CHAPTER III

RESEARCH METHODOLOGY

A. Research Design

In this research, the researcher used quantitative approach with correlation design. Quantitative research is officially about collecting numerical data to explain particular phenomenon. Quantitative approach research is a research method that use for study about population or sample, and use instruments to collect the data, the data analysis will analyze in statistics in order to testing the hypothesis that already existed. The quantitative research will help the researcher to describe the significant correlation between students' vocabulary mastery and their reading comprehension.

Arikunto says that "Correlation research aims to find out whether there is a relationship and if there is, how close the relationship is and whether or not the relationship is meaningful".[1] According to Creswell defines a correlation as a statistical test to determine the tendency or pattern for two (or more) variables or two sets of data to vary consistently. The researcher' aim is to find out the relationship between two variables, vocabulary mastery and reading comprehension.

[1] Suharsimi Arikunto, *Prosedur Pendidikan Penelitian Suatu Pendekatan Praktik* (Jakarta: Rineka Cipta, 2010), p.313

B. Time and Place of Research

This research was conducted on November, 16th 2020 – December, 4th 2020. The research was conducted on the eleventh grade of SMAN 2 Pandeglang Academic Year 2020/2021. There are three main reasons of selecting this setting. First, the researcher gains access to carry out research. The permission is very crucial because it eases the researcher to design research planning and conduct research. Second, the facilities and number of students are adequate to support this study. Third, the researcher found the students' problems in vocabulary mastery is very vital because without acquiring vocabulary students will not be able in reading comprehension.

C. Population and Sample

1. Population

According to Sugiyono "Population is a generalization area consisting of: objects/subjects that have certain qualities and characteristics determined by the researcher to be studied and then draw conclusions".[2]

Population in this research are the eleventh grade students of SMAN 2 Pandeglang in the academic year 2020/2021. There are twelve classes. The population of the research can be seen as follows:

[2] Sugiyono, *Metode Penelitian Pendidikan* (Bandung:Alfabeta, 2015), p.61

Table III.1

Population of the Research

No	Class	Number of Students
1	XI Science 1	36
2	XI Science 2	34
3	XI Science 3	34
4	XI Science 4	36
5	XI Science 5	35
6	XI Science 6	34
7	XI Social 1	36
8	XI Social 2	36
9	XI Social 3	29
10	XI Social 4	36
11	XI Social 5	34
12	XI Social 6	36
Population		416

2. Sample

Sugiyono states that "The sample is part of the number and characteristics of the population". The sample must be representative of the population. For sampling, it must be representative and can represent the population. Arikunto states that "if the total population is more than 100 students, the sample can be taken between 10-15% or more."

In this research to determine the number of samples, the researcher used the Taro Yamane formula as follows:

$$n = \frac{N}{N.d^2 + 1}$$

n = number of samples

N = number of population

d^2 = predefined precession (Bambang Suwarsono, 2007: 44)

By using the formula above, the number of samples used is:

N = 939 and d = 10%

$$n = \frac{N}{N.d^2 + 1} = \frac{416}{416.0{,}1^2 + 1} = 80.62015 = 80 \text{ respondents}$$

So the samples taken in this study were 90 students from the eleventh grade at SMA 2 Pandeglang.

In this research, the researcher determined the sample using random sampling technique because the researcher could not choose all of class for the sample. The researcher used lottery method to determine the sample because it was more efficient to the students. Lottery method to determine the sample since it was more effective and efficient.

The researcher wrote 12 classes in small papers then the researcher rolled small papers and inserted into a box. Finally, the writer chose rolled paper randomly and rolled paper took as the sample. It was gotten XI Science 1, Science 2 and Science 6 as a sample of this research.

D. Research Variables

This research contains two variables that assumed have a relationship. They were dependent variable and independent variable.

1. According to Sugiyono "Independent variable is variable that affect or are the cause of changes or the emergence of the dependent variable." The independent variable in this research is vocabulary mastery.
2. Sugiyono says that "Dependent variable is a variable that is influenced or which is the result, because of the independent variable." The dependent variable in this research is reading comprehension.[3]

E. The Technique of Data Collection

The researcher used test as the techniques to collect the data for this research. This test is used to determine or measure students' vocabulary mastery and reading comprehension. Vocabulary mastery and reading comprehension instruments were measured by a multiple choice test. This test is conducted and distributed online using Google Forms.

F. Research Instruments

The instrument used in this study was a test with a list of questions as a tool to determine respondent data. The form of the test items is arranged in

[3] Sugiyono, *Metode Penelitian Pendidikan* (Bandung:Alfabeta, 2015), p.61

questions to be answered by the respondent with a choice of answers according to the condition of each student.

Previously, the instruments that had been prepared were tested first to calculate their validity and reliability. The validity of this instrument is the validity of content or content validity. Which aims to ensure the quality of these instruments.

1. Vocabulary Mastery Instrument

Vocabulary test is used to measure the students' vocabulary mastery. This test takes the form of multiple choices with absolute response provisions, giving the right answer a score of 1 and the wrong answer a score of 0. Because there are 30 items, the minimum score is 0 and the maximum score is 30. As for the answer choices, the respondent is presented with 5 answer choices, where 1 is the correct answer, while 4 others are wrong answers. This test is used because it is more practical both in implementation and examination, and more objective in the assessment system.

a. Vocabulary Mastery Lattice

The instrument lattice for measuring vocabulary mastery is presented in this section consisting of two instrument lattices, which are tested and the final instrument lattice used to measure vocabulary

mastery variables. The lattice of the instrument is presented in table below:

Table III.2

Blue Print of Vocabulary Mastery

Indicator	Item number	Total
Word classes a. Noun b. Verb c. Adjective d. Adverb	 22, 23, 24, 25, 26, 27. 19, 20, 21. 14, 15, 16, 17, 18. 28, 29, 30.	17
Word meaning a. Synonym b. Antonym c. Hyponym	 1, 2, 3, 4, 5. 6, 7, 8, 9, 10. 11,12, 13.	13
Total		30

b. Calibration

To test the validity and reliability of the instrument from the vocabulary mastery variable, an instrument trial or instrument calibration was carried out on 30 students who were not included as the research sample. Research instrument calibration includes:

1. Validity

Validity is the standard which used to measure the appropriateness of an instrument. To know whether the data are valid or not, the researcher used content validity. Arikunto (2001: 166) states that the validity test knows a description of the

accuracy of the measuring instrument used and the ability of the measuring instrument to be measured.

The scores for the answers of each instrument are presented according to an interval scale so that the validity of the research instrument is tested through the Product Moment formula, as follow:

$$r_{yx} = \frac{n\sum x.y - \sum x.\sum y}{\sqrt{n\sum x^2 - (\sum x)^2}\sqrt{n\sum y^2 - (\sum y)^2}}$$

r_{yx} = correlation coefficient of data x to data y.

x = score of a specific item for each student.

y = total question score for each student.

n = number of trial samples.

To calculate the validity of the vocabulary mastery question items using the product moment person correlation formula, where the criteria for receiving instrument items are valid or not used the instrument validity test with r table determined by one-sided test with a significant level (α) = 0.05 and the degree of confidence (df) = k-2 (k = the number of respondents in the trial).

The testing criterion is if r count> r table, then the item is valid. After calculating the validity, the items are said to be valid

if the value of r count is greater than the value of r table (r count> r table) for a significant level α = 5% and n = number of sample members.

Based on the test results of the vocabulary mastery instrument for 30 eleventh grade students, the results were:

Table III.3

Results of the Calculation of Vocabulary Mastery Validity Test

number	**r_{count}**	**r_{table}**	**status**
1	0,568	0,361	valid
2	0,514	0,361	valid
3	0,558	0,361	valid
4	0,56	0,361	valid
5	**0,261**	0,361	**drop**
6	0,467	0,361	valid
7	0,376	0,361	valid
8	0,458	0,361	valid
9	0,611	0,361	valid
10	0,545	0,361	valid

number	r_{count}	r_{table}	status
11	0,833	0,361	valid
12	0,761	0,361	valid
13	0,412	0,361	valid
14	0,836	0,361	valid
15	0,546	0,361	valid
16	0,514	0,361	Valid
17	0,831	0,361	Valid
18	0,485	0,361	Valid
19	0,535	0,361	Valid
20	0,731	0,361	Valid
21	0,63	0,361	Valid
22	0,577	0,361	Valid
23	0,555	0,361	Valid
24	0,593	0,361	Valid
25	0,665	0,361	Valid
26	0,544	0,361	Valid
27	0,395	0,361	Valid
28	0,452	0,361	Valid
29	0,527	0,361	Valid
30	0,363	0,361	Valid

Based on the results of the instrument test, it can be concluded that there are 29 valid vocabulary mastery variable questions because it has a value of r count> r table, and there is 1 question that was dropped because it has a value of r count <r table, question number 5. So it can be concluded that the valid questions on the vocabulary mastery variable used in the research amounted to 29 questions.

2. Reliability

In addition to calculating the validity of the instrument, the reliability is also calculated. An instrument is said to be reliable if the instrument can be trusted as a measuring device. Thus an instrument is said to be reliable if the results of the instrument show accuracy.

The reliability test of the vocabulary mastery instrument was carried out for all test items using the Alpha Croanbach formula (Riduwan, 2004: 125):

$$r_{AC} = \frac{k}{k-1}[1 - \frac{\Sigma(S_i)^2}{(S_x)^2}]$$

k = number of valid items

S_i^2 = variant score of instrument items

S_x^2 = total score variant

The amount of the variance of a research sample can be formulated:

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$$

s^2 = variant

s = standard deviation

xi = the value of x to i

$\bar{x}$ = average

n = the number of samples

The higher the reliability coefficient of a test, the smaller the probability of measurement error that occurs. According to Sugiyono, a scale can be said to be reliable, if the r_{AC} coefficient is more than> 0.60, then the data is said to be reliable.

As a criterion for determining the level of reliability, the following classifications are used:

Table III.4

The Level of Acceptable Reliabillity

Criteria	Reliability Instrument
0.91 – 1.00	Level of reliability is very high
0.71 – 0.90	High level of reliability
0.41 – 0.70	Medium level of reliability
0.21 – 0.40	Low level of reliability
< 0.20	Level of reliability is very low

Based on the test results of the vocabulary mastery instrument which was calculated using the Alpha Chronbach reliability coefficient, it was found that the reliability of the vocabulary mastery instrument was 0.924 which means it has a very high degree of reliability.

Table III.5
Reliability Test Results

Cronbach's Alpha	N of Items
0.924	30

3. The Difficulty Level of the Question Items

The level of difficulty index or Proportional Correct is denoted by p. The formula is:

$$p = \frac{JB}{N}$$

JB = the number of students who answered correctly

N = the number of students who took the test

The difficulty index of the questions ranges from 0 to 1, meaning that p = 0 means that none of the respondents can answer the questions correctly, on the other hand, if p = 1, all respondents will answer the questions correctly. The criteria for

the level of difficulty used in this analysis were: if $p < 0.70$ for easy category, $0.30 < p < 0.70$ for medium category, and $p < 0.30$ for hard category.

Based on the results of the vocabulary mastery instrument trial for 30 students, the results of the calculation of the difficulty level of the vocabulary mastery instrument were as follows:

Table III.6
The Difficulty Level of Vocabulary Mastery

Number	r_{count}	Status
1	0,6	Medium
2	0,7333	Easy
3	0,7	Medium
4	0,5667	Medium
5	0,6	Medium
6	0,6333	Medium
7	0,7	Medium
8	0,8667	Easy
9	0,2667	Hard
10	0,6333	Medium
11	0,6	Medium
12	0,4667	Medium
13	0,5333	Medium
14	0,5	Medium
15	0,3	Medium

Number	r_{count}	Status
16	0,7333	Easy
17	0,5333	Medium
18	0,5	Medium
19	0,5667	Medium
20	0,5333	Medium
21	0,3333	Medium
22	0,6	Medium
23	0,5333	Medium
24	0,4667	Medium
25	0,3333	Medium
26	0,5	Medium
27	0,8	Easy
28	0,4333	Medium
29	0,5	Medium
30	0,7333	Easy

Based on these results, it can be concluded that the vocabulary mastery instrument has a hard level with 1 questions (3,33%) in the hard category, 24 (80%) in the medium category, and 5 in the easy category (16,67%).

4. Discriminating Power of Question Items

Discriminating power of question is the ability of the questions to discriminating between students who have mastered and those who have not mastered the material asked by the questions. To calculate the discriminating power of the question, the formula is used:

$$D = P_A - P_B, \qquad \text{with} \qquad P_A = \frac{B_A}{J_A} \text{ dan } P_B = \frac{B_B}{J_B}$$

DP = Index of discriminating power.

JA = The number of participants in the upper group test.

JB = The number of participants in the lower group.

BA = The number of participants in the upper group who answered correctly.

BB = The number of participants in the lower group who answered correctly.

PA = The proportion of participants who answered correctly.

PB = The proportion of the lower group of participants who answered correctly.

Based on the test results of the vocabulary mastery instrument, the calculation of the discriminating power of the instrument is as follows:

Table III. 7

Number	**Discriminating power**	**Status**
1	0,5333	Good
2	0,5333	Good
3	0,4667	Good
4	0,4667	Good
5	0,2667	Sufficient
6	0,4667	Good
7	0,3333	Sufficient
8	0,2667	Sufficient
9	0,5333	Good
10	0,4667	Good
11	0,8	Very good
12	0,8	Very good
13	0,4	Sufficient
14	0,8667	Very good

Number	Discriminating power	Status
15	0,4667	Good
16	0,5333	Good
17	0,8	Very good
18	0,3333	Sufficient
19	0,4667	Good
20	0,8	Very good
21	0,5333	Good
22	0,5333	Good
23	0,5333	Good
24	0,5333	Good
25	0,5333	Good
26	0,4667	Good
27	0,2667	Sufficient
28	0,4667	Good
29	0,4667	Good
30	0,2667	Sufficient

Based on these results it can be concluded that the vocabulary mastery instrument has discriminating power with very good categories of 5 questions (16,67%), good categories 18 questions (60%), and sufficient categories 7 questions (23.33%).

2. Reading Comprehension Instrument

Reading Comprehension test takes the form of multiple choices with absolute response provisions, giving the right answer a score of 1 and the wrong answer a score of 0.

There are 5 types of text in the questions, narrative text, recount text, description text, text analytical exposition, and explanatory text.

Because there are 30 items, the minimum score is 0 and the maximum score is 30. As for the answer choices, the respondent is presented with 5 answer choices, where 1 is the correct answer, while 4 others are wrong answers.

a. Reading Comprehension Lattice

Table III.8

Indicators	Item number	Total
Find the main idea	33, 40, 45, 47, 55, 57	6
Finding factual infomation	36, 37, 39, 43, 46, 48, 49, 50, 51, 52, 53, 58,	12
The purpose of the text	37, 41, 42, 60	4
Identify the meaning of	34, 44, 56	3

vocabulary		
Making inference	31, 38	2
Identify inference	32, 54, 3	3
Total		30

b. Calibration

1. Validaty

The validity of the items for multiple choice questions was tested using Pearson's product moment technique (Safari, 2005: 35):

$$r_{yx} = \frac{n\sum x.y - \sum x.\sum y}{\sqrt{n\sum x^2 - (\sum x)^2}\sqrt{n\sum y^2 - (\sum y)^2}}$$

r_{yx} = correlation coefficient of data x to data y.

x = score of a specific item for each student.

y = total question score for each student.

n = number of trial samples.

The ryx value obtained from the calculation is then compared with the r table. The criterion is if ryx > rtabel then the question item is said to be valid.

To calculate the validity of the reading comprehension instrument items using the Pearson product moment correlation formula, where the criteria for receiving instrument items are valid or not used the instrument validity test with r table

determined by one-sided test with a significant level (α) = 0.05 and the degree of confidence (df) = k-2, k = the number of respondents in the trial. A reading comprehension instrument item is a valid instrument if it has a value of r count > r table.

Table III.9
The results of the calculation of the validity test of reading comprehension

number	r_{count}	r_{table}	status
1	0,624	0,361	valid
2	0,594	0,361	valid
3	0,671	0,361	valid
4	0,52	0,361	valid
5	0,583	0,361	valid
6	0,561	0,361	valid
7	0,51	0,361	valid
8	0,479	0,361	valid
9	0,521	0,361	valid

number	r_{count}	r_{table}	status
10	0,442	0,361	valid
11	0,839	0,361	valid
12	0,447	0,361	valid
13	0,409	0,361	valid
14	0,461	0,361	valid
15	0,397	0,361	valid
16	0,49	0,361	valid
17	0,407	0,361	valid
18	0,407	0,361	valid
19	0,717	0,361	valid
20	0,578	0,361	valid
21	0,409	0,361	valid
22	0,433	0,361	valid
23	0,839	0,361	valid
24	0,645	0,361	valid
25	0,388	0,361	valid
26	0,472	0,361	valid
27	0,39	0,361	valid
28	0,614	0,361	valid
29	0,528	0,361	valid
30	0,615	0,361	valid

Based on the test results of the instrument, it can be concluded that there are 30 valid reading comprehension variable instruments because it has a value of r count> r table, so it can be

concluded that the valid instrument on the reading comprehension variable used in the study is 30 questions.

2. Reliability

This reading comprehension instrument uses a Likert scale, so the instrument reliability testing is carried out for all test items using the Alpha Croanbach formula:

$$r_{AC} = \frac{k}{k-1}[1 - \frac{\Sigma(S_i)^2}{(S_x)^2}]$$

k = number of valid items

S_i^2 = variant score of instrument items

S_x^2 = total score variant

The amount of the variance of a research sample can be formulated:

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$$

s^2 = variant

s = standard deviation

xi = the value of x to i

$\bar{x}$ = average

n = the number of samples

A scale can be said to be reliable, if the r_{AC} coefficient is more than > 0.60, then the data is said to be reliable.

Based on the test results of the reading comprehension instrument which was calculated using the Alpha Chronbach reliability coefficient, it was found that the reliability of reading comprehension instruments was 0.913, which means that it had a very high degree of reliability.

Table III.10
Reliability Test Results

Cronbach's Alpha	N of Items
0.913	30

Furthermore, the calculation results obtained are interpreted with a guideline table to provide interpretation of the correlation coefficient. The guideline table used is the guideline table according to Sugiyono (2016: 231) as follows:

Table III.11
Guidelines for Providing Interpretation of Correlation Coefficients

Coefficient Interval	Correlation Level
0.80-1.00	Very strong

0.60-0.79	Strong
0.40-0.59	Medium
0.20-0.39	Low
0.00-0.19	Very low

Source: Sugiyono, 2016[4]

3. The Difficulty Level of the Question Items

The level of difficulty index or Proportional Correct is denoted by p. The formula is:

$$p = \frac{JB}{N}$$

JB = the number of students who answered correctly

N = the number of students who took the test

The difficulty index of the questions ranges from 0 to 1, meaning that p = 0 means that none of the respondents can answer the questions correctly, on the other hand, if p = 1, all respondents will answer the questions correctly. The criteria for the level of difficulty used in this analysis were: if p < 0.70 for

[4] Sugiyono, *Metode Penelitian Kuantitatif, Kualitatif dan R&D*, (Bandung: PT Alfabet, 2016), p.231

easy category, $0.30 < p < 0.70$ for medium category, and $p < 0.30$ for hard category.

Based on the test results of the reading comprehension instrument, the results of the calculation of the difficulty level of the reading comprehension instrument obtained the following results:

Table III.12

Number	r_{count}	Status
1	0,767	Easy
2	0,7	Medium
3	0,667	Medium
4	0,433	Medium
5	0,633	Medium
6	0,767	Easy
7	0,6	Medium
8	0,7	Medium
9	0,667	Medium
10	0,567	Medium
11	0,633	Medium
12	0,533	Medium
13	0,667	Medium
14	0,8	Easy
15	0,567	Medium
16	0,5	Medium
17	0,733	Easy

Number	r_{count}	Status
18	0,733	Easy
19	0,6	Medium
20	0,5	Medium
21	0,667	Medium
22	0,567	Medium
23	0,633	Medium
24	0,767	Easy
25	0,567	Medium
26	0,5	Medium
27	0,667	Medium
28	0,7	Medium
29	0,6	Medium
30	0,533	Medium

Based on these results it can be concluded that the reading comprehension instrument has a difficulty level with the medium category 24 questions (80%), and the easy category 6 questions (20%).

5. Discriminating Power of Question Items

Discriminating power of question is the ability of the questions to discriminating between students who have mastered and those who have not mastered the material asked by the questions. To calculate the discriminating power of the question, the formula is used:

$$D = P_A - P_B, \qquad \text{with} \qquad P_A = \frac{B_A}{J_A} \text{ dan } P_B = \frac{B_B}{J_B}$$

DP = Index of discriminating power.

JA = The number of participants in the upper group test.

JB = The number of participants in the lower group.

BA = The number of participants in the upper group who answered correctly.

BB = The number of participants in the lower group who answered correctly.

PA = The proportion of participants who answered correctly.

PB = The proportion of the lower group of participants who answered correctly.

Ngalim Purwanto (2004: 144) interprets the discriminating power of question items as follows:

DP : 0,00 – 0,20 : Bad

DP : 0,20 – 0,40 : Sufficient

DP : 0,40 – 0,70 : Good

DP : 0,70 – 1,00 : Very good

Based on the test results of the reading comprehension instrument, the calculation of the discriminating power of the instrument is as follows:

Table III.13

Number	Discriminating power	Status
1	0,467	Good
2	0,6	Good
3	0,533	Good
4	0,333	Sufficient
5	0,467	Good
6	0,467	Good
7	0,4	Sufficient
8	0,467	Good
9	0,533	Good
10	0,333	Sufficient
11	0,733	Very good
12	0,4	Sufficient
13	0,4	Sufficient
14	0,267	Sufficient
15	0,333	Sufficient
16	0,467	Good
17	0,267	Sufficient
18	0,267	Sufficient
19	0,667	Good
20	0,6	Good
21	0,4	Sufficient
22	0,333	Sufficient
23	0,733	Very good
24	0,467	Good

Number	Discriminating power	Status
25	0,333	Sufficient
26	0,333	Sufficient
27	0,4	Sufficient
28	0,333	Sufficient
29	0,4	Sufficient
30	0,533	Good

Based on these results it can be concluded that the reading comprehension instrument has discriminating power with very good categories of 2 questions (6,67%), good categories 16 questions (53,33%), and sufficient categories 12 questions (40%).

G. Technique of Data Analysis

After collecting the data through testing vocabulary mastery and reading comprehension, the researcher then analyzed the data. For the technique of data analysis, the researcher applied a quantitative analysis.

1. Descriptive Statistics

In descriptive analysis, data presentation techniques will be carried out in the form of frequency distribution tables, graphs/diagrams for each variable.

In addition, each variable will also be processed and analyzed for the size of the concentration and location such as mean, mode and median as well as deviation measures such as range, variance, standard deviation, inclination and kurtosis.

The steps for creating a frequency distribution table and presenting polygon and histogram graphs are carried out in the following steps:

a. Specifies the range (R), the largest data minus the smallest data.

b. Specifies the number of classes (k) with the Struges rule,

   K = 1 + 3,3 log n, n = amount of data

c. Determine the length of the interval class (P), i.e.

$$P = \frac{\text{Range}}{\text{Many classes}}$$

d. Determines the lower end of the first class interval, i.e. the smallest data.

e. Make a complete frequency distribution table, by determining the lower end (UB) and the upper end (UA) of each class interval calculating the amount (frequency) of data for each interval class.

f. Make a histogram graph, by first determining the bottom edge (TB) and top edge (TA) for each interval class, namely:

   TB = UB - ½ data unit, and TA = UA + ½ data unit.

g. Graphs the frequency polygon, by first determining the mean (Yi) of each interval class, namely Yi = ½ (UA-UB).

Meanwhile, the size of the center, location and deviation can be determined by the following formulas:

1. Determine the Mean / average (Y), with the formula:

$$Y = \frac{\sum Y_i . fi}{n}$$

2. Determine the Mode (Mo), with the formula:

$$Mo = b + p\left(\frac{b_1}{b_1 + b_2}\right)$$

Mo = Mode

p = Class length

b = Lower limit of the mode class, is the interval class with the most frequency

b1 = The mode class frequency minus the previous closest interval class frequency

b2 = The mode class frequency minus the closest interval class frequency thereafter

3. Determine the Median (Me), with the formula:

$$\text{Me} = \text{b} + \text{p}\left(\frac{\frac{1}{2}n - F}{f}\right)$$

Me = Median

n = Amount of data

F = Sum of all frequencies before class median

f = Median class frequency

b = Lower limit of the median class

p = Median class length

4. Variance (SD) and Standard Deviation, with the formula:

$$SD = \sum_{i=1}^{k} \frac{Yi^2.fi}{n} - \left(\sum_{i=1}^{k} \frac{Yi.fi}{n} \right)^2$$ dan Simpangan Baku (S) = $\sqrt{SD}$

To make it easier for researcher, the descriptive statistical calculations in this study will be completed using the help of the SPSS 22.0 computer program.

2. Test Data Analysis Prerequisite

The analysis prerequisite test is needed to determine whether data analysis for hypothesis testing can be continued or not. In this research, the analysis prerequisite test used consisted of the normality test and the linearity test.

a. Normality Test

Normality test is used to test whether in a regression model the dependent variable, the independent variable or both have a normal distribution or not. A good regression model is if the data distribution is normal or close to normal. The normality test was carried out using SPSS.

Normality testing can be seen with the normal Q-Q Plot graph and the Kolmogorov Smirnov test. The histogram graph compares the observed data with a distribution that is close to normal. Normality can be detected by looking at the distribution of data (points) on the

diagonal axis of the graph or by looking at the residual histogram. Can also be explained:

a. If the data spreads around the diagonal line and follows the direction of diagonal line or the histogram graph shows a normal distribution pattern, then the regression model meets the normality assumption.
b. If the data spreads far from the diagonal and or does not follow the direction of the diagonal line or the histogram graph does not show a normal distribution pattern, the regression model does not meet the normality assumption.

In the Kolmogorov Smirnov test, if the significance is greater than > 0.05, it means that the data is normally distributed. Conversely, if the significance is smaller than < 0.05, it means that the data are not normally distributed.

b. Linearity Test

Linearity test aims to see significantly whether two variables have a linear relationship or not. Linearity test is usually used as a prerequisite for analysis or linear regression. Tests on SPSS using the Test for Linearity at a significance level of 0.05. Two variables are said to have a linear relationship if the significance (Linearity) is less than 0.05.

3. Statistic Hypothesis

Hypothesis testing uses partial correlation and multiple correlation techniques, along with simple linear regression and multiple linear regression.

Statistically, the hypotheses are:

H_0: $r = 0$: H_0 is accepted if $r_{count} < r_{table}$ there is no significant correlation between students' vocabulary mastery and their reading comprehension.

H_a: $r \neq 0$: H_a is accepted if $r_{count} > r_{table}$ there is a significant correlation between students' vocabulary mastery and their reading comprehension.