

CHAPTER III

RESEARCH METHODS

A. Research Design

According to Creswell and Creswell (2018), quantitative research is a method used to investigate problems by gathering numerical data and analyzing it using statistical procedures. This approach enables research to test hypotheses and examine relationships among variables using measurable.

In this study, the quantitative method was chosen to evaluate the effectiveness of the Total Physical Response (TPR) method in enhancing students' vocabulary mastery, as indicated by the pre-test and post-test results. This was in line with Creswell and Creswell's (2018) explanation that quantitative research is appropriate when the goal is to assess relationships and effects through numerical analysis. The method fit this research, which involved statistical comparisons between the experimental and control groups' test scores.

This research employed a quasi-experimental design involving a pre-test and post-test administered to two groups. The experimental group was taught using the Total Physical Response (TPR) method, while the control group received instruction using conventional teaching methods. The pre-test and post-test results were then analyzed to determine the effectiveness on students' vocabulary mastery. This research

employed a quasi-experimental design with two groups: an experimental group that received treatment using the Total Physical Response (TPR) method and a control group that received lecture-methods.

The research consisted of three phases. In the pre-test phase, all participants from both groups took a pre-test to measure their vocabulary mastery before the treatment begins, serving as a baseline for comparison. During the Treatment Phase, which lasted for eight weeks, the experimental group is taught using the Total Physical Response (TPR) method, emphasizing vocabulary learning through physical movement and direct interaction, while the control group continued with conventional teaching methods without TPR. The researcher acted as the instructor for both groups to ensure consistency in factors such as teacher, duration, and materials. In the Post-test Phase, after the treatment period ends, all participants took the same post-test as in the pre-test to assess changes in their vocabulary mastery. The post-test results from both groups were then compared to determine whether the TPR method had a significant impact on improving students' vocabulary mastery.

As stated by Ary et al. (2019), quasi-experimental research is commonly used in educational settings where random assignment is not possible. This type of research still allows for the manipulation of the independent variable, but

participants are not randomly assigned to groups. Instead, intact groups—such as existing classes—are used, which makes it more practical and feasible for classroom-based studies. This approach enables researchers to evaluate the effect of an intervention while maintaining control over other influencing factors.

Table 1. The Schedule of Non – Equivalent Control Group Design.

Class	Pre – test	Treatment	Post - test
Experiment	O_1	X	O_2
Control	O_3	-	O_4

Source : Creswell & Creswell (2018),

B. Place and Time of Research

The research was conducted at SDN 182 Bengkulu Utara, located on Jl. Utama Bukit Berlian, Kecamatan Ulok Kupai, Kabupaten Bengkulu Utara. SDN 182 Bengkulu Utara is a public elementary school under the supervision of the Education Office of North Bengkulu Regency.

The research was planned to be conducted during the even semester of the 2024/2025 academic year, around August to September 2025, over a period of four meetings.

C. Population and Sample

a. Population:

According to Arikunto (2016), population refers to the entire research subject or group from which data can be

collected. It includes all individuals or objects that share similar characteristics relevant to the research objective. Based on this definition, the population in this research consisted of all fifth-grade students at SDN 182 Bengkulu Utara. The school has three fifth-grade classes, namely VA, VB, and VC, with a total number of 64 students. Therefore, all fifth-grade students were considered as the population of this research.

Table 2 Population of the Research

No	Grade Level	Class	Number of Students
1.	Fifth Grade	VA	21
2.	Fifth Grade	VB	22
3.	Fifth Grade	VC	21
	Total		64

Source : SDN 182 Bengkulu Utara, 2025

b. Sample:

This research used Cluster Random Sampling technique. in which all members of the population were included as the sample. Therefore, the sample consists of all 64 fifth-grade students from classes VA, VB, and VC at SDN 182 Bengkulu Utara. To implement the treatment using the Total Physical Response (TPR) method, class VA (21 students) was chosen as the experimental group, while

class VB (22 students) served as the control group. Class VC (21 students) was excluded from the treatment and control due to school scheduling constraints and was not involved in the experiment.

Table 3 Research Sample

Group Type	Class	Number of Students
Experimental	VA	21
Control	VB	22
Total		43

Source : SDN 182 Bengkulu Utara, 2025

D. Oprational Definition of Variables

1. Total physical Response (TPR) Method

Total Physical Response (TPR) is a method developed by Asher (2020) that emphasizes learning language through physical movement in response to verbal input. Learners hear commands and respond with actions, helping to internalize vocabulary through kinesthetic memory. This method is especially effective for young learners who benefit from active engagement and physical activity in language learning.

2. Vocabulary Mastery

According to Nation (2022), vocabulary mastery refers to the ability to recognize, understand, and

appropriately use words in context. It involves both receptive (listening and reading) and productive (speaking and writing) vocabulary knowledge. In this study, vocabulary mastery is measured by students' ability to recall and use target words correctly in written and oral tasks.

E. Data Collection Technique

Data collection is an important stage in research, where the researcher directly interacts with the research subjects to obtain the required information or data (Sugiyono, 2003). In this study, the researcher collected data by administering a vocabulary test to the fifth-grade students of SDN 182 Bengkulu Utara. The test consisted of a pre-test and a post-test. The pre-test was conducted before the treatment, while the post-test was administered after the implementation of the Total Physical Response (TPR) method to determine the improvement in students' vocabulary mastery. The data from both tests were then compared to identify differences in learning outcomes between the control and experimental classes. In addition, the researcher also observed students' initial vocabulary ability before the treatment was given. The following are three techniques used in the data collection process:

a. Test

A test refers to a measurement instrument used to assess various aspects such as personality, aptitude, intellectual level, interest, or individual achievement. In this research, the test was used to measure students' vocabulary mastery in learning English.

a. Pre-test (Initial Test)

Before the Total Physical Response (TPR) method was implemented, this study explains that the researcher first administered a pre-test to the students. This test aims to assess the students' initial vocabulary mastery in English, particularly on the topic "Parts of Our Body" based on the fifth-grade English curriculum.

The test will consist of matching questions, where students are asked to match English words related to body parts with the appropriate pictures. This format is chosen to align with the students' proficiency level and the selected learning material. The vocabulary tested will cover basic body parts such as head, eyes, ears, nose, mouth, hands, feet, etc.

Table 4. Scores of the Research

The Range of Score	Scores Category	
85 - 100	Excellent	A
75 - 84	Good	B
56 - 74	Fair	C
<55	Poor	D

b. Treatment

After the pre-test, students will received treatment through the implementation of the Total Physical Response (TPR) method over four meetings. The material taught during the treatment will focused on “Parts of Our Body”, which is part of the fifth-grade English curriculum.

In each meeting, the students learned vocabulary related to body parts such as head, eyes, nose, ears, mouth, hands, feet, and others through physical movements that represent each word. This technique helped students better comprehend and retain vocabulary by actively engaging their bodies in the learning process. By associating words with movements, the TPR method enhanced students memory and understanding in a fun and interactive way.

c. Post-test (Final Test)

After the treatment period ended, a post-test was administered to evaluate the extent to which students' vocabulary mastery has improved following the implementation of the TPR method. The post-test will use the same type of questions and the same topic as the pre-test, namely matching words to pictures of body parts.

To ensure fairness and minimize memorization, the items in the post-test will be presented in a different order from the pre-test. This approach ensures that improvements in students' scores reflect genuine vocabulary acquisition rather than rote memorization.

b. Vocabulary Test

A vocabulary test was used as the research instrument to determine how well students understood and mastered English vocabulary. This test evaluated the students' ability to recognize, comprehend, and use words correctly in various situations. In this research, the test aimed to determine the effectiveness of the Total Physical Response (TPR) method in improving students' vocabulary mastery. The participants of this study were fifth-grade students at SDN 182 Bengkulu Utara who were selected as the research sample.

The research began with the administration of a pre-test consisting of 35 multiple-choice questions. This test aimed to identify the students' initial vocabulary knowledge before receiving the treatment. After that, the students participated in learning activities using the TPR method for four meetings. During the treatment process, the students learned vocabulary through physical movements that represented the meaning of each word. After completing the learning sessions, a post-test was administered, containing the same number of questions as the pre-test. This test aimed to determine whether there was an improvement in students' vocabulary mastery after being taught through the TPR method.

Each question had one correct answer. The scoring system was simple: a score of 1 was given for each correct answer, while a score of 0 was given for each incorrect answer. After all students completed the pre-test and post-test, the researcher collected and calculated their scores.

Finally, the researcher compared the pre-test and post-test results to determine whether there was an improvement in students' vocabulary mastery. The data were then statistically analyzed to determine whether the Total Physical Response (TPR) method was effective in improving students' English vocabulary mastery.

c. Documentation

Documentation was one of the data collection techniques used to obtain information from written sources such as reports, notes, archives, journals, or other relevant documents. In this study, documentation was used to support the research process by collecting data related to the research title (Amelia et al., 2023). The collected documentation included the list of students who participated in the study, their scores from both the pre-test and post-test, and photographs taken during the learning process.

F. Research Instruments

The instrument used in this research is a vocabulary test designed to assess students' English vocabulary mastery before and after the implementation of the Total Physical Response (TPR) method.

The test focuses on the topic "Parts of Our Body", which is adapted from the fifth-grade English curriculum. Initially, the test consisted of 60 items, which were developed and then administered in a try-out. Based on the try-out results, only the valid items were selected for the final version of the test.

The final test items require students to match vocabulary words with corresponding images of body

parts, ensuring alignment with the lesson material and the students' proficiency level.

Table 5. Vocabulary Mastery Score Criteria

Score Range	Category	Description
90 – 100	Excellent	The students shows excellent understanding and accurate use of vocabulary.
75 – 89	Good	The students understands and accurate use of vocabulary.
60 – 74	Fair	The students has limited vocabulary knowledge and makes several mistakes.
0 – 59	Poor	The students lacks basic understanding of vocabulary and frequently misuses it.

Source : Adapted from Heaton, J. B (1982). Introduction to English Language Testing

G. Data Analysis Technique

1. Validity Test

Before the researcher conducts the research, the researcher conducts the validity of the instrument first. The instrument to be tested must show suitability in the aspect to be tested, according to (Kimberlin & Winterstein, 2008) Validity is the extent to which a study accurately measures what it wants to measure (Kouam Arthur William, 2024),

Table 6. Test Instrument Validity Criteria

No	Cronbach's Alpa	Validity Level
1.	0, 81 – 1,00	Highly
2.	0,61 – 0, 80	High
3.	0,41 – 0,60	Fairly
4.	0, 21 – 0, 40	Low
5.	0,00 – 0, 20	Very Low

Source : Widodo et al, 2023

In addition, researcher also analyzed the test results used computer statistics such as SPSS version 25. And the test results consisted of 60 questions containing three vocabulary indicators. Following were the results of the data analysis of the instrument validity test.

Instrument validity is used to determine whether the test items are appropriate to measure students' vocabulary mastery. In this study, the validity of the instrument was analyzed using the Pearson Product Moment correlation formula. The instrument was tried out before being used in the actual research. The detailed calculation results of instrument validity are presented in Appendix 8. Based on the results, the valid test items were used as the research instrument, while the invalid items were revised or omitted.

In this validity test, the researcher used 60 multiple choice questions with predetermined indicators. This trial involved 30 students at SD 76 Kota Bengkulu as the test subject for the instrument. After the trial was completed, the researcher analyzed the data by calculating the validity of the question. From results of the analysis, it was found that 35 out of 60 question.

The validity test was conducted to determine whether the vocabulary test items were appropriate to measure students' vocabulary mastery on the topic of parts of the body. The indicators of vocabulary mastery and the validity test results were analyzed using the Pearson Product Moment correlation. The detailed indicators of vocabulary validity test and the calculation results are presented in Appendix 9.

2. Reliability of the test

According to Kouam Arthur William reliability refers to the consistency and stability of the results (Kouam Arthur William, 2024). After tested the questions, there were questions that have validity criteria, so after the validity test the next step was the reliability test. The reliability value is measured using Cronbach Alpha, if the Cronbach Alpha value > r-table then the variable can be said to be realizable.

Table 7. Test instrument Reability Criteria

No	Cronbach's Alpa	Reability Level
1.	0,00 – 0,20	Lack of Reliable
2.	0,20 – 0,40	Somewhat Reliable
3.	0,40 – 0,60	Fairly Reliable
4.	0,60 – 0,80	Reliable
5.	0,80 – 1,00	Highly Reliable

Source : Sugiono, 2019

Researchers analyzed the reliability value using the SPSS version 25 program. This reliability test was carried out to see whether the research instrument provided consistent results. The calculation uses the Cronbach's Alpha method, which measures the extent to which the

questions in the instrument are related to each other. The calculation results show the reliability coefficient value (Cronbach's Alpha) as follows:

Table 8. Instrument Reability

Reliability Statistics	
Cronbach's Alpha	N of Items
0,896	60

In this reliability test, the researcher used 60 multiple-choice questions with predetermined indicators, and the test of this question instrument was carried out at SDN 76 Kota Bengkulu which certainly had an accreditation level equivalent to the school that would be the main research location. This trial involved 30 students as test subjects for the question instrument. After the trial was completed, the researcher analyzed the data by calculating the test result data, by calculating the reliability. After processing the reliability data of the test questions, the reliability test coefficient was 0.883, which is based on the category above that the item is said to be highly reliable if it reaches a value of 0.600 or higher (Hari Sugiharto Setyaedhi, 2024). Therefore, it can be assumed that this test has a good level of reliability (Herlawan et al., 2021).

3. Item Difficulty Test

One important component in making research instruments is the item difficulty test. Analyzing the level of difficulty of items means examining test questions in terms of difficulty so that it could be obtained which questions easy (Magdalena, Fauziah, et al., 2021). This test was conducted to evaluate the effect of Wordscapes application on students vocabulary mastery. This test was conducted to analyze the difficulty level of the game so that each question was in accordance with the students abilities. In this process, data was collected from students who completed the game vocabulary test and then analyzed the results. To determine whether the question was easy, difficult, or moderate, the difficulty level was calculated by dividing the number of students who answered correctly by the total number of test takers. The general formula is:

$$P = \frac{Rh}{Nh} + \frac{Ri}{Ni} \times 100\%$$

Description:

P : Difficulty level in percent

Nh : Number of test takers in the highest score group

Rh : Number of correct answers in the highest score group

Ni : Number of the test takers in the low score group

Ri : Number of correct answers in the low score group

$$P : \frac{337}{8} + \frac{204}{8} \times 100 \%$$

$$P : 56,3\%$$

Table 9. Interpretation of difficult question categories

Question Category	Interpretation
0% - 15 %	Very Difficult
16 % - 30 %	Difficult
31% - 70 %	Medium
71% - 85 %	Easy
86% - 100 and	Very Easy

Source : (Gde Yadnyawati, 2019)

In testing research questions, it was important to bratelle balance the level of difficulty of the test items made. It was intended that the questions given are neither too easy nor too difficult. Based on the calculation of the level of difficulty, the tested questions reached 61.6%. According to the applicable criteria, this figure is included in the medium category, because the range of 31% - 70% is categorized as medium difficulty. Therefore, the question is suitable for use in research.

4. Item differentiability test

Differentiating power referred to the extent to which the question could distinguish between students who had high abilities (upper group) and those who had low abilities (lower group) in understanding the material being tested (Pradita et al., 2023). The purpose of this test was to

determine the extent to which vocabulary test items had the ability to differentiate between students who had high and low ability in mastering English vocabulary. Items with high discriminating power will showed a clear difference between students who have mastered the vocabulary. In this research, the test of discriminating power was conducted by dividing students into two groups based on their vocabulary test scores. Students in the upper group received the highest vocabulary test score, and students in the lower group received the lowest vocabulary test score. The general formula is:

$$DP = \frac{BA}{BJ} - \frac{BB}{JB} = NM$$

Description :

DP = Question discriminating power

BA = Number of correct answers from high score group test takers

BB = Number of correct answers from low score group test takers

JA = Number of high score group test takers

JB = Number of low score group test takers

NM = Score Maximum

$$DP = \frac{337}{8} - \frac{204}{8} : 30$$

$$DP = 0,50$$

Table 10. Interpretation of item Differentiability Test

Differentiability Test	Interpretation
0,00 -----0,20	Poor
0,21 ----- 0,40	Enough
0,41 ----- 0,70	Good
0,71 -----1,00	Very Good

Source : (Desman et al., 2025)

The Distinguishing Power Test is conducted to ensure that the test given could effectively identify the level of vocabulary mastery of each student. In this research, the results of the calculation of Distinguishing Power reached 121,66%. When compared to the eligible category, which is in the range of $0.21 < DP(0.33) < 0.40$, it could be concluded that the Distinguishing Power in this research was sufficiently qualified.

5. Normality Test

Normality testing is a statistical method to determine whether data follows a normal distributions, which was an important requirement in parametric statistical analysis (Sihotang, 2023). A test was often used to determine if the data was normal. Normality tests were conducted to ensure the instrument was normal and whether the research is generally outstanding. Researchers

use SPSS to test and establish normality. Then there are testing criteria in the Normality Test, namely :

- a. Sig Value (P Value) $< 0,05$ conclude that the data was not normality distributed.
- b. Sig Value (P Value) $> 0,05$ conclude that the data is normality distributed.

6. Homogeneity Test

The homogeneity test is a test to ascertain whether data from two or more groups have the same variation (variance). This was important for valid statistical analysis (Sihotang, 2023). The homogeneity test determines whether or not population data was homogeneous. Use the following principles to determine homogeneity :

- a. Significance test ($\alpha = 0.05$)
- b. If $\text{sig} > \alpha$, then the variance of each samples is the same (homogeneity)
- c. If $\text{sig} < \alpha$, then the variance of each samples is the not same (not homogeneity) to determine in the research uses SPSS.

7. T – Test

The T-test is an effective tool, but it may be affected under certain conditions, such as when the data is not normally distributed or when it is used on data that does not match (Novak, 2022). The analysis will be carried out using the SPSS T-test. The test was conducted to determine

how significant the partial effect of the independent variable was on the dependent variable when the other independent variables were assumed to be constant. The formula to be used is as follows:

- a. The independent variable affect the dependent variable uniquely if the significance level is less than 0,05
- b. The independent variable does not affect the dependent variable if the significance level is greater than α (0,05)

8. F – test

George W. Snedecor named this statistical tool the F-test, in honor of its inventor, Sir Ronald Fisher, a renowned statistician. The F-test is one of the important tools in statistics used to compare two or more variances, i.e., how much the data is spread out in different groups (Odek & Opuodho, 2023). One of the most commonly used applications of the F-test is in Analysis of Covariance (ANCOVA). ANCOVA is used to compare the means of two or more groups by taking into account one or more control variables (called covariates) that can affect the results of the analysis. With covariates, ANCOVA can reduce the influence of outside variables that are not the main focus of the research, so that the results of comparisons between groups become more accurate and objective. This technique is very useful in experimental

research, especially when trying to determine the effectiveness of a particular treatment or method.

9. The effect size

The effect size is a measure used to describe how much influence or relationship exists between two variables in research. This measure is important because it provides information about how large or strong the effect is, not just whether or not it exists. Two commonly used effect sizes are Cohen's d , which measures the mean difference between two variables in research. This measure is important because it provides information about how large or strong the effect was, not just whether or not it exists. Two commonly used effect sizes are Cohen's d , which measures the mean difference between two groups, and Pearson's r , which measures how strong the relationship was between two variables (Funder & Ozer, 2019). The following was the category of effect size values used to measure how much influence or relationship between two variables in research:

Table 11. Value of the effect size

No.	Value	Effect size
1.	0,00-0,1,95	Very weak effect
2.	0,20-0,395	Weak effect

3.	0,40-0,595	Modest effect
4.	0,60-0,749	Strong effect
5.	0,80-1,00	Very strong effect

Source : Jacob Cohen, 2019

10. Statistical Hypothesis

Hypothesis testing functions in testing H_0 was accepted or rejected and testing the alternative H_a was accepted or rejected. Hypothesis testing can also be done using SPSS version 25 with basic guidelines for collecting the following test results:

- a. If the value of Sig. (2-tailed) < 0.05 it is said that the hypothesis is accepted.
- b. If the value of Sig. (2-tailed) > 0.05 it is said that the hypothesis is rejected.