

## CHAPTER III

### RESEARCH METHOD

#### A. The Research Design

This research was experimental research. Experiment is testing an idea (or practice or procedure) to determine whether it influences an outcome or dependent variable.<sup>1</sup> In this research, the writer used quasi-experimental design. Quasi Experiment is used when the researcher needs to use intact group.<sup>2</sup> Kerlinger in Cohen stated that quasi-experiment is an apt description applied to much educational research where the random selection or random assignment of schools and classrooms is quite impracticable.<sup>3</sup> The research design form was the pre-test post-test non-equivalent group design. There were two variables used in this research. The first was the use of dialogue memorization technique which was symbolized as (X) and the second was students' speaking skill which was symbolized as (Y). In conducting this research, the writer involved the second grade students of MTs Darul Hikmah, the classes were divided into an experimental class and a control class. Both of the classes were given a pre-test to know the students' skill in speaking. After that, the experimental class was given the treatment by using dialogue memorization technique for 6 meetings while the control class was given conventional technique. At the end, both of the classes were tested again to find out the students' result in a post-test.

---

<sup>1</sup>John W. Creswell, *Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research*, (Boston: Pearson Education, 2008), 299

<sup>2</sup> Ibid, p. 313

<sup>3</sup> Louis Cohen, et al, *Research Method in Education*, Sixth ed, (London: Routledge, 2007), p. 282

According to Louis, the type of this research can be illustrated as follows:<sup>4</sup>

**Table III.1**  
**Research Design**

| Class        | Pre-test | Treatment | Post-test |
|--------------|----------|-----------|-----------|
| Experimental | O1       | X         | O2        |
| Control      | O3       | -         | O4        |

Where:

X : Treatment

O1: Pre-test of Experimental Class

O2: Post-test of Experimental Class

O3: Pre-test of Control Class

O4: Post-test of Control Class

Thus, the writer adapted the design above in which the subject of the research was the second grade students of MTs Darul Hikmah Pekanbaru. Then, the treatment was done for six meetings.

### **B. The Subject and the Object of the Research**

The subject of this research was the second grade students of MTs Darul Hikmah Pekanbaru, in the academic year of 2012/2013. The object of this research was dialogue memorization technique and speaking skill.

### **C. The Location and the Time of the Research**

This research was conducted at Madrasah Tsanawiyah Darul Hikmah Pekanbaru, located on Manyar Sakti Street. The research was done for 6 meetings, started from August 24<sup>th</sup> to September 12<sup>th</sup> 2013.

### **D. The Population and the Sample of the Research**

The population in this research was the second grade students at MTs Darul Hikmah Pekanbaru in 2012/2013 academic years, which consisted of 10

---

<sup>4</sup> Ibid p. 283

classes. The total of population was 292 students. The writer limited the population only 2 classes of the second grade after doing Cluster Sampling, VIII B1 as an experimental class and VIII B2 as a control class. Those were as the sample of the research by number 52 students; 26 students for experimental class and also 26 students for control class. In order to make it clearer, see the table below :

**Table III.2**

**The Population of the Second Grade Students  
at MTs Darul Hikmah Pekanbaru**

| <b>No</b>    | <b>Class</b> | <b>Male</b> |
|--------------|--------------|-------------|
| 1            | VIII B1      | 26          |
| 2            | VIII B2      | 26          |
| 3            | VIII B3      | 30          |
| 4            | VIII B4      | 30          |
| 5            | VIII B5      | 30          |
| <b>TOTAL</b> |              | <b>142</b>  |

| <b>No</b>    | <b>Class</b> | <b>Female</b> |
|--------------|--------------|---------------|
| 1            | VIII A1      | 30            |
| 2            | VIII A2      | 30            |
| 3            | VIII A3      | 30            |
| 4            | VIII A4      | 30            |
| 5            | VIII A5      | 30            |
| <b>TOTAL</b> |              | <b>150</b>    |

*Source: the documentation of Junior High School of Darul Hikmah Islamic Boarding School Pekanbaru Academic Year 2012-2013*

*Note: A is for female classes*

*B is for male classes*

The writer took all of the population as the sample, in case the population is below 100 individuals.<sup>5</sup>

Then, to determine the control and experimental classes, the writer used cluster sampling. The description of the sample can be seen as follows:

**Table III.3**  
**The Population and Sample**

| <b>NO</b> | <b>CLASS</b>        | <b>STUDENTS</b> | <b>SAMPLE</b> |
|-----------|---------------------|-----------------|---------------|
| <b>1</b>  | <b>Experimental</b> | <b>26</b>       | <b>26</b>     |
| <b>2</b>  | <b>Control</b>      | <b>26</b>       | <b>26</b>     |
|           | <b>TOTAL</b>        | <b>52</b>       | <b>52</b>     |

According to the table, the writer took 52 students as sample in this research.

### **E. The Technique of Collecting Data**

In this research, to collect the data of speaking skill, the writer used oral test technique which was done twice in pre-test and post-test. The students' voice was recorded in a camera digital.

The test consisted of pre test and post test. The pre test was given to the students in the experimental and control classes in order to know the basic of speaking skill from the students. The writer had given treatment to the students in the experimental class before the writer gave post test based on the procedures of Dialogue Memorization. The treatment was given to know the effect of Dialogue Memorization Technique toward students' speaking skill. Furthermore, there was no treatment in control class, so the writer used the conventional strategy that was used by the English teacher. At last, the

---

<sup>5</sup> Jack R. Fraenkel, Norman E. Wallen, *How to Design and Evaluate a Research in Education*, (New York: McGraw-Hill Companies Inc., 2006), p. 104

writer gave post-test to the experimental and control classes. The result of post test was as the final data in this research.

The students' speaking skill was measured by using oral language scoring rubric which was adopted from Adam and Frith in Hughes as follows:<sup>6</sup>

**Table III.4.**  
**The Category Level of Speaking Ability**

| Aspects       | Level6 | Level5 | Level4 | Level3 | Level2 | Level1 |
|---------------|--------|--------|--------|--------|--------|--------|
| Accent        | 6      | 5      | 4      | 3      | 2      | 1      |
| Grammar       | 6      | 5      | 4      | 3      | 2      | 1      |
| Vocabulary    | 6      | 5      | 4      | 3      | 2      | 1      |
| Fluency       | 6      | 5      | 4      | 3      | 2      | 1      |
| Comprehension | 6      | 5      | 4      | 3      | 2      | 1      |

**Table III.5**  
**Aspect Description for Accent**

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

**Table III.6**  
**Aspect Description for Grammar**

| Grammar |  |
|---------|--|
| 6       | No more than two errors during the interview   |
| 5       | Few errors, with no patterns of failure  |
| 4       | Occasional errors showing imperfect control of some pattern but no weakness that causes misunderstanding |

<sup>6</sup> Arthur Hughes. *Testing For Language Teacher*, ( Cambridge:Cambridge University Press. 2003 ) pp.111-113

|   |   |
|---|---|
| 3 | Frequent errors showing some major patterns uncontrolled and causing occasional irritation and misunderstanding |
| 2 | Constant errors showing control of very few major patterns and frequently preventing communication              |
| 1 | Grammar almost entirely inaccurate except in stock phrases  |

**Table III.7**  
**Aspect Description for Vocabulary**

| <b>Vocabulary</b> |   |
|-------------------|---|
| 6                 | Vocabulary apparently as accurate and extensive as that of an educated native speaker   |
| 5                 | Professional vocabulary broad and precise, general vocabulary adequate to cope with complex practical problems and varied social situations               |
| 4                 | Professional vocabulary adequate to discuss special interest, general vocabulary permits discussion of any non-technical subject with some circumlocution |
| 3                 | Choice of words sometimes inaccurate, limitations of vocabulary prevent discussion of some common professional and social topics                          |
| 2                 | Vocabulary limited to basic personal and survival areas (time, food, transportation, family, etc)   |
| 1                 | Vocabulary inadequate for even the simplest conversation  |

**Table III.8**  
**Aspect Description for Fluency**

| <b>Fluency</b> |   |
|----------------|---|
| 6              | Speech on all professional and general topics as effortless and smooth as a native speaker        |
| 5              | Speech is effortless and smooth, but perceptively non-native in speed and evenness                |
| 4              | Speech is occasionally hesitant, with some unevenness caused by rephrasing and grouping for words |
| 3              | Speech is frequently hesitant and jerky, sentence may be left uncompleted                         |
| 2              | Speech is very slow and uneven except for short or routine sentence                               |
| 1              | Speech is so halting and fragmentary that conversation is virtually impossible                    |

**Table III. 9**  
**Aspect Description for Comprehension**

| <b>Comprehension</b> |   |
|----------------------|---|
| 6                    | Understand everything in both formal and colloquial speech to be expected of an educated native speaker |

|   |   |
|---|---|
| 5 | Understand everything in normal educated conversation except for very colloquial or low-frequency items, or exceptionally rapid or slurred speech |
| 4 | Understands quite well normal educated speech when engaged in a dialogue, but requires occasional repetition or rephrasing                        |
| 3 | Understand careful, somewhat simplified speech when engaged in a dialogue, but may require considerable repetition and rephrasing                 |
| 2 | Understands only slow, very simple speech on common social and touristic topic; requires constant repetition and rephrasing                       |
| 1 | Understand to little for the simplest type of conversation  |

The score levels given to the students are about from level 1 to level 5

To collect the data, the speaking result was evaluated by using five components and each component had score or level. Each component had 20 as the highest score and the total of the components was 100. In this research, the writer took 80 as the highest score. Then the score was interpreted in the following category:<sup>7</sup>

1. 80 – 100 = A (Very good)
2. 66 – 79 = B (Good)
3. 56 – 65 = C (Enough)
4. 40 – 55 = D (Less)
5. 30 – 39 = E (Bad)

## 1. Procedures of the Research

In conducting the experimental research, the writer did some research procedures of both experimental and control groups. The research was carried out for six meetings. The research procedures are as follows:

### a. Conducting Pre-test

The pre-test was carried out to know the basic of students' speaking skill in both experimental and control groups. The test was choosing the topic and presenting the topic based on the topic in the text book used by the students.

### b. Conducting Treatment

---

<sup>7</sup> Suharsimi Arikunto. *Dasar-dasar Evaluasi Pendidikan : Edisi Revisi*. (Jakarta: Bumi Aksara, 2009), p. 245

The treatment was given only to the experimental group. The treatment was given based on the Dialogue Memorization procedures. The treatment was given as follows:

- a. The teacher asks the students to memorize the dialogue through mimicry.
  - b. The teacher asks the students to take role of one person in the dialogue.
  - c. The teacher asks the students to learn the one person's line, and memorize the other person's part.
  - d. After the dialogue has been memorized, teacher asks the students to perform the dialogue for the rest of the class in pairs.
- c. Conducting Post-test

The writer gave the post-test to both experimental and control groups after conducting the treatment for six meetings. The post-test was conducted to know the improvement of students' speaking skill after practicing dialogue memorization. The post-test was the similar model to the pre-test to know the students' speaking skill.

**Table III.10**  
**Topic of the Teaching Speaking in Each Meeting**

| No | Meeting | Topic                     |
|----|---------|---------------------------|
| 1  | I       | Getting Information       |
| 2  | II      | Agreeing and Disagreeing  |
| 3  | III     | Asking and Giving Opinion |
| 4  | IV      | Inviting People           |
| 5  | V       | Describing someone's job  |

|   |    |                 |
|---|----|-----------------|
| 6 | VI | Offering Things |
|---|----|-----------------|

## 2. Validity of the Test

According to Brown, a test is a method of measuring a person's ability, knowledge, or performance in a given domain<sup>8</sup>. According to Gronlund (in Brown), validity is the extent to which inferences made from assessment result are appropriate, meaningful, and useful in terms on the purpose of the assessment. According to Hughes, a test is said to be valid if it measures accurately what it is intended to measure<sup>9</sup>. Based on some experts above, the writer concluded 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.<sup>10</sup>

In this research, the writer used content validity. Brown states that content validity is if a test actually samples the subject matter about in which conclusions are to be drawn, and if it requires the test taker to perform the behavior that is being measured.<sup>11</sup> In conclusion, the test was given to the students based on the material that they have learned.

## 3. Reliability of the Test

Reliability is the measuring of test that is consistent and dependable.<sup>12</sup> It means, the test should consistently measure the person's ability. Furthermore Brown states that

---

<sup>8</sup> H. Douglas Brown, *Language Assessment: Principles and Classroom Practices*, (San Francisco:San Francisco State University, 2003), p. 3.

<sup>9</sup> Arthur Hughes, *Testing for Language Teacher*, (Cambridge: Cambridge University Press, 2005), p. 26.

<sup>10</sup> L.R. Gay and Peter Airasian.:*Educational Research Competencies for Analysis and Application. Sixth Edition.*( New Jersey : Pearson Education, 2000 ) pp. 163-167.

<sup>11</sup> H. Douglas Brown, *Op Cit.*, p. 22.

<sup>12</sup> *Ibid*, p. 20.

there are two scoring process 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 a common occurrences for classroom teachers because of the unclear scoring criteria, bias toward particular ‘good’ and ‘bad’ students, or simple carelessness.<sup>13</sup>

In this research, the writer used inter-rater reliability. It means that the scores of the test were evaluated more than one person. The students’ speaking scores were evaluated by two raters. Gay said that inter judge reliability can be obtained by having two or more judges independently score to be compared to the score of both judges. Then the scores of the rater 1 were correlated with the score of the rater 2. The higher correlation, the higher the inter judge reliability will be.<sup>14</sup> 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 16 Version:

**Table III.11**  
**Correlations**

|                            | RATER1 | RATER2 |
|----------------------------|--------|--------|
| RATER1 Pearson Correlation | 1      | .855** |
| Sig. (2-tailed)            |        | .000   |
| N                          | 26     | 26     |
| RATER2 Pearson Correlation | .855** | 1      |
| Sig.(2-tailed)             | .000   |        |
| N                          | 26     | 26     |

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

<sup>13</sup> *Ibid*,p:28

<sup>14</sup> L.R.Gay and Peter Airisian, *Op Cit*, p.175-176

From the output above, it can be seen that  $r$  calculation is 0.855 correlated to  $r$  table,  $df= 50$ . At level 5%  $r$  table is 0,273, while at level 1%  $r$  table is 0,354. Thus, the  $r$  observation is obtained higher than  $r$  table, either at level 5% or 1%. So the writer concludes that there is a significance correlation between score of rater 1 and score of rater 2. In the other words, the speaking test is reliable. The reliability of speaking test is very high.

To know the validity of the test, the writer used content validity. Referring to Bambang, if a measurement is as the representative of the ideas or the appropriate material that will be measured called content validity<sup>15</sup>. It means the test had fulfilled the validity of the content. In other words, the materials of the test had been taught at the second grade of MTs Darul Hikmah Pekanbaru. The materials were taken from guidance book for the students and other related resources. The writer prepared some topics based on the topics discussed at the time. The topic would be chosen randomly by students and they presented it in front of the class. And their voices were recorded by using camera digital.

## **F. The Technique of Analyzing the Data**

### **1. Different Score**

This analysis was used to find out the improvement of students' speaking skill that occurred before and after learning process that was calculated by D factor ( Different Score ). It is the difference between the pre-test and post-test.  $D=$  different score between pre-test and post-test.

---

<sup>15</sup> Ag.Bambang Setiyadi, *Metode Penelitian Pengajaran Bahasa Asing; Pendekatan Kuantitatif dan Kualitatif*, (Yogyakarta: Graha Ilmu, 2006 ) P:23

## 2. The Reliability and Validity of the Test

The test measuring students' speaking skill had to have reliability and validity. According to Gay, reliability is the degree to which a test consistently measures whatever it is measuring.<sup>16</sup> It is reflected in the obtaining how far the test or instrument test that enable to measure the same subject on different occasions indicating the similar result. In short, the characteristic of reliability is sometimes termed consistency.

In this research, to know the reliability of the speaking test, the writer used inters rater reliability. The writer had two raters in order to score the students' speaking skill. Gay says that inter judge reliability can be obtained by having two (more) judges independently score to be compared to the score of both (more) judges. The higher correlation, the higher the inter judge reliability will be. The writer used SPSS.16 Version to see the score correlation between the raters.

$r$  product moment can be obtained by considering the degree of freedom (df) as below:

$$df = N - nr$$

$N$  = Number of cases

$nr$  = The total variable correlated

Statistically the hypotheses are:

$$H_0 : r_0 < r_t$$

$$H_a : r_0 \geq r_t$$

$H_0$  was accepted if  $r_0 < r_t$  or there was no significant correlation between score from rater 1 and rater 2.

---

<sup>16</sup> L.R. Gay, op.cit., P. 169

$H_a$  was accepted if  $r_0 \geq r_t$  or there was a significant correlation between score from rater 1 and rater 2.

To know the validity of the test, the writer used content validity. Referring to Bambang, if a measurement is as the representative of the ideas or the appropriate material that will be measured called content validity.<sup>17</sup> It means the test had fulfilled the validity of the content.

### 3. T- test

In order to find out whether there was a significant effect on students' speaking skill between those taught by using dialogue memorization technique and those were not, the data were statistically analyzed. In this research, the writer used SPSS.16 Version to calculate the data. The result of T-test analyzing could be seen on the SPSS output. The significant level chosen in analyzing the score  $t_0$  (t- observed) was 5% or 0.05.

$H_a$  was accepted if:  $t_0 > t_t$

or if probabilities  $< 0.05$

It means there was a significant effect on students' skill in speaking between those taught by using dialogue memorization technique and those were not.

$H_0$  was accepted if:  $t_0 < t_t$

or if probabilities  $> 0.05$

It means, there was no a significant effect on students' skill in speaking between those taught by using dialogue memorization technique and those were not.

---

<sup>17</sup> Ag. Bambang Setiyadi, op.cit, p.23