

III. METHODS

This chapter discusses the research method used in this study. The discussion includes design, data, data resource, instrument, procedure, data analysis, and hypothesis testing.

3.1. Design

This research was a quantitative in nature. The researcher used one group pretest-posttest design (Hatch and Farhady, 1982: 20). One group pretest-posttest design was a research design where one group of participants was pre-tested on the dependent variable and then post-tested after the treatment condition has been administered. The pre-test was conducted to measure students' reading comprehension achievement before treatment and the post-test was conducted to find the students' reading comprehension achievement after being taught using PQ4R strategy. Then, the means score of both pre-test and post-test was compared to find out the progress before and after the treatment.

The researcher used one class as the experimental class and another class as the try out class. The researcher used simple random probability sampling to determine one experimental class and one try out class.

The research design is represented as follows:

T1 X T2

Where :

T1 Pre-test

T2 Post-test

X Treatment

(Hatch and Farhady (1982:24))

This research was conducted in nine meetings. The first meeting was for try-out test and the second meeting was for pre-test. The third until eighth meetings were for treatment by using PQ4R strategy. After that, the ninth meeting was for the post-test. If the average score of pretest (mean) was higher than the average score (mean) of the posttest, it indicated that PQ4R strategy did not affect the students' reading comprehension achievement. However, if the average score (mean) of the posttest was higher than the average score of pretest (mean), it showed that PQ4R strategy affected the students' reading comprehension achievement because it could increase the students' reading comprehension achievement.

3.2. Population and Sample

The population of this research was the second grade students of SMPN 1 Sukoharjo in 2014/2015 academic year. There were seven classes at the second grade students in SMPN 1 Sukoharjo. Each class consisted about 32-36 students. This research employed two classes, the first class was VIII A as the try out class and the second class was VIII B as the experimental class. Those classes were selected randomly because there was no priority class of the second grade in SMPN I Sukoharjo. It was

applied based on the consideration that every class in the population has the same chance to be chosen and in order to avoid the subjectivity in the research (Setiyadi, 2006: 39).

3.3. Variables

In order to assess the influence of the treatment in this research, variable can be defined as dependent and independent variables. Hatch and Farhady (1982: 15) state that the independent variable is the major variable that a researcher hopes to investigate and the dependent variable is the variable that the researcher observed and measured to determine the improvement of the independent variable.

The research consists of the following variables:

1. Preview, Question, Read, Reflect, Recite, Review (PQ4R) strategy as independent variable (X).
2. Reading Comprehension Achievement as dependent variable (Y).

3.4. Instrument

The instrument of this research was reading tests. The reading tests were focused on examining students reading achievement that was using Preview, Question, Read, Reflect, Recite, Review (PQ4R) strategy. The researcher administered a pre-test, and post-test. Then, the data were analyzed from the result of those two activities which can be clarified as follows:

1. Pre-test

Pre-test was conducted in order to find out the students' reading comprehension achievement before the treatment. This test was multiple choice in which the students were asked to choose one correct answer from the options a, b, c, or d. In this test, the students were given 30 items were taken from the result of try-out test and it was conducted within 60 minutes.

2. Post-test

After conducting the treatment, the posttest was administered. It was done in order to know the students' achievement after the treatment. This test consisted of 30 items of multiple choices and would be done within 60 minutes. It can be stated that when the students get higher score in post-test, it means that they have learnt certain topics well.

If the scores were same, or if the post-test score was lower than the pre-test score, it can be inferred that the topics were not learned well in the teaching and learning process.

3.5. Data Collecting Technique

In collecting the data, the researcher used the procedure that can be described as follows:

1. Planning

The procedure of planning this research can be described as follows:

a. Preparing the Try-out Test

In the first meeting, the researcher gave a test in try out class. In try out class, the researcher provided 40 items in 90 minutes. It was done in order to know whether the test items were applicable or not, by finding out the validity, reliability, level of difficulty, and discrimination power. Split-half method was used to measure the reliability in which requires the researcher to provide the items into two same groups, first half and second half.

b. Preparing the Pre-Test

In the second meeting, the researcher administered the pre-test in experimental class. This test was administered to find out the students' basic reading comprehension before treatment. It used an objective test in form of multiple choices with 30 items in 60 minutes. 30 items were taken from the result of try-out test.

c. Preparing the treatment

After giving the pre-test for the students, the researcher conducted treatment for six times by using PQ4R strategy. The materials of this research were based on the School Based Curriculum 2006 for second grade student, that is recount text. There were three lesson plans in the process of teaching reading, which involved recount text inside. Hopefully, those lesson plans in teaching reading process were able to generate a good reading comprehension.

d. Preparing the Post-Test

In the last meeting, the post-test was distributed in the experimental class to determine the result of students' reading comprehension after being taught by using PQ4R strategy. The researcher used an objective test in form of multiple choice items in 60 minutes. It consisted of 30 items to find out whether there was a significant increase on the students' reading comprehension achievement after the treatments.

2. Application

After making a plan, there were some steps applied as follows:

a. In the first meeting, the try out test was given.

The test papers were distributed to the students and the students were asked to do the test. And the last, they were asked to hand in their test. This test was in the form of multiple choices that consisted of 40 items and was allocated within 90 minutes.

b. In the second meeting, the Pre-Test was given.

In this test, the researcher distributed multiple choices test that consisted of 30 items of recount text in 40 minutes.

c. Conducting treatment

After giving the pre-test, the treatment was conducted in sixth meetings with 2 x 45 minutes in each meeting. The students were given different assignments for each session and the materials were about recount text. And then, there were five recount texts taught to the students.

d. In the last meeting, the Post-Test was given.

The post-test was in the form of multiple choices which consisted of 30 items of recount text in 60 minutes.

3.6. Try Out of the Instruments

In doing this research and proving whether the test items were applicable or not, the researcher tried out the test to find out validity, reliability, level of difficulty, and discrimination power of the test. It was conducted in order to determine whether 40 items had a good quality or not before being given for the pre-test and the post-test. There are four criteria of a good test that should be met: validity, reliability, level of difficulty, and discrimination power.

3.6.1. Validity of the Test

Validity refers to the extent to which the test measures and to what it was intended to measure. A test is considered valid if the test measures the object to be measured and suitable with the criteria (Hatch and Farhady, 1982: 250). A test must aim to provide true measure of the particular skill which it is intended to measure.

There are four kinds of validity:

1. Face validity, concerns with the layout of test.
2. Content validity, depends on a careful analysis of the language being stated.
3. Construct validity, measure certain specific characteristic in accordance with a theory of a language.

4. Criterion-related validity, concerns with measuring the success in the future as in replacement test.

According to the types of validity above, the researcher used content and constructs validity. Both of them are explained as follows:

a. Content Validity

Content validity was intended to know whether the test items were good reflection of what will be covered or not. The test items which are adapted from the materials that have been taught to the students should be constructed as to contain a representative sample of the course (Heaton, 1988). The following ways were used to prove whether the test had a good content validity: 1) It was adopted from Educational goal stated on 2006 English curriculum and the syllabus for second year of SMP students, 2) It represented the material taught in the class. The material of the test was also taken from the subject matter content that had been taught in the class.

b. Construct Validity

Construct validity investigates the research instrument appropriateness to the research object. Related to this research, the test items should be questioning the five aspects of reading such as main idea, specific information, reference, inference, and vocabulary. As Nuttal (1985) states that reading have five sorts reading skills that should be mastered by the reader to comprehend the text deeply. The researcher made the same percentage of all aspects to know the basic knowledge about five aspects of

reading when they answered the questions in the try out test. The researcher provided the table of specification for try out test as follows:

Table 3.1 Reading Specification of Try out Test

No.	Reading Skills	Items Numbers	Percentage of items
1.	Determining main idea	1,6,14,21,24,35,38,33	20%
2.	Identifying details	2,7,10,15,27,31,32,39	20%
3.	Determining inference	3,12,16,19,23,29,30,36	20%
4.	Reference	4,8,17,22,25,28,34,37	20%
5.	Understanding vocabulary	5,9,11,13,18,20,26,40	20%
Total		40 items	100%

Nuttal (1985)

3.6.2. Reliability of the Test

Reliability refers to the extent to which the test is consistent in its score and gives us an indication of how accurate the test score are (Hatach and Frhady, 1982: 224). To test the reliability of the instruments, the writer used *split-half method* in which the reading tests were divided into halves. By splitting the test into two equal parts (first half and second half); it was made as if the whole tests have been taken twice. The test is determined by using Pearson Product Moment which measures the correlation coefficient of the reliability between odd and even number (reliability of half test) in the following formula:

$$r_{xy} = \frac{\sum xy}{\sqrt{(\sum x^2)(\sum y^2)}}$$

Where:

r_{xy} : coefficient of reliability between odd and even numbers item

x : odd number

y : even number

$\sum x^2$: total score of odd number items
 $\sum y^2$: total score of even number items
 $\sum xy$: total score of odd and even number

After getting the reliability of half test, the researcher used Spearman Bowns Prophecy formula (Hatch and Farhady, 1982: 247) to determine the reliability of the whole tests, as follows:

$$r_k = \frac{2 r_{xy}}{1 + r_{xy}}$$

Where :

r_k : the reliability of the whole tests

r_{xy} : the reliability of half tests

The criteria of reliability as follows:

0.90 – 1.00 = high
 0.50 – 0.89 = moderate
 0.00– 0.49 = low

(Hatch and Farhady, 1982: 127)

The result of the reliability in the try out test can be seen in Appendix 12. It was 0.97.

The criteria was, 0.90-1.00 = high, 0.50-0.89 = moderate, 0.00-0.49 = low. So, it can be stated that the reliability of the test was high.

3.6.3. Level of Difficulty

Level of difficulty relates to how easy or difficult the item taken from the point of view of the students who take the test. It was important since test items which were too easy (that all students get right) can tell us nothing about differences within the test population (Shohamy, 1985: 79).

Moreover, the difficulty level of an item shows how easy or difficult that particular item done by the participants (Heaton, 1975: 182).

It is calculated by the following formula:

$$LD = \frac{U+L}{N}$$

Where:

LD : level of difficulty

U : the number of upper group who answer correctly

L : the number of lower group who answer correctly

N : the total number of students in upper and lower groups

The criteria are as follows:

<0.30 : difficult

0.30 – 0.70 : average

> 0.70 : easy

(Shohamy, 1985: 79)

Based on the criteria above, there were 11 easy items in the try out test (1, 2, 3, 4, 7, 8, 9, 10, 13, 18, and 23). There were 23 average items (5, 6, 11, 14, 16, 17, 19, 20, 21, 22, 24, 26, 27, 29, 30, 31, 33, 34, 35, 36, 37, 38, and 39). And, there were 6 easy items (12, 15, 25, 28, 32, and 40). The item which were difficult and had negative discrimination power were omitted, the average and easy items which had satisfactory level and good discrimination index were used in the pre-test and the post-test (see Appendix 11)

3.5.4. Discrimination Power

Discrimination Power refers to the extent to which the items are able to differentiate between high and low level students on that test. Discrimination power is used to differentiate between the students who have high ability and those who have low ability. The discrimination power is calculated by this following formula:

$$DP = \frac{U-L}{\frac{1}{2}N}$$

Where :

DP : 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

The criteria are:

DP: 0.00 - 0.19 = Poor items

DP: 0.20 - 0.39 = Satisfactory items

DP: 0.40 - 0.69 = Good items

DP: 0.70 - 1.00 = Excellent items

DP: - (Negative) = Bad items, should be omitted

(Heaton, 1975: 180)

Based on criteria above, there were 14 items in the try out test which did not fulfill the standard of discrimination power, since those items had discrimination index under 0.19 which meant that the items had bad and poor discrimination power and it was also found that the discrimination power of 25 items were between 0.20-0.39, belonging satisfied items, and 1 item was between 0.40-0.69, belonging good items.

3.7. Scoring System

The scoring system that is used in this research is dividing the right answer by total items timed 100. In scoring the students' result of the pre-test and the post-test, the formula by Arikunto (1997: 212) is employed:

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

Were :

S : score of the test
 R : number of right answer
 N : total number of items on test

3.8. Data Analysis

In order to know the students' progress in comprehending the text, the students' score are computed by doing three activities:

1. Scoring the pre-test and the post-test.
2. Tabulating the result of the test and calculating the mean of the pre-test and the post-test. The mean is calculated by applying the following formula:

$$M = \frac{\sum X}{N}$$

Notes:

M = mean (average score)
 $\sum x$ = the total students' score
 N = total number of students

(Hatch and Farhady: 1982)

3. Drawing conclusion from the tabulated results of the test given, that is by statistically analyzing the data using statistical computerization i.e paired T-Test of statistical Package for Social Science (SPSS) to test whether the increase of students gain is significant or not, in which the significance is determined by $p < 0.05$. It is used as the data from one sample (Hatch and Farhady, 1982: 117). In order to know whether the students get any progress, the formula was as follows:

$$I = X2 - X1$$

Notes

I = the increase of students reading comprehension achievement

X2 = the average score of post-test

X1 = the average score of pr-test

3.9. Hypothesis Testing

The result of the pretest and the posttest was compared in order to know the gain. The researcher used Paired Sample T-Test towards the average score of pretest and posttest. Moreover, the result of t-test was used to determine the difference on the students' reading comprehension achievement before and after the treatment and to prove whether the proposed hypothesis was accepted or rejected. In this case, the researcher used significant level of 0.05 in which the probability of error in the hypothesis is only about 5%.

The hypotheses are drawn as follows:

1. H_0 There is no difference of the students' reading comprehension achievement before and after being taught by using Preview, Question, Read, Reflect, Recite, and Review (PQ4R) strategy.
 H_1 There is a difference of the students' reading comprehension achievement before and after being taught by using Preview, Question, Read, Reflect, Recite, and Review (PQ4R) strategy.
2. H_0 The aspect of reading comprehension which increased the most was not understanding vocabulary.
 H_1 The aspect of reading comprehension which increased the most was understanding vocabulary.

The criteria for accepting the hypothesis are as follows:

H_0 is accepted if the t-value is lower than T-ratio.

H_1 is accepted if the t-value is higher than T-ratio.