III. RESEARCH METHOD

3.1 Research Design

To conduct this research, the researcher used *Pretest Posttest Control Group Design*. This design belongs to true experimental designs. True experimental designs have three basic characteristics: (1) a control group is present, (2) The sample are randomly selected and assigned to the groups, and (3) a pretest is administered to capture the initial differences between the groups (Hatch and Farhady 1982:22).

The researcher used this design because she wants to give special treatment to the experimental class using retelling story in teaching reading comprehension. There were two classes of this experimental study; one was experimental class which got treatment through retelling story and another as a control class which got treatment through translation.

The pretest was administered first before the treatment. It was intended to measure the students' basic ability of both in order to ensure their entry point. Control class was needed for comparison purposes because it lets the writer interpret her findings more confidently. Both of them got the same materials.

Based on Hatch and Farhady (1982: 22), the researcher used the following design:

G1 (Random) : T1 X1 T2

G2 (Random) : T1 X2 T2

Notes:

G_1 = experimental Group

G_2 = control Group

T_1 = the pretest

T_2 = the posttest

X_1 = treatment by the researcher (Teaching reading through retelling story technique)

X_2 = treatment by the teacher (Teaching reading through translation technique)

3.2 Population and Sample

The population of the research was the first year students of SMAN 1 Gunung Sugih. The researcher was chose the first year students in the second semester of academic year 2011/2012. There were four classes of the first year students, that was XA, XB, XC, XD and each class consisted of 32 students. Their ages range from 15-16 years old.

The class as the sample was taken through lottery, because all the classes have the same opportunities to be chosen as the sample of this research. One was the experimental class I, and the other one was the experimental class II. In this case, the researcher asked the leader of each class to take a small piece of paper in order to know which the class would be as experimental class I or experimental class II.

3.3 Data Colleting Technique

In collecting the data, the researcher used reading test as the instrument. There were two kinds of test, pre-test and post-test. Pre-test administered in order to measure the student's reading comprehension achievement before the treatment,

and post-test administered after presenting the treatment in order to know the achievement of reading comprehension. The test designed based on School Based curriculum for the ten grade students.

3.4 Try-Out

The try-out was administered to determine the quality of the test that was used in taking the data. Before conducting the pre-test and post-test, a try out test was carried out. This test was administered in order to determine the quality of the test as instrument of the research.

3.4.1 The Validity

Validity is the extent to which a test measures what it claims to measure. It is vital for a test to be valid in order for the results to be accurately applied and interpreted. Validity isn't determined by a single statistic, but by a body of research that demonstrates the relationship between the test and the behavior it is intended to measure. A test can be said valid if it measures the object to be measured and suitable for the criteria (Hatch and Farhady, 1982: 251).

There are four basic types of validity: content validity, criterion-related validity, face validity, and construct validity, (Hatch and Farhady, 1982:251). To determine the validity of the test, the researcher emphasizes only on content 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 on adequacy of the sample and not simply on the appearance of the text. To assure the researcher of

content validity of a test, the content of whatever the test will measure must be carefully defined.

Table 3.1. Tabel of specification of try out test

No	Objectives	Item Numbers	Total Items	Percentage
1	Identify the main idea	1, 9, 15, 19, 26, 27	6	18%
2	Vocabulary	6, 7, 17, 18, 24, 25, 33, 35	8	22%
3	Specific information	4, 10, 12, 13, 14, 21, 23, 28, 30, 32	10	28%
4	Inference	2, 3, 11, 20, 22, 29, 31	7	20%
5	Reference	5, 8, 16, 34	4	12%
TOTAL			35	100%

3.4.2 The Reliability

Reliability refers to the consistency of a measure. A test was considered reliable if we got the same result repeatedly. For example, if a test was designed to measure a trait (such as introversion), then each time the test was administered to a subject, the results should be approximately the same. Unfortunately, it was impossible to calculate reliability exactly, but it can be estimated in a number of different ways.

According to Hatch and Farhady (1982:243), the reliability of a test can be defined as the extent to which a test procedures consistent result when administered under similar conditions. To estimate the reliability of the test, the split-half method is used in order to analyze the odd (x) and even (y) of the test items. To measure the coefficient of the reliability between odd and even group, the researcher will use Spearman-Brown formula, that is:

$$R_{xy} = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{(N\Sigma X^2 - (\Sigma X)^2)(N\Sigma Y^2 - (\Sigma Y)^2)}}$$

Notes:

R : coefficient of reliability between odd and even numbers
N : number of the students
X : square of x
Y : square of y
∑X : total score of odd number
∑Y : total score of even number
∑Y : total score of even number
(Hatch and Farhady, 1982:198)

The criteria are:
0.80 up to 1.00 is very high.
0.60 up to 0.79 is high.
0.40 up to 0.59 is average.
0.20 up to 0.39 is low.
0.0 p to 0.19 is very low.

3.4.3 Level of Difficulty

The difficulty level of an item shows how easy or difficult that particular item done by the participants (Heaton,1975:182). Level of difficulty was generally expressed as the fraction (or percentage) of the students who answered the item correctly. It was calculated by the following formula:

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

Notes:
LD : the level of difficulty
R : the number of the students who answer correctly
N : the total of the students in the higher and lower group
(Heaton,1975:182)

The criteria of the difficulty level are
< 0.30 = difficult
0.30- 0.70 = average
> 0.70 = easy

3.4.4 Discrimination Power

The discrimination power (D) is the proportion of the high group students getting the items correct minus the proportion of the low-level students who getting the items correct. The discrimination power of an indicate item the extent, to which the item discriminates between the test taker from the less able. The formula of the discrimination power is:

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

Notes:
D : discrimination power
U : the number of students from the upper who answer correctly
L : the number of students from the lower who answer correctly
N : the number of the students
(Shohamy, 1985:82)

The criteria of discrimination power are:

1. If the value is positive, it has positive discrimination because large number or more knowledge students than poor students get the item correct. If the value is zero, it means that there is no discrimination.

2. If the value is negative, it has negative discrimination power because lower and higher level of the students gets the item correct.

3. In general, the higher discrimination index is better. In the classroom situation most items should be higher than 0.20 index.
(Shohamy, 1985:82)

Ha = L-ratio is higher than L-table (the distribution of the data is not normal)

From the calculation, the distribution of the data is normal (L-ratio is smaller than L-table)

3.5 Procedures of Taking the Data

There were some procedures that were applied for taking the data, they were:

1. Determining the population and the sample.

The researcher took two classes to determine the experimental class I and experimental class II.

2. Administering try-out.

   The try out administered to determine the quality of the test.

3. Administering pre-test.

   The researcher and the teacher administer the pre-test on both groups experimental class I and experimental class II.

4. Conducting treatment.

5. Administering the post-test.

   The researcher and the teacher administered the post-test, experimental class I and experimental class II.

6. Scoring the student's work.

   The researcher scored the learner's work in order to get the data.

7. Analyzing the data.

   After collecting the data, the researcher analized the data.

8. Testing hypothesis.

   After analyzing the data, the researcher tested the hypothesis.

3.6 Scoring System

In scoring the result of students' test, the researcher used Percentage Correct (Lyman, 1971:95). The score is 1 for each correct number. In order to make scoring easier, each correct answer divided by total number of the test and multiplied by 100 to make maximum score 100. The percentage correct score used in reporting the result of classroom achievement tests.

The researcher calculated the average of the pre-test and post test by used this formula:

$$X_{\%c} = 100\frac{R}{T}$$

(Lyman, 1971: 95)

Where:

$X_{\%c}$ = percentage of correct score

R = number of right answers

T = total number of items on test

3.7 Data Analysis

The writer computes the students using the following steps:

- Scoring the pre-test and post-test.
- Tabulating the results of the test and calculating the score of the pre-test and post-test.
- 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. *Independent Groups T-Test of Statistical Package for Social Science (SPSS) version 20.0 for windows* to test whether the increase of students' gain is significant or not, in which the significance was determined by $p < 0.05$. It is used as the data come from the two samples (Hatch and Farhady, 1982:111).

3.8 Treatment of the Data

In order to determine whether the data were good or not, the researcher analized the data by:

1. Scoring the pre-test and post-test
2. Tabulating the result of the thesis and calculating the mean of the pretest and posttest. To compute the average score or mean of the pretest and posttest, the researcher used a very simple statistic formula as follows:

$$\bar{X} = \frac{\Sigma x}{N}$$

Notes:

$\bar{X}$: mean (average score)
∑x : total number of the student's score
N : total number of the students
(Hatch and Farhady, 1982:5)

3. Calculating from the tabulated results of the pretest and posttest administered, that was by statistically analized the data using ***t-test*** to test whether or not the difference between pretest and posttest was significant. It was used as the data comes from the same sample or known as paired data (Hatch and Farhady, 1982).
4. Administering the Normality Test

   This test was used to measure whether the data in two classes were normally distributed or not. The data are tested by One-sample Kolmogorov-Smirnov Formula (SPSS 20).

   The criteria of normal distribution are:

The hypothesis is accepted if the result of the normality test is higher than 0.05 (sign > α). In this case, the researcher used level of significance of 0.05

5. Administering the Homogeneity Test

   This test was used to know whether the data of the posttest from the experimental class 1 and from the experimental class 2 were homogeneous or not. The data tested by Independent Sample Test (SPSS 20). The criteria for the homogeneity of pre test are:

   H_1: There is no significant difference in the level of ability (equal)

   H_0: There is a significant difference in the level of ability (not equal)

   The criteria for the hypothesis is: H_1 is accepted if the result of Homogeneity test of pre test is higher than 0.05 (Sign > α).

3.9 Hypothesis Testing

After collecting the data, the writer recorded and analyzed them in order to find out whether there is an increasing in students' ability in reading comprehension of folktale or not after the treatment. The writer used Independent Group T-test to know the level of significance of the treatment effect.

The formulation is:

$$t_{obs} = \frac{\overline{X}_e - \overline{X}_c}{S_{(\overline{X}_e - \overline{X}_c)}}$$

With:

$$S_{(\overline{X}_e - \overline{X}_c)} = \sqrt{\left(\frac{S_e}{\sqrt{n_1}}\right)^2 + \left(\frac{S_c}{\sqrt{n_2}}\right)^2}$$

$\overline{X}_e$: Mean from the difference pre-test and post-test of experimental class and control class

$\overline{X}_c$: Mean from the difference pre-test and post-test of experimental class and control class

$S_{(\overline{X}_e - \overline{X}_c)}$: Standard error of differences between means

n : Subjects on sample

(Hatch and Farhady, 1982:111)

The criteria are:

If the t-ratio is higher than t-table : H_1 is accepted

If the t-ratio is lower than t-table : H_0 is accepted