III. RESEARCH METHOD

This chapter discusses about the methods of research used in this study, they are: research design, population and sample, data collecting technique, research procedures, scoring system, and data analysis.

3.1 Research Design

The research is quantitative by nature. There was one class as the participants' class, and the research design was one group pre-test post-test design. The researcher used short story in order to increase the students' listening achievement. The treatment was given three times. Here is the design:

$$T1 \ X \ T2$$

Where :
T1   = Pretest, which is administered before receiving the treatment
X   = Treatment, which is teaching listening by using short story
T2   = Posttest, which is administered after receiving the treatment
              (Hatch and Farhady, 1982:20 in Setiyadi,2006:132)

The first activity done by the researcher was to administer try out to XI social science 2 to make the instrument more valid and then the researcher administers pretest to the experimental class in order to find out the input of the students before they got the treatment. After that, the researcher conducted the treatment that was using short story to stimulate and increase the students' listening comprehension achievement. The treatment was given three times in assumption

three times treatment was enough and also considered the curriculum from the government that stated that learning narrative story was only three meetings. The next step was administering the post test to experimental class to know the result of the treatment.

3.2 Population and Sample

The population of the research was the second grade of MA Daarul Ma'arif Natar. There were two classes of the second grade students and each class consists of 30 students. Between those classes, the researcher took only one class as the sample of the study using random sampling to make them possible to be chosen. The class, that was the participants' class of the research, was XI-1 social program.

3.3 Data Collecting Technique

In collecting the data the writer used the following technique:

1. **Administering the Pre-test**

    The pre-test was given before the treatment, in order to find out how far the competence of the students in listening comprehension or their input before the treatment and to find out the experimental class' listening comprehension achievement, the test was multiple choices that consist of thirty items. The materials tested, was related to the curriculum used in the school and suitable with their level.

2. **Administering the Post-test**

Post-test was given after the treatment in order to find out whether there was any significant increase of students' listening comprehension achievement. The test was multiple choices consisted of thirty items. The materials tested, were relatied to the curriculum used in the school and suitable with their level. The post-test was done after three meetings of the treatments. The result of the post-test of the participant class was analyzed.

3.4 Procedure of the Research

The researcher used the following steps in order to collect the data:

1. **Determining research problem**

Based on the researcher's background of problem in the first chapter, it was assumed that short narrative story could be used to improve the students listening comprehension achievement. And the researcher tried to find out whether there was significant improvement of learner's achievement in listening comprehension skill before and after being taught using short story

2. **Selecting instruments materials**

The research used three short stories for treatments. The material was based on KTSP 2006. The stories were taken from students' handbook and also from the internet.

3. **Determining research instruments**

The scope of the research was macro skill of listening comprehension to obtain general idea of the story and specific information for lower level

listening. Thus, the test was listening comprehension test in order to find out the students listening comprehension ability.

4. **Determining Sample**

The sample was the students chosen, that was the second grade of social science 1, The writer selected the sample by using random sampling with the assumption that the second year classes of MA Daarul Ma'arif Natar had the same characters and level of English Proficiency.

5. **Conducting try-out**

The try-out had been conducted before the pre-test was administered. This was expected to measure the validity and reliability of pretest and posttest, to ensure the data used by the researcher was valid and reliable to use as a research instruments. This test was multiple choice tests and was conducted in 90 minutes. There were 30 items of multiple choices with four options and one of them was as the correct answer, the test items could be reduced or kept depends on its relability and validity. The aim of try -out was to determine the quality of the test used as the instrument of the research, and to determine which item should be revised for the pre-test and the post-test. This research used the result of the try-out test to measure the level of difficulty and discrimination power, to find out the validity and reliability of the test.

A good test should meet four criteria: a good validity, reliability, and level of difficulty and discrimination power.

Criteria of Good Tests

The tests would be said good if they had several criteria, they had validity, reliability, level of difficulty, and discrimination power.

1. **Validity of Test**

A test can be said valid if it measures what is supposed to be measured and nothing else (Heaton, 1977:159). There were several types of validity, but the researcher only focused on construct validity, content validity, and face validity.

a. **Construct Validity**

According to Thorndike and Daniel (2002:247) construct validity relates to the scientific value or appropriateness of a measure, to be constructing into valid, a measure must relate constructs to the real world observations. Thus, construct validity was concerned with whether the test is actually in line with the theory of what listening comprehension means or not.

Construct validity examines whether the test is actually in line with the theory of what it means to know the language, whether the test is actually a reflection of what it means to know a language (Shohamy, 1985: 74-75). It is concerned with the theory of testing certain language skill; in this research, listening. The table of specification as follows:

b. Content Validity

The next one was content validity. It concerned for an educational test achievement, it means every test item should come from both domain, cognitive and content process and cover all aspects of the domain (Thorndike and Daniel, 2002:247). Moreover Hatch & Farhady (1981:250) states that this kind of validity concerns on adequency of the sample, not simply on the appearance of the test.

Content validity refers to the good reflection of the material that will be tested. It means that the test should represent the material that has been discussed before. To get the content validity of the test, the writer adopted the test based on the students' handbook and the curriculum used. The writer arranged the test based on the material that had been taught to the students.

2. Reliability of Tests

Reliability concerns with stability, consistency, or reproducibility of measurement. The item was said reliable if it resulted the same in scale value for an individual every time the tests was administered. According to Heaton (1991:162), reliability is a necessary characteristic of any good tests, if the tests are given to the same person in other time without any treatment or language learning then it produces different significance results it is no where reliable.

To find out the reliability of the test the researcher used split-half method. Split-half method means the test was divided into halves; the first half and the second half. The first half contained the question number 1 to 15 and the second half covered question 16 to 30 (Hatch & Farhady, 1982:246). To measure the coefficiency of theirs reliability, the researcher has used Pearson Product Moment, and here was the formula:

$$r_1 = \frac{\sum x}{\sqrt{(\sum x^2)\,(\sum y^2)}}$$

Where :
r1 = coefficient reliability between 1st half and 2nd half groups
x   = total number of 1st half group
y   = total number of 2nd half group
x2 = square of x
y2 = square of y

Then to know about the coefficiency of their correlation as whole the researcher used Spearman Brown's formula and here is the formula:

$$r_k = \frac{2r_1}{1 + r_1}$$

Where :
r_k = full test reliability
r_l = half test reliability

The criteria of reability are:

0.80-1.00   = very high

0.60-0.79   = high

0.40-0.59   = average

0.20-0.39   = low

0.00-0.19   = very low   (Hatch and Farhady,1982:246)

3. Level of Difficulty

To see the level of difficulty the researcher used the following formula:

$$F = \frac{R}{N}$$

Where :

FV = Index of difficulty
R   = Number of the students answers correctly
N   = Total number of the students

Heaton (1975:182) states the criteria of the index of difficulty. Here are

the criteria:

LD<0.30   = Difficult

LD=0.30-0.70   = Satisfactory

LD>0.70   = Easy

4. Discrimination Power

Discrimination power was the proportion of the high group students

getting the items correct minus the proportion of the low level students

who getting the items correct. Here was the formula to calculate it:

$$D = \frac{U - L}{1/2\,N}$$

Where :
DP = Discrimination Power
U   = Number of upper group who answer correctly
L   = Number of lower group who answer correctly
N   = Total number of the students

Heaton (1975:182) states criteria of the discrimination power result, and here were the criteria:

DP: 0.00-0.19 = Poor

DP: 0.20-0.39 = Satisfactory

DP: 0.40-0.69 = Good

DP: 0.70-100 = Excellent

DP: - (Negative) = Bad items (Be omitted)

If the result was positive, it would have positive discrimination power because upper students could answer while poor could not answer. When the result was zero then the item had no discrimination power. The worst was when the result was negative because the lower students could answer more many than upper students.

5. **Administering pretest**

The test aim was to know the input or the state of students' ability in listening comprehension before they were given the treatment. The test was used by the researcher was multiple choice questions with four alternative answers for each question. One was the key answer and the last three were distracters.

6. **Giving treatment**

There was three times treatments in this research. The short story was used as the media in teaching listening to the students.

7. **Administering post test**

The next step administerd the post test to the experimental class. The type of the test was similar to the pretest. The urgency of giving the test was to find

out whether there was any significant increase of the students' listening comprehension achievement.

8. **Analyzing result of both pretest and post test**

The next step of the research analyzed the data. The way of analyzing them was by using matched t-test formula in order to compare the two means of the same students. In analyzing them, the researcher used the sevententh version of SPSS (Statistical Package for Social Science). It was a computer program to analyze the data.

9. **Testing the hypotheses**

The last step was to find out whether there is any significant increase of the students' achievement in listening comprehension by using top down strategy in short narrative story. There are two hypotheses; (H_0) zero hypotheses and (H_1) progressive hypotheses.

H_0 : There is no significant increase of students' listening comprehension achievement after thought using top down strategy in short narrative story at the second grade of MA Daarul Ma'arif Natar

H_1 : There is significant increase of students' listening comprehension achievement after thought using top down strategy in short narrative story at the second grade of MA Daarul Ma'arif Natar

The criteria of those hypotheses were:

H$_0$: The hypotheses is accepted if the significance level or *p alpha level at>* 0.05 or in SPSS *p alpha level p>* 0.01

H$_1$: The hypotheses will be analyzed by using SPSS 16, in which the H$_1$ approved if *p<* 0.05 or at *p* 0.05 or in SPSS approved if *p<* 0.01 or at *p* 0.01

3.5 Scoring System

Mikado and Matsumoto in Danahar (1994:4) suggest that in order to improve the reliability of listening comprehension test, the number of questions is improved and also the question of multiple choices introduce more. Thus, the researcher used multiple choices in order to gain the objectivity of the result, in form of thirty multiple choice and four alternative answers. One was as key answer and the other three was the distracters. In evaluating the students' listening comprehension scores, the researcher used this formula:

$$S = \frac{R}{N} \times 100\%$$

Note:

S = score of the test

R = right answer

N = number of item test

3.6 Data Analysis

In analyzing the data, the researcher followed these steps:

1. Scoring the pretest and posttest

2. Tabulating the results of pretests and posttest and calculating of both means

3. Drawing conclusion from the tabulated results of the pre-test and post-test administered, that was by statistically analyzing the data using statistical computerization, i.e., Matched T-Test of Statistical Package for Social Science (SPSS) version 17.0 for Windows to test whether the improvement gained by the students was significant or not, in which the significance was determined by $p < 0.05$. It was used as the data come from the two samples (Hatch and Farhady, 1982:111)..0