using just-a-minute (JAM) game and those who are not at SMAN 1 Lubuk Dalam Siak Regency. The writer used the oral presentation test to collect the data. The test consisted of a topic of speech that was taken from the students' text book and other sources.

The test consisted of pre test and post test. The pre test was carried out to know the background knowledge of the students' speaking ability in form of hortatory exposition for both experimental and control group. The researcher had given treatment to the students in the experimental class before the researcher gave post test based on the procedures of just-a-minute (JAM) game. The treatment was given to know the significant difference between the students' speaking ability taught and taught without using just-a-minute (JAM) game at SMAN 1 Lubuk Dalam Siak Regency.

Furthermore, there was no treatment in control class, so the researcher used the oral presentation test. At last, the researcher gave post-test to the experimental and control class. The speaking test consisted of five components. They are accent, grammar, vocabulary, fluency, and comprehension. In this

research, for assessing and scoring the students' speaking, there were some components that had to be considered. Hughes describes the rating as follows:[5]

a. Accent

TABLE III.3
ACCENT

Score	Requirement
1.	Pronunciation frequently unintelligible.
2.	Frequent gross error and a very heavy accent make understanding difficult, require frequently repetition.
3	"Foreign accent" requires concentrated listening, and mispronunciations lead to occasional misunderstanding and apparent errors in grammar of vocabulary.
4	Marked "Foreign accent" and occasional mispronunciations which do not interfere with understanding.
5	No conspicuous, mispronunciations, but would not be taken for a native speaker.
6	Native pronunciation, with no trace of "foreign accent"

b. Grammar

TABLE III.4
GRAMMAR

Score	Grammar
1	Grammar almost entirely inaccurate except in stock phrase.
2	Constant errors showing control of view major patterns and frequently preventing communication.
3	Frequent errors showing some major patterns uncontrolled and causing occasional irritation and misunderstanding.
4	Occasional errors showing imperfect control of some pattern but no weakness that causes misunderstanding.
5	Few errors, with no patterns of failure.
6	No more than two errors during the interview.

[5] Arthur Hughes, *Testing for Language Teacher* (Cambridge: Cambridge University, 2003),p.131.

c. vocabulary

TABLE III.5
VOCABULARY

Score	Requirement
1	Vocabulary inadequate for even the simple 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.
5	Professional vocabulary broad and precise; general vocabulary adequate to cope with complex practical problems and varied social situations.
6	Vocabulary apparently as accurate and extensive as that of an educated native speaker.

d. Fluency

TABLE III.6

FLUENCY

Score	Requirement
1	Speech is so 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.
5	Speech is effortless and smooth, but perceptively non-native an speed and evenness.
6	Speech on all professional and general topics as effortless and smooth as a native speaker's.

e. Comprehension

TABLE III.7

COMPREHENSION

Score	Requirement
1	Understands too title for the simplest types of conversation.
2	Understands only show, very simple speech on common social and touristic topics; requires constant repetition and rephrasing.
3	Understands careful, somewhat simplified speech when engaged in a dialogue, but may require considerable repetition and rephrasing.
4	Understands quite well normal educated speech when engaged in a dialogue, but requires occasional repetition or rephrasing.
5	Understands everything in normal educated conversation except for very colloquial or low-frequency items, or exceptionally rapid or slurred speech.
6	Understand everything in both formal and colloquial speech to be expected of an educated native speaker.

Note: for non-native speaker, number 5 is the highest score.

The speaking result was evaluated by concerning with five components and each component had score or level. Each component had 20 as the highest score and the total of all components was 100. The specification is as follows:

TABLE III. 8
THE SPECIFICATION OF THE TEST SCORE

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

F. Procedures of Research

In conducting the experimental research, the researcher did some research procedures for both experimental and control group. The research was carried out for eight meetings. The research procedures are as follows:

a. Conducting Pre-test

Pre-test was given by the researcher to both experimental and control group to know the basic ability of students' speaking. The researcher gave a text about hortatory exposition. The students should read it about five until ten minutes. After that, the researcher asked the students to give arguments about the topics that were already given to the students. They should present it about 3 minutes maximally. The researcher recorded them on videos.

b. Conducting Treatment

The treatment was given only to the experimental group. The treatment was given based on the just-a-minute (JAM) game procedures.

c. Conducting Post-test

The researcher gave the post test to both experimental and control group. In experimental class, the post-test was conducted to know the significant difference between the students' speaking ability taught by using just-a-minute (JAM) game. The post-test was using the procedures of just-a-minute (JAM) game. While, in control class, Post test was given to the students after teaching English for several times. It was to know the students' speaking ability taught without using just-a-minute (JAM) game. The procedure was the same as the pre-test, but the topic of text was different.

TABLE III.9
Topic of the Teaching Speaking in Each Meeting

No	Meeting	Topic
1	I	(Giving pre-test about several topic) general theme : Mobile Phone
2	II	Corruption
3	III	Drugs
4	IV	Smoking
5	V	Televison
6	VI	Tv"s ads
7	VII	Wearing helmet
8	VIII	(Giving post-test about several topic) general theme : internet

G. Validity and Reliability of the Test

1. Validity of the Test

The test used for testing the students' speaking ability has to have validity. According to Brown, a test is a method of measuring a person's ability, knowledge, or performance in a given domain.[6] Furthermore, Hughes states that a test is said to be valid if it measures accurately what it is intended to measure.[7] Based on the experts above, the researcher concludes that a test can be said valid if it is really measured person's ability appropriately.

Furthermore, Gay states that there are three kinds of validity. They are content validity, criterion-related validity, and construct validity.[8] In this research, the researcher used content validity. According to Bambang, if a measurement is as the representative of the ideas or the appropriate material that will be measured called content validity.[9] It means the test had fulfilled the validity of the content. In other words, the materials of the test have been taught at the second year of SMAN 1 Lubuk Dalam. They were familiar materials and near to the students' daily life. They were appropriate to the students' knowledge, insight and experience. Moreover, the materials were taken from the book guide for the students and other related resources. The test was based on the materials studied by the students at the moment. The pretest and posttest of this research were valid,

---

[6] H. Douglas Brown, *Language Assessment: Principles and Classroom Practices* (San Francisco:San Francisco State University, 2003), p. 3.

[7] Arthur Hughes, op. cit., p. 26.

[8] L.R. Gay and Peter Airasian, op. cit., pp. 163-167.

[9] Ag. Bambang Setiyadi, *Metode Penelitian Pengajaran Bahasa Asing; Pendekatan Kuantitatif dan Kualitatif.* Edisi Pertama. Yogyakarta: Graha Ilmu.2006. p.23.

because the students' speaking was measured by two raters by using the standard score of assessing speaking from Hughes. In making the test, the researcher had consulted first with the researcher's supervisor.

2. Reliability of the Test

Reliability is the measuring of test that is consistent and dependable.[10] It means that the test should consistently measure the person's ability. Furthermore, Gay states that reliability is the degree to which a test consistently measures whatever it is measuring.[11] It is reflected in obtaining how far the test or instrument test is able to measure the same subject on different occasions indicating the similar result. Furthermore, Brown states that there are two scoring processes in reliability. They are inter-rater reliability and intra-rater reliability. Inter-rater reliability occurs when two or more scores yield inconsistent scores of the same test. Intra-rater reliability is common occurrences for classroom teachers because of the unclear scoring criteria, bias toward particular 'good' and 'bad' students, or simple carelessness.[12] Based on the theories above, It is clear that reliability is used to measure the quality of the test scores and the consistency of the test. In this research, to know the reliability of the speaking test, the researcher used inter rater reliability, because the researcher had two raters in order to score the students' speaking ability. Gay said that inter judge reliability can be obtained by having two (more) judges independently score to be compared to the score of

[10] H. Douglas Brown, Ibid, p. 20.
[11] L.R. Gay and Peter Airasian, op. cit., p. 169.
[12] H. Douglas Brown, Ibid, p. 21.

both judges. Then the scores of the rater 1 correlated with the scores of the rater 2. The higher correlation, the higher the inter judge reliability.[13] The following table describes the correlation between score of rater 1 and the score of the rater 2 by using pearson product moment correlation formula through SPSS 17 Version:

TABLE III.10

Correlations

		rater1	rater2
rater1	Pearson Correlation	1	.573**
	Sig. (2-tailed)		.004
	N	23	23
rater2	Pearson Correlation	.573**	1
	Sig. (2-tailed)	.004	
	N	23	23

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

From the output above, it was found that r calculation was 0.573, the significance was 0.004 and the total number of students was 23. By orienting the Number of Significance, it was probably > 0.05, there was no correlation. If probably was < 0.05, there was a correlation. From the data, the number of significance was 0.004. it was lower than 0.05. it means that there was a correlation between rater1 and rater2. From r calculation of the data, it shows that 0.573 includes into moderate correlation from the Product moment correlation coefficient table.[14] From the output above, the speaking test was reliable.

[13] L.R. Gay and Peter Airasian, *loc cit*., p. 176.

[14] Hartono. *Statistik untuk penelitian (*Yogyakarta*: Pustaka pelajar,2010),* p 8.

H. The Technique of Analyzing the Data

1. Normality Test

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 lilliefors technique with SPSS 17 version.

TABLE III.11
NORMALITY OF TEST

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
experiment	23	100.0%	0	.0%	23	100.0%
Control	23	100.0%	0	.0%	23	100.0%

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	Df	Sig.	Statistic	Df	Sig.
Experiment	.124	23	.200*	.960	23	.466
Control	.172	23	.076	.900	23	.026

a. Lilliefors Significance Correction

*. This is a lower bound of the true significance.

For normality test, if the significance level > 0.05, the data distribution was normal, and if the significance level < 0.05 the data distribution was not normal, from the first output, it shows that the valid data was 23 and there was no missing data. From the second output, it shows the result of normality test, the significance level of experimental class was 0.200, while the significance level of

control class was 0.076. It can be stated that 0.200 > 0.05 and 0.076 > 0.05. It means that the data distribution was normal.

2. **Analysis data T-Test (Independent sample T-Test)**

In this research, the data were analyzed by using statistical method. First, in order to answer the formulation of the problem, the writer used students' scores of the experimental and the control group as the data of the research which were measured by two raters. In order to analyze the category of speaking ability, the writer used the category standard as follows[15]:

TABLE III.12
The Scale of the Students' Speaking ability

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

Next, the writer analyzed the data by using t-test[16] to know whether the result of the research is statistically significant the data were calculated by using SPSS 17.0. First, the writer found the difference between post-test score and pre test score in order to know the gain of each group.[17] Post-test score was subtracted pre-test score and it was equivalent with gain score. After that, the writer should find the t-score to analyze the data. These scores were analyzed statistically by using independent sample T-Test from SPSS 17.0 version.

[15] Suharsimi Arikunto,*Dasar-Dasar Evalusi Pendidikan* (Jakarta: Bumi Aksara, 2013),p. 251.
[16] Hartono. loc. cit., p. 178.
[17] Suharsimi Arikunto. loc.cit., pp. 350-352.