

III. RESEARCH METHOD

This research was quantitative research. And in conducting the research, the researcher used quasi experimental method. This experimental method dealt with only one group, the experimental class. The data that was obtained was number (score) that were assessed and analyzed by using statistical analysis.

3.1 Research Design

In conducting the research, the researcher used quasi experimental method. This experimental method dealt with only one group, the experimental class. The experimental class was the class which got treatments through Words Categorization Method that was presented by the researcher. The class received pre test, treatments which were done three times, and post test. The result of pre test and post test was compared to know the significant.

The design was presented as follows:

G: T1 X T2

G : Experimental class

T1 : Pre-test

T2 : Post-test

X : Treatment (Setiyadi, 2006:132)

3.2 Research Population and Sample

The population of this research was the eighth grade students of SMPN I Seputih Surabaya. Because it was impossible to use the whole class as the experimental class, therefore the researcher chose one class as the experimental class. The selection of the research sample used simple random sampling. After using dice to determine the research population, the researcher got 8E as the experimental class and 8D as the tryout class.

3.3 Data Collecting Technique

Several techniques were employed to collect the data of this study, these included: administering the pretest and posttest. Through these techniques, necessary information about the use of word categorization method was expected to be obtained. To know the validity, reliability, discrimination power and difficulty level, the researcher used try out test. The analysis of those four aspects was used in making the items for pretest and posttest. The next step was comparing means of pre test and post test to see the significance.

3.3.1 Research Instrument

The Instrument was a tool used by the writer in the research. In this research, the researcher used a test to get the data about students' vocabulary mastery after being given the treatment. The test was about vocabulary. The vocabularies that were used in the test related to the following themes: health, travelling, flora and fauna. The number of items for pretest and posttest was just the same, that was 30 items.

3.3.1.1 Pre – Test

Pre-test was conducted before presenting the treatments to know how far the students' vocabulary mastery. The researcher used multiple choices because by using this test, the target vocabulary was easier to be measured. The total number of the test was thirty items. The items were based on the materials that would be taught in the class. The materials were about flora and fauna, travelling, season and health. The questions that were given were in the kinds of multiple choices. The answer was the vocabulary that related with the material. For the "health" topic, there were so many kinds of vocabulary related to it, for example oculist, surgeon, veterinarian, etc. In this research, the researcher used the pre-test to get the data about vocabulary mastery in any classes of word like noun, verb, and adjectives. The example of pretest could be seen in Appendix 3.

3.3.1.2 Post – Test

After the teaching program was done for three meetings, the posttest was conducted in order to know the progress of students' vocabulary mastery after being taught by using Word Categorization Method in the treatments. The type of post test was multiple choices test which consisted of thirty items. The question for posttest was just the same as pretest. The different was just in the number of item that had been changed. The example of posttest items could be seen in Appendix 6.

3.3.2 Research Procedure

This research was based on the following procedures:

3.3.2.1 Administering The Try-Out Test

It was conducted to measure the validity, reliability, difficulty level and discrimination power of the items that would be used in pretest and posttest and to make sure whether the test was good or bad for students. The test was tried out to the students whose level is equal to the sample of the research. In this case, it had been chosen 8D. It was administered to find out the quality of the test before it was used, whether the items were good or not in validity, reliability, level of difficulty, and the discrimination power. This exam used multiple choice test consisted of 40 items to be done in 80 minutes. The maximum score was 100.

3.3.2.2 Administering The Pre-Test

Pretest was the test that was done after the tryout test. After the items in try out test were analyzed, the items would be repaired to be used in pretest. The researcher administered the pre test in order to find out the students' basic ability before treatments. The subject of pretest was the experimental class that was 8E. In this test, the researcher asked students to do the objective test that consisted of thirty items.

3.3.2.3 Conducting Treatments

Treatments were the application of the method that was used by the researcher in the class. Treatments were done after conducting pretest. It was done in order the

researcher could compare the result before and after conducting the treatments. In this case the treatments conducted in three meetings. It required eighty minutes for each meeting. In each treatment, there were different sub topics that were presented. The themes that were used are health, flora, fauna, and travelling. The selecting of the themes was adapted with the grade of the students in experimental class and the curriculum of the recent textbooks.

3.3.2.4 Administering The Post Test

The post test administered after treatments. This test was used to know whether the treatments could significantly improve students' vocabulary achievement or not. In this test, the students were asked to do the multiple choice test that consisted of thirty items. The items were not different with the items that were used in pretest item, but the researcher just changed the number of the questions.

3.3.2.5 Analyzing The Test Result (Pre Test And Post Test)

After scoring students' work finished, the researcher compared the result of pre test and post test to see whether the score of post test higher than pre test and to see the correlation of the two items.

3.4 Try Out Test

Before the instrument used in the research, the researcher administered try out test to investigate the validity, reliability, difficulty level and discrimination power of the instrument. Try out test comprised forty multiple choice questions related to the materials. The test materials were adapted from several textbooks used in the

eighth grade of junior high school. The test was administered in 8D before the pretest. The example of tryout test could be seen in Appendix 1. There were some considerations needed in order to find out the quality of the instruments such as: validity, reliability, level of difficulty, and discrimination power of the test.

3.4.1 Validity and Reliability Test

In the research, data was the most important one, because it represented the variable that was researched. It had function to prove the hypothesis. Whether the data was true or not, it influenced the quality of the research. Whereas the true or not of the data depended on the instrument of the research. To examine the instrument, the researcher used four criteria: validity, reliability, difficulty level and discrimination power. In this chapter, the researcher would explain about validity and reliability first, the last two criteria would be discussed later.

3.4.1.1 Validity Test

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 (Kendra 2010:46). It meant that the test measure what was claimed to measure. To know the validity of the test, in this research, the writer compared the r value and r table :

If r value $>$ r table, the item was valid but if r value $<$ r table, it would be not valid.

Besides that the researcher also saw the validity from content validity, construct validity and internal validity.

a. Content Validity

According to Weir (1990:22) content validity concerned with whether the test was sufficiently representative and comprehensive for the test. Content validity could be found by relating the material of the test with the curriculum for Junior High School. It meant that the test was designed based on the curriculum in the school. In this case, to know whether the instrument had fulfilled the criteria of content validity the researcher had checked in Competence-based English Developing competencies in English for grade 8 junior high school Grafindo media pratama. The researcher had also consulted the instrument test to the English teacher at the school that had been chosen as a place for research.

b. Construct Validity

Still according to Weir (1990:22), construct validity concerned with whether the test was actually in line with the theory of what it meant to know the language. Construct validity focused on the kind of the test that was used to measure the ability. It meant that the test item should really measure the student's vocabulary mastery. In this case whether to know the instrument had fulfilled the criteria of construct validity the writer consulted the instrument test to the English teacher at the school that had been chosen as a place for research. The table specification was as follows:

Table Specification.

Theme / Class of word	Health	Flora and Fauna	Travelling	Total
Noun	10	8	7	25
Verb	2	2	5	9
Adjective	1	1	4	6
Total	13	11	16	40

The researcher wanted to increase students' vocabulary achievement by using a method named word categorization method. This method was supported by a media named clipping picture puzzle. Actually, at least there are four aspects in vocabulary : verb, noun, adjective and adverb. But in this method, the researcher wanted to give more portion in noun because the media was more suitable to visualize noun rather than other classes of word. As can be seen from the table above.

3.4.1.2 Reliability Test

According to Cronbach (1971) reliability referred to whether the test was consistent in its scoring and gave us an indication of how accurate the test scores. In other words, the reliability was a test that was consistent and dependable. To obtain the reliability of the instrument, this study used SPSS to analyze the reliability of the test. The steps to run the reliability in SPSS were as follows:

Entering the data → Analyze → Scale → Reliability Analysis. The criteria of Reliability test were:

0.81 – 1.00 = Very high

0.61 – 0.80 = High

0.41 – 0.60 = Sufficient

0.21 – 0.40 = Low

0.0 – 0.20 = Very low (Gronlund, 1982:126)

3.4.2 Difficulty Level and Discrimination Power

Besides validity and reliability, the researcher tested the instrument that would be used in pretest by using difficulty level and discrimination power too. Those two would be discussed one by one.

3.4.2.1 To Find Difficulty Level

Difficulty level was used to make sure that the instrument was not too easy and difficult for the students. The formula was :

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

Where:

LD : level of difficulty

R : number of students who answer correctly

N : total number of students

The criteria are:

LD < 0,30 = difficult

LD = 0,30 – 0,70 = satisfactory

LD > 0,70 = easy (Oller, 1979)

3.4.2.2 To Find The Discrimination Power

Discrimination power would make sure that the instrument could discriminate the upper group and lower group. The formula was :

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

Where:

DP = Discrimination power

U = The number of upper groups students who answer correctly

L = The number of lower groups students who answer correctly

N = Total number of students

The criteria are:

0,0 – 0,20 = poor

0,21 – 0,40 = satisfactory

0,41 – 0,70 = good

0,71 – 1,00 = excellent (Brown,1998)

3.4.3 Result of Try Out Test

The try out test was given to know the quality of the test as the instrument of this research. The try out test consisted of 40 items in multiple choice form. The data gained from the try out test were analyzed to decide the level difficulty,

discrimination power, the reliability and validity of the test. Concerning with the validity, reliability, level of difficulty and discrimination power it was finally decided that the result showed 30 items were good and the rest (10) items were bad and should be dropped. The lowest score in the level of difficulty was 0.17 and the highest score was 0.73 while the lowest score in the discrimination power was 0.00 and the highest score was 0.53. To see the complete data of level of difficulty and discrimination power, it can be seen in Appendix 8.

To see the table of validity, it could be seen in Appendix 9. To find the validity of the test, the researcher compared the r value and r table. If r value was higher than r table, therefore the item was valid but if the r value was lower than the r table, the item was not valid. To get the r value, the researcher used Pearson Formula in Microsoft excel whether the r table could be seen in Appendix 11.

To find out the reliability of the test, the researcher used SPSS. The computation of the data showed that the reliability of the test was 0.74, which means that the reliability (r) was high since the range of this criteria was 0.61 – 0.80 (Complete table of the reliability can be seen in Appendix 9). It indicated that this instrument would produce consistent result when administered under similar conditions to the same participants, and in different time.

3.4 Data Analysis

There were some steps in analyzing the data, firstly the researcher would score the pre test and post test, then the data would see whether it was normal distributed or not, and for the last was hypothesis testing that would be done by using SPSS.

3.4.1 Scoring Technique

As mentioned earlier, the test was in the form of multiple choice questions. If it was compared with other format, the multiple choice format could be used to assess a greater variety of learning target (Nitko and brookhart,2007:152). In scoring students' result on the test, the researcher used "Arikunto's formula". The highest score would be 100.

The formula was:

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

Where:

S = The score of the test

R = the total of the right answer

N = number of items (Arikunto, 2007)

3.4.2 Normality Test

The normality test was a test that was used to know whether the data was in normal distributed or not. It was counted by using SPSS. After knowing the data was normal distributed or not, it can be continued by determining the testing procedure whether parametric or non parametric. To run normality test, the steps were as follows:

Entering the data → analyze → Descriptive Statistics → Explore.

To see the normality, it could be seen in three ways; variants coefficient, Skewness ratio, and Kurtosis Ratio.

3.4.2.1 Variants Coefficient

The data was called as normal distribution data if the variants coefficient < 30 %.

This value should be counted first with this formula:

Variants Coefficient = $\frac{\text{Std Deviation}}{\text{Mean}} \times 100 \%$

Mean

3.4.2.2 Skewness Ratio

The data was called as normal distribution data if the Skewness ratio was in the range of -2 until 2. The formula to count this was as follows :

Skewness ratio = $\frac{\text{Skewness}}{\text{Skewness Std.error}}$

Skewness Std.error

3.4.2.3 Kurtosis Ratio

It was almost the same as Skewness Ratio. The data was called as normal distribution data if the range was between -2 and 2. The formula was:

Skewness ratio = $\frac{\text{Kurtosis}}{\text{Kurtosis Std.error}}$

Kurtosis Std.error

3.4.3 Hypothesis Test

This test was used to test whether the hypothesis proposed by the writer accepted or not. To test the hypothesis, the researcher used SPSS. The steps were as follows:

Entering the data → Analyze → Compare means → Paired samples T-test.

The writer's hypotheses are:

Ho: There is no significant improvement in students' vocabulary achievement before and after being taught through word categorization method.

H₁: There is any significant improvement in students' vocabulary achievement before and after being taught through word categorization method.

While the criteria of the test were:

If the significant > 0.05 , Ho is accepted, but if the significant < 0.05 , Ho is refused.