

III. RESEARCH METHODS

This research is intended to find out whether Role Play Technique can be used to improve students' speaking performance. This chapter includes the research design, the population and sample, data collecting technique, research procedures, validity and reliability of the test, and criteria for evaluating students' speaking ability, data analysis and hypotheses testing.

3.1 Research Design

This research was a quantitative study which used *one group pretest-posttest* design. There was one class as the experimental class which was chosen purposively. It is carried out to see and find out the result of implementing Role Play Technique in improving students' speaking performance. The design used in this research was one group pretest-posttest design. It means before the first teaching, pretest was carried out and after three teachings, a posttest was conducted. The research design could be presented as follows:

T1 X T2

Note: T1 : Pre-test

T2 : Post-test

X : Treatment by using components of Role Play Technique

(Setiyadi, 2006:133)

In conducting this research, the researcher used inter-rater reliability and computed the average score the oral test of the pre-test and post-test.

3.2 Population and Sample

The population of the research was the first year students of SMA N 1 Natar that consists of 9 classes, and X 8 was taken as the sample by purpose since based on observation in pre-research, it was found that the students had crucial problems in speaking performance.

3.3 Data Collecting Technique

In collecting the data the researcher purposed to take two components that are the product and the process. The first component is to answer the first research question and the second component is to answer the second research question provided by the researcher.

3.3.1 Pre-test

In this oral test, the students were assigned to practice dialogue concerning topics of describing jobs and describing daily schedules model of conversation. To do this, they may talk about their hobbies, favorite activities and work.

Here are the procedures:

1. The students worked in pairs which consist of two students.

2. Each pair should make a conversation based on the topic given (describing peoples' jobs).
3. Each pair had 5 minutes to practice their dialogue
4. Each pair was free to express their ideas and to speak as clearly as possible since their voice was recorded by the teacher.

These following words or expressions could be used in their dialogue,

ex: Andrea is a receptionist.

She works by answering the phone.

She works in an office.

It is interesting for her because she gets a lot of knowledge from that.

The researcher administered the pre-test which takes 80 minutes. The purpose of the test was to find out how far the students ability in mastering speaking skill. In administering the pre-test, the researcher provided a topic to the students and let them to create a conversation based on the topic provided in their own language. The test was done in pairs. Then the researcher recorded the students' conversation in a cassette. The topics were "describing people's jobs and describing daily schedules".

3.3.2 Post-test

In this oral test, the students were assigned to practice dialogue concerning topics of daily schedules model of conversation. To do this, they might talk about their routine works, favorite activities and school.

Here are the procedures:

1. The students worked in pairs which consist of two students.
2. Each pair should make a conversation based on the topic given (daily activities).
3. Each pair had 5 minutes to practice their dialogue.
4. Each pair was free to express their ideas and to speak as clearly as possible since their voice was recorded by the teacher.

These following words or expressions can be used in your dialogue,

ex: I usually get up at 6:00 in the morning on weekdays.

I sometimes go to bed at midnight on weeknights.

I always start work at 11:30 at night.

I never stay up until midnight on Mondays.

The researcher administered the post-test which takes 80 minutes. The purpose of this test is to know the students' improvement in speaking after the researcher gives treatments by using Role Play Technique. The technique of giving post-test is similar to the pre-test because the researcher wants to know the students' improvement after three treatments.

The students' record result of pre-test was scored by two raters, the researcher himself and another English teacher in order to have reliable test result. In scoring the students result, the rater used the oral ability scale proposed by Heaton (1978). The researcher provided the cassettes which consist of students' oral record.

3.3.3 Interview

The interview was done in order to get valid data and to find out the students' and the teacher's perception about the teaching learning process. These interviews are aimed at getting an accurate data from both sides, and it was very useful to be as a media cross check on both students' and teacher's opinion about the Role Play Technique. Fraenkel (1990: 385) describes that interview is an important way for a researcher to check the accuracy of the impressions of what he or she has gained through observation. The purpose of interviewing people is to find out their mind, what they think or how they feel about something. To help the researcher collect the data, he used voice-recorder to save the data so that he can analyze it and also to re-ensure that the scores given are suitable with the performance.

3.3.4 Observation Sheet

In this observation, the researcher was a non-participant observer. He asked the English teacher who had been previously confirmed by him to teach speaking skills through Role Play Technique. Then, the researcher and other rater observed teaching learning process based on the Role Play Technique which has been provided in the observation sheet. Setiyadi (2006:101) specifies the purpose of observation is to explain the situation being investigated: activities, person, or individuals involved in an activity and relationship among them.

3.4 Research Procedures

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

3.4.1 Selecting speaking material

In selecting speaking material the researcher used the syllabus at the second year of SMA N 1 Natar on KTSP (an English operational curriculum which is arranged and applied by each education unit) which is the newest curriculum used by the school. The topics chosen are “*Describing People’s Jobs and Describing Daily Schedules*” in the form of dialogue.

3.4.2 Determining the Instrument of the Research

The instrument of the research is speaking test. The researcher conducted the speaking test for pre-test and post-test. These tests aim at gaining the data that is the students’ speaking achievement score before the treatment and after the treatment in performing short dialogue.

3.4.3 Determining Population and Sample

The population of the research was the first year students of SMA N 1 Natar that consisted of 9 classes and X 8 which consisted of 40 students was taken as example. The sample was purposively selected based on pre-observation and interview from the teacher since the class had crucial problem in speaking performance and should be improved.

3.4.4 Conducting Pre-test

Pretest was given before the researcher applies the treatments (teaching speaking through components of Role Play Technique). The test is speaking test in the form of dialogue. The material tested is in the form of dialogue and based on KTSP.

Pretest is given to know how far the competence of the students in speaking skill before treatments. The test was held for 80 minutes, and the topics chosen are like and dislike and daily activities in the form of dialogue. The form of the test is subjective test since there is no exact answer though the researcher uses inter-rater reliability to make the test reliable.

3.4.5 Giving the Treatment

The researcher presented the material for treatment through applying Role Play Technique. There were three times treatments in this research. Each treatment was held for 80 minutes. In selecting the speaking material the writer used the syllabus of the second year of SMA N 1 Natar based on KTSP which is the newest curriculum used by the school. The topics chosen were describing peoples' jobs and describing daily schedules in the form of dialogue.

3.4.6 Conducting Post-test

The researcher administered the post-test after treatment, which last 80 minutes. It aims to know the progress of students' speaking achievement after being given the treatment using the components of Role Play Technique.

In conducting the post-test the researcher provided some topics and let them make a short dialogue in pair based on the topics provided. The test was done orally and directly. The teacher called each pair one by one in front of the class to perform their dialogue. The researcher asked the students to speak clearly since their voice was recorded during the test.

3.4.9 Analyzing, Interpreting, and Concluding the Data Gained

After collecting the data that is students' utterance in performing the dialogue and listening from the recorder, the data analyzed referring the rating scale namely pronunciation, fluency, comprehensibility. Then interpretation of the data is done. First, the researcher scored the pre-test and post-test, then tabulated the result of the test and calculate the mean of pretest and post-test. Finally the researcher drew the conclusion from the tabulated result of pre-test and post-test that was statically analyzed by using Repeated Measure T-Test of SPSS (statistical package for social science) version 17.0 for social science. The data was gained from one group and the researcher intended to find out whether there is a significant improvement of students speaking achievement.

3.5 Validity of the test

A test could be said valid if the test measures the object to be measured and suitable with the criteria (Hatch and Farhady, 1982:250) in Setiyadi 2006.

According to Hatch and Farhady (1982:281) there are types of validity. They are content validity and construct validity.

3.5.1 Content Validity

Content validity is the extent to which a test measures a representative sample of the subject matter content (Hatch and Farhady, 1982:251). The focus of this content validity is 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 the researcher selected a representative sample of test purpose and the materials given are suitable with the curriculum since the researcher has taken the school-based syllabus during his pre-research which was already approved by the teacher as an interrater so that the materials are in line with indicators achieved. In this case, the researcher gave the speaking material based on the second semester syllabus of the first grade of SMAN 1 Natar. The materials are *asking and describing people's jobs and describing daily schedules* inform of *descriptive text*. To measure students' ability and to find out whether there is an improvement, the researcher provided speaking pre-test and post test consisting of materials based on the syllabus from the school.

3.5.2 Construct Validity

Construct validity is to measure the ability which supposed to be measured. If we attempt to measure that ability in particular test, then the part of the test had construct validity only if we are able to demonstrate that we are measuring just that

ability. It means that the test item should really test the students whether they have improved speaking ability after being taught. The test items have to measure the fluency, accuracy, and comprehensibility.

To measure construct validity, the researcher uses inter-rater agreement of agreeing or disagreeing materials. 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 were two raters who judge the validity of the test. In this research the researcher became one of the raters and the other one was Dian Oktaviana, S.Pd., the English teacher at SMA N 1 Natar.

In addition, Arikunto (1996:64) states that a test is valid if the instrument can measure what should be measured. 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 he wants to test how well the test predicts or estimates a particular performance.

3.5.3 Reliability

Speaking performance cannot be measured by multiple choice or questions answered in a paper yet it should be performed and the evaluation is often subjective. 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 (1996) states that a test has a good reliability if the instrument of the test can

indicate the stability of the scores. In other words, there have no high differences among the scores. Here are the tables to score students' speaking performance:

1. Putting the data of pre-test and post-test score on the table below:

Elements	Score		Total	
Accuracy	R1	R2	R1	R2
Fluency				
Comprehensibility				
N=			X1=	X2=

2. Row data of oral test

No.	Elements	Rater 1		Rater 2	
1	Accuracy	Pre-test	Post-test	Pre-test	Post-test
2	Fluency				
3	Comprehensibility				
...					

3. Inter-rater reliability of pre-test

No.	Elements	R1	R2	d1	d2
1	Accuracy				
2	Fluency				
3	Comprehensibility				
...					

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

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

1. Uses a speaking criteria based on what Heaton (1991). The focusses of speaking skills that were scored are:
 1. Fluency
 2. Accuracy
 3. Comprehensibility
2. Involves second experienced rater in using the profile to give judgement for each students' speaking performance. The second rater is the teacher who has experience in rating students' speaking performance. This is meant to provide a consistent and fair judgement.

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

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

Notes:

R : Reliability

N : Number of students

d : The difference of rank correlation

1-6 : Constant number

(Sudijono, 2006)

The researcher considers it reliable for the test if the test has reached range 0.60-0.79. Here is 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

In scoring the test, the researcher used inter-rater method to score the students' result in pre-test and post-test. Beside the researcher, the English teacher was asked to score the students' result. Therefore, there were two raters who scored the students' result. After two raters gave the score by using criteria proposed by Heaton, the result of the two raters were combined and divided by two to get the average score for each student. For example, if rater 1 gave score 80, rater 2 gave 70, then all scores were combined and divided by two and the final score was 75.

Let us see the example of students' score:

No.	Students' Code	R1	R2	Average Score
1	ANS	80	70	75
2	RYN	72	70	71
3	ZAK	70	68	69

1. Interrater Reliability of Pretest

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

$$R = 1 - \frac{6.193}{40(40^2-1)}$$

$$R = 1 - \frac{1.158}{40(1600-1)}$$

$$R = 1 - \frac{1.158}{40.1.599}$$

$$R = 1 - \frac{1.158}{63.960}$$

$$R = 1 - 0.018105$$

$$R = 0.981895$$

From the data calculated above, it is found that the two raters have a very high reliability.

2. Interrater Reliability of Posttest

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

$$R = 1 - \frac{6. 278}{40(40^2-1)}$$

$$R = 1 - \frac{1.668}{40(1600-1)}$$

$$R = 1 - \frac{1.668}{40.1.599}$$

$$R = 1 - \frac{1.668}{63.960}$$

$$R = 1 - 0.0260788$$

$$R = 0.9739212$$

From the data calculated above, it is found that the two raters have a very high reliability.

3.6 Criteria of Evaluating Students' Speaking Performance

In evaluating the students' speaking score, the writer and another rater listened to the students' record and use the oral ability scale proposed by Heaton (1991). In scoring the test, the researcher implemented analytical scoring which cover, accuracy, fluency and comprehensibility separately. To find out the average score of students, the researcher also uses formula as follow:

Table 3.1 Table of Specification of Speaking Assessments

This table is to describe each aspect of speaking.

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 in general meaning.

Table 3.2 The Guideline of Grading System

This table is as the guideline to assess students' speaking performance.

Rating	Accuracy	Fluency	Comprehensibility
6	Pronunciation only very slightly influenced by mother-tongue	Speaks without too great an effort with a fairly wide range of expression.	Easy for listener to understand the speaker attention and general

		Searches for words occasionally but only one or two unnatural pauses	meaning
5	Pronunciation is slightly influenced by mother-tongue. Most utterances are correct.	Has to make an effort at times to search for words. Nevertheless smooth very delivery on the whole and only a few unnatural pauses	The speakers' intention and general meaning are fairly clear. A few interruption by listener for the sake of clarification are necessary.
4	Pronunciation is still moderately influenced by mother-tongue but no serious phonological errors	Although she/he has made an effort and search for words, there are not too many unnatural pauses. Fairly smooth delivery mostly.	Most of the speakers' utterances are easy to follow. His intention is always clear but several interruptions are necessary to help him to convey the message or to see the clarification.
3	Pronunciation is influenced by the mother-tongue but only a few serious phonological errors	Has to make an effort for much of time. Often has to search for the desired meaning. Rather halting delivery and fragmentary.	The listener can understand a lot of what is said, but he must constantly seek clarification. Cannot understand many of the speakers' more complex and longer sentences.
2	Pronunciation is influenced by the mother-tongue with errors causing a breakdown in communication	Long pauses while he searches for the desired meaning. Frequently halting delivery and fragmentary. Almost	Only small bits(usually short sentences and phrases) can be understood- and then with considerable effort by

		gives up for making effort at times	someone who us used to listening to the speaker.
1	Serious pronunciation errors. No evidence of having mastered any of the language skills and area practiced in course.	Full of long and unnatural pauses. Very halting and fragmentary delivery. At times gives up making effort.	Hardly anything of what is said can be understood. Even when the listener makes a great effort or interrupts, 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

3.7 Data Analysis

The researcher analyzed the data statistically using repeated measure t-test. t-test is probably is the most widely used statistical test for the comparison of two means because it can be used with very small sample size (Hatch and Farhady, 1982:108). Repeated measure T-test is to analyze the data of students' speaking improvement. The researcher only used one class for experimental class and there is no control group. He counted the data by using SPSS 17.0 (Statistical Package for Social Science).

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

1. Normality Test

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

H^0 : the data is not normally distributed.

H^1 : the data is normally distributed

2. Hypothesis Test

In administering hypothesis test, T-test is used. Its function is to find out the difference between two scores compared which are given before and after treatment is significant or not. The data are analysed by using Repeated Measures T-test. This test is used to compare two types of data taken from the same samples (Setiyadi, 2006:170). The normality test and hypothesis test was counted by using SPSS (Statistical Package for Social Science).

3. Scoring system

The focuses of speaking skills assessed are:

1. Fluency
2. Accuracy
3. Comprehensibility

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

For example, if student gets 4 for accuracy, 5 for fluency and 3 for responsibility, the score was:

Accuracy : $4 \times 5 = 20$

Fluency : $5 \times 5 = 25$

Comprehensibility : $3 \times 5 = 15 +$

$$55 + 10 = 65$$

Therefore the score was 65.

Besides that, the observation and the interview were constructed. The interview in the experimental class was in the form of open questions (the questions must be in the form of explanation or description rather than “yes” or “no” answers, to avoid the students of being reluctant to answer the questions given) and formal type, was constructed; to analyze its qualitative data in order to find out the factors that make students improve their speaking performance through role play technique, matrix analysis, in this case description analysis was used, since the researcher used his own idea, including his own interpretation toward the data. (Setiyadi, 2006:262).

3.8 Hypotheses Testing

The hypotheses are:

H⁰: there is no any significant improvement toward students' speaking achievement after being taught through Role Play Technique.

H¹: there is a significant improvement toward students' speaking achievement after being taught through Role Play Technique.

3.9 Statistical Testing

Repeated Measures T-test was used as the data came from the same sample or known as paired data (Hatch and Farhady, 1982:114).

The data were analyzed by using T-test. The formula was as follow:

$$t = \frac{X1 - X2}{S_{D-}} \quad \text{in which} \quad S_{-} = \frac{S_D}{\sqrt{n}}$$

Notes:

X1 = Mean of pre-test

X2 = mean of post-test

S_{D-} = Standard error of differences between two means (denominator)

S_D = Standard deviation

n = Number of students

(Hatch and Farhady, 1982:116)

3.10 Schedule of the Research

No.	Activities	Time	Day	Date
1.	Pre-Research	08.00 a.m. - 09.20 a.m.	Saturday	March 3 rd , 2012
2.	Observation	08.00 a.m. - 09.20 a.m.	Tuesday	April 10 th , 2012
3.	Pre-test	07.20 a.m. - 08.40 a.m.	Saturday	April 21 th , 2012
4.	First Treatment	08.40 a.m. - 10.00 a.m.	Tuesday	April 24 th , 2012
5.	Second Treatment	08.10 a.m. - 09.30 a.m.	Saturday	April 28 th , 2012
6.	Third Treatment	08.10 a.m. - 09.30 a.m.	Tuesday	May 1 st , 2012
7.	Post-test	07.20 a.m. - 08.40 a.m.	Saturday	May 05 th , 2012
8.	Interview	08.10 a.m. - 09.30 a.m.	Tuesday	May 08 th , 2012