

III. RESEARCH METHOD

3.1 Research Design

In conducting this research, the researcher used experimental method. The researcher used *Control Group Pretest-Posttest* design (Setiyadi, 2006, p.143). This experimental method deals with two groups; one is an experimental class and another as a control class. Each group received pre-test, treatments and post-test. Furthermore, one experimental class got treatment through *role play technique* and the control class got treatment through *jigsaw technique*.

This research design can be represented as follows:

G ₁	T ₁	X ₁	T ₂
G ₂	T ₁	X ₂	T ₂

Notes:

G₁ is an experimental class using role play

G₂ is a control class using jigsaw

T₁ is pre-test

X₁ is a role play treatment

X₂ is a jigsaw treatment

T₂ is post-test

3.2 Samples

The samples of this research were the second year students of SMP N 3 Bandar Lampung. There were six classes of the second year students and there was no rank for each class. It means that the ability of each class were same. The researcher took two classes as the sample of the research. Each classes both in the experimental and control classes were consisting 22 students. In determining the experimental classes, the classes were randomly selected by lottery and assigned to the class sample.

3.3 Variables

There are three variables in this research; they are one dependent and two independent variables;

1. The dependent variable is students' speaking ability
2. The first independent variable is role play technique, and
3. The second independent variable is jigsaw technique.

3.4 Research Procedures

The procedures of this research are as follows:

1. Determining the problem

The sample of this study was the second year students of SLTP 3 Bandar Lampung and most classes consisted only 25-30 students. The sample of this research was two classes, in determining the experimental classes, in order to get the same characteristic of students' speaking ability, the researcher randomly selected by lottery and assigned to the group.

2. Administering the pre-test

This test was given to experimental class and control class in order to know the students' speaking ability. It also administered in order to know the equality and the difference of the two classes. The researcher asked two raters to score the result.

3. Preparing the materials to be taught

The researcher arranged the materials would be discussed to each class by preparing media.

4. Implementing the techniques *role play* and *jigsaw* both in experimental classes.

In this term, the researcher applied two techniques to both classes. The experimental class taught by using Role Play and the other experimental class taught by Jigsaw. The experiment taught in two meetings for each class.

5. Administering the post-test to evaluate the result of the experiment

The test was given after the experiment to both classes in order to know the students' achievement after they receive the treatment. The researcher asked two raters to score the result.

6. Analyzing the data

The data analyzed by using normality test, homogeneity, and hypothesis test.

3.5 Data Collecting Techniques

In collecting the data, the researcher used the following steps:

1. Pre-test

The goal is in order to know the students' speaking ability before they are given the treatment. It is administered in order to know the equality and the difference of the two classes. It can be used to see whether the two classes have equal background knowledge or not. In this case the researcher used an individual presentation by giving the topics and information first then they should choose and prepare the presentation after that they should perform it in front of the class one by one.

2. Post-test

After conducting the treatment, the researcher gave post-test to both classes. It is the same like in the pre-test, the researcher used an individual presentation to know the result of the experimental classes, whether they have development or not.

3.6 Validity and Reliability

In this research, the researcher tried out the test to one of different classes to prove the tests had good quality or not. The test is absolutely considered as a good test if it has a good validity and reliability.

1. Validity of test

Hatch and Farhady (1982) says that validity is a matter of degree. It means that the test can be highly valid for one purpose but not for another. So here, we will consider that the test measures what is claimed to measure. To measure the test

has good validity, the researcher analyses the test from content and construct validity.

Content validity is the extent to which a test measures a representative sample of the subject matter content. The focus of content validity is one on the adequacy of the sample and not simply on the appearance of the test. It indicates that the items of the test should represent the material being discussed. Therefore we should select a representative sample for test purpose. In this research the topics chosen were *my best friend*, *jobs description*, and *my hobbies* inform of *descriptive text*. Those topics were based on the school curriculum as a matter of tailoring the lesson to students' need.

Construct validity is to measure the ability, which it is supposed to be measured. If we attempt to measure that ability in a particular test, then that part of the test would have construct validity only if we were able to demonstrate that we were indeed measuring just that ability. It means that the test items should really test the students whether they have improved speaking ability that have been taught or not; or the test items should really measure the fluency, accuracy, and comprehensibility.

To measure construct validity the researcher uses inter raters' validity. It means that the items of the test should represent the material being discussed. The material is measured by inter rater validity to find the degree of agreement. In inter rater validity; there are two raters who judged the validity of the test.

In this research the researcher became one of the raters and the other inter rater was Astuti, S.Pd. She is one of the English teachers at SMP N 3 Bandar Lampung.

Moreover Arikunto (1986:64) says that a test is valid if the instrument can measure what it should be measured. Besides, Shahomy (1985: 75) adds that it also examines whether the test is a good representation of the material, which needs to be tested. In this research, the researcher focused on the construct validity, since she wants to test how well the test predicts or estimates a particular performance.

2. Reliability of Test

Reliability refers to the extent to which the test is consistent in its score, and it gives an indication of how accurate the test scores are. Arikunto (1992) states that a test has a good reliability if the instrument of the test can indicate the stability of the scores; in the other words, among the scores there have no high differences.

To ensure the reliability of scores and to avoid the subjectivity of the researcher, inter rater reliability is applied in this research. Inter rater reliability is used when

score of the test is independently estimated by two or more judge. To achieve such reliability, in judging the students' speaking performance, the researcher:

1. Uses a speaking criteria based on what Heaton (1991). The focuses of speaking skills that had been assessed are;
 1. fluency
 2. accuracy
 3. comprehensibility
2. Involves second experience rater in using the profile to give judgment for each students' speaking performance. The second rater is English teacher who has experience in rating students' speaking. This is meant to provide a consistent and fair judgment.

Thus, to determine the level of reliability of the scoring system, the Spearman Rank Correlation is applied on the data. The formula of this is:

$$R = 1 - \frac{6 \cdot \sum d^2}{N(N^2 - 1)}$$

Notes:

R : Reliability

N : Number of students

D : The different of rank correlation

1-6 : Constant number

(Sudijono, 2006)

The researcher considers it is reliable for the test if the test has reached range 0.60-0.79. The standard of reliability:

- | | |
|----------------------------|---------------------------|
| A. a very low reliability | ranges from 0.00 to 0.19 |
| B. a low reliability | ranges from 0.20 to 0.39 |
| C. an average reliability | ranges from 0.40 to 0.59 |
| D. a high reliability | ranges from 0.60 to 0.79 |
| E. a very high reliability | ranges from 0.80 to 0.100 |

Slameto (1998:147)

The researcher considers that both raters achieved the reliability if the inter rater reliability has reached ranger 0.60 to 0.79 (a high reliability).

In this research, it was found that the result of inter-reliability of pre-test and post-test was as follows:

Control Group

Inter-raters Reliability in Pre-test

$$R = 1 - \frac{6 \cdot \sum d^2}{N(N^2 - 1)}$$

$$R = 1 - \frac{6 \cdot (10)}{22 \cdot (22^2 - 1)}$$

$$R = 1 - \frac{60}{10626}$$

$$R = 1 - 0.0056$$

$$R = 0.99$$

It means that both raters have a very high reliability.

Inter-raters Reliability in Post-test

$$R = 1 - \frac{6 \cdot \sum d^2}{N(N^2 - 1)}$$

$$R = 1 - \frac{6 \cdot (9)}{22 \cdot (22^2 - 1)}$$

$$R = 1 - \frac{54}{10626}$$

$$R = 1 - 0.0051$$

$$R = 0.99$$

It means that both raters have a very high reliability.

Experimental Group

Inter-raters Reliability in Pre-test

$$R = 1 - \frac{6 \cdot \sum d^2}{N(N^2 - 1)}$$

$$R = 1 - \frac{6 \cdot (10)}{22 \cdot (22^2 - 1)}$$

$$R = 1 - \frac{60}{10626}$$

$$R = 1 - 0.0056$$

$$R = 0.99$$

It means that both raters have a very high reliability.

Inter-raters Reliability in Post-test

$$R = 1 - \frac{6 \cdot \sum d^2}{N(N^2 - 1)}$$

$$R = 1 - \frac{6 \cdot (7)}{22 \cdot (22^2 - 1)}$$

$$R = 1 - \frac{42}{10626}$$

$$R = 1 - 0.0039$$

$$R = 0.99$$

It means that both raters have a very high reliability.

3.7 Data Analysis

After collecting the data, the researcher analyses the data by using the following procedures:

1. Normality Test

Normality test is used to measure whether the data in experimental class and control classes are normally distributed or not.

H_0 : The data is not distributed normally

H_1 : The data is distributed normally.

In this research, the criteria for the hypothesis are H_0 is accepted if $p > \alpha$ and the researcher used level of significant 0.05.

2. Homogeneity Test

The Homogeneity test is used to know whether the data in the experimental class and control class are homogenous or not.

In this research, the research used Independent Sample Test (SPSS 15) to know the homogeneity of the test.

H_0 : The data is not homogenous

H_1 : The data is homogenous.

In this research, the criteria for the hypothesis are H is accepted if $F\text{-ratio} > \alpha$ and the researcher used level of significant 0.05.

3. Hypothesis Test

In administrating hypothesis test, t-test is used. Its function is to know the difference between two scores compared is significant or not. The data are analysed by using Independent Group T-test. This test is used when we want to compare the means of two different groups and the data from the two groups are taken from different situations too (Setiyadi, 2001:60). The normality test, homogeneity test, and the hypothesis test will be counted by using SPSS formula (Statistical Package for Social Science).

The criteria are:

1. Non-directional two tailed hypothesis

The possible hypotheses are:

a) The null hypothesis (H_0)

There is no difference between Role Play technique (U1) and Jigsaw technique (U2) towards the students' speaking ability achievement. $H_0 = U1 = U2$

b) The alternative hypothesis (H_1)

There is a difference between Role Play technique (U1) and Jigsaw (U2) technique towards the students' speaking ability achievement.

$$H_0 = U1 = U2$$

Testing formulation:

Accept H_0 if $t\text{-table} < t\text{-ratio} < t\text{-table}$

Reject H_0 if $-t\text{-table} > t\text{-ratio} > t\text{-table}$

(Hatch and Farhady, 1982: 126)

4. Scoring system

The focuses of speaking skills that had been assessed are;

3. fluency
4. accuracy
5. comprehensibility

These criteria based on what Heaton (1991) proposed. The score is in scale 0-6 for each skill. In order to make scoring easier each scale will be multiplied by 5 and plus 10 to make maximum score 100.

Example if student gets 4 for accuracy, 3 for fluency and 4 for comprehensibility so the score will be:

$$\text{Accuracy} \quad : 4 \times 5 = 20$$

$$\text{Fluency} \quad : 3 \times 5 = 15$$

$$\underline{\text{Comprehensibility}} \quad : 4 \times 5 = 20 +$$

$$55 + 10 = 65$$

Therefore, the score is 65.

In scoring the test, the researcher used inters rater method to score the students' result in pre-test and post-test. Beside the researcher, the other persons who have competency in English will be asked to score the students' result. So, there were two raters who will score the students' result. After two raters gave score by using speaking criteria proposed by Heaton, the result of the two raters added and divided by two to get average score for each student. For example, if rater 1 gives score 70, rater 2 gives score 80, all the scores will be added and then divided by three and the final score is 75.

Let us see the following data:

No	Students' Code	R1	R2	Average Score
1	TJS	75	78	76,5
2	MSR	80	82	81
3	MDAS	60	65	62,5

Scoring Criteria

The criteria for marking speaking are for rater proposed by Heaton (1991). The speaking test that measured is individual performance delivering one's daily activity. The elements of speaking which had been used in this research were as follows:

Table 3.1 Table of Specification of Speaking Assessments

Elements of Speaking Assessment	Descriptions
Accuracy	Covering pronunciation, grammar and vocabulary
Fluency	Covering fairly wide range of expression and responding well without difficulty.
Comprehensibility	Understanding the speaker intention and general meaning.

Table 3.2 The Rubric of Grading System

Rating	Accuracy	Fluency	Comprehensibility
6	Pronunciation is only very slightly influenced by mother tongue. Two or three a minor grammatical and lexical errors.	Speaks without too great an effort with fairly wide range of expression. Searches for words occasionally one or two unnatural pause.	Easy for listener to understand the speaker intention and general meaning. Very few Interruptions or clarification required.
5	Pronunciation is slightly influences by mother tongue. A few minor grammatical and lexical errors but most utterances are correct.	Has to make an effort at times to search for words. Nevertheless, smooth delivery on the whole and only a few unnatural pauses.	The speaker's intention and general meaning are fairly clear. A few interruptions by the listener for the sake pf clarification are necessary.
4	Pronunciation is still moderately influenced by the mother tongue but no serious phonological errors, a few grammatical and lexical errors but only one or two major errors causing confusion.	Although he has to make an effort and search for words, there are no too many unnatural pauses. Fairly smooth delivery mostly. Occasionally fragmentary but succeed in conveying the general meaning. Fair range of expression.	Most of what speaker says is easy to follow. His intention is always clear but several interruptions are necessary to help him to convey the message or to seek clarification.
3	Pronunciation is influenced by the mother tongue but only a few serious phonological errors some of which cause confusion.	Has to make an effort for much of the time, often has to search for desired meaning. Rather halting delivery and fragmentary. Range of expression often limited.	The listener can understand slot of what is said m but he must constantly seek of clarification. Cannot understand many of the speaker's more complex or longer sentences.
2	Pronunciation is seriously influenced b y mother tongue with error causing a break down in	Long pauses while he searches for the desired meaning. Frequently fragmentary and halting	Only small bits (usually shorts sentences and phrases) can be understood and then

	communication. Many “basic” grammatical and lexical errors.	delivery .almost gives up in making the effort at the times, limited range of expression.	with considerable effort by some one who is used to listening to the speaker.
1	Serious pronunciation errors as well as many basic grammatical and lexical errors. no evidence of having mastered any of the language skills and areas practiced in the course,	Full of long and unnatural pauses, very halting and fragmentary delivery. At times gives up making the effort. Very limited range of expression.	Hardly anything of what is said can be understood. Even when the listener makes great efforts or interrupts s, the speaker is unable to clarify anything he seems to have said.

The interpretation of grading system is as follows:

- 6 : Excellent
- 5 : Very Good
- 4 : Good
- 3 : Fair
- 2 : Poor
- 1 : Moderate