CHAPTER III
METHODOLOGY

A. Research Design

This study was used with a quantitative method. According to Creswell, the quantitative approach was a research method aimed to be used for objectively tested theories by analyzed relationships between variables or compared groups. These variables were measured using specific instruments, and the collected numerical data was analyzed through statistical procedures (Creswell. 2022). In general, quantitative research was a method in which numerical data was used to study and understand a particular problem or phenomenon. The numerical data that was collected was used to help researchers gain insights and make decisions in various fields.

According to Yuliana, experimental research was a type of study that was conducted through experiments. This research was classified into the quantitative category and was aimed at determining the effect of the treatment (independent variable) on the outcome (dependent variable) was examined under controlled conditions (Yuliana, 2022). The goal was to examine the effect of certain variables using standardized measuring instruments. The design that was used was Quasi-Experimental, which was characterized by the presence of a control group; however, external factors that might have

affected the experimental results could not be fully controlled.

The study was implemented using a quasi-experimental design with a pretest-posttest control group approach to examine the impact of video-based multimedia on students' English vocabulary mastery at SMPN 52 Bengkulu Utara. The research sample was composed of two classes: an experimental group that was given vocabulary instruction using video-based multimedia as the teaching method, and a control group that received traditional vocabulary instruction. Both groups were administered a pretest to measure their initial vocabulary knowledge, followed by an intervention period that lasted several weeks. Vocabulary lessons for the experimental group were delivered using video-based multimedia resources, while the control group was taught using conventional methods. After the intervention, both groups were given a posttest to assess changes in their vocabulary mastery. The data collected from the pretest and posttest were analyzed using descriptive statistics and an independent t-test to determine the significance of the difference in vocabulary acquisition between the two groups. In addition, feedback from the experimental group was gathered through questionnaires to evaluate students' engagement and motivation while using video-based multimedia. This study aimed to contribute to the growing body of knowledge on the use of digital tools in language learning by highlighting the potential benefits of

video-based multimedia in improving vocabulary mastery in the classroom.

Table 1. Pre-Test Post-Test Nonequivalent Control Group Design

Group	Pre-test	Treatment	Post-test
Control class	O1	X	O2
Experiment class	O3		O4

Description:

E : Experimental Group

K : Control Group

X : Learning Method

01 : Pretest in experimental Class

02 : Posttest on the experimental class

03 : Pretest in control class

04 Posttest on control class

B. Place and Time of Research

This research was conducted at SMP Negeri 52 North Bengkulu, which was located in Desa Tanjung Harapan, Kecamatan Ulok Kupai, Kabupaten Bengkulu Utara. The implementation of the research was carried out once a research permit had been issued by the university.

C. Population and Sample

1. Population

The next stage in the research procedure was to determine which subjects or students would The population and sample were determined for this study. According to Ary (2010), a population refers to all members of a clearly defined group, which may include people, events, or objects..' According to Sugiyono in Supardi, the population was a generalization region consisting of objects or subjects that were characterized by certain quantities and traits defined by the researcher in order to be studied and from which conclusions were drawn.

Based on the opinions mentioned above, it could be concluded that the population referred to the objects or subjects located in a specific region and meeting certain criteria related to the research problem or objectives. The population in this study was taken from the seventh-grade students of SMP Negeri 52 North Bengkulu in the academic year 2024/2025.

The population of this research consisted of 60 seventh-grade students of SMP Negeri 52 North Bengkulu. In selecting the sample, the researcher chose class VII7 as the control class, which consisted of 30 students, and class VII8 as the experimental class, which also consisted of 30 students.

The population of this study consisted of 60 seventh-grade students of SMPN 52 North Bengkulu, divided into two classes.:

Table 2. Total Students

No	Class	Total
1.	7A	30
2.	7B	30
Total amount		60

2. Sample

The sample was part of the population that had been observed (Ary, 2010). Purposive sampling was used by the researcher in this study. According to Sugiyono (2018), purposive sampling was a sampling technique carried out with certain considerations. Purposive sampling was a non-probability sampling method, which meant that not all members of the population were given the same opportunity to be selected.

Purposive sampling was used by the researcher to select the experimental and control classes based on the English teacher's suggestion or assessment that both classes had the same level of English proficiency. As a result, class 7A, consisting of 31 students, was chosen as the experimental class, and class 7B, consisting of 30 students, was selected as the control class.

Table 3. total students in the experiment and control class

No	Class	Male	Female	Total
1.	Experimental Group	17	14	30
2.	Control group	14	16	30
	Total			60

D. Operational Definition Variables

Operational definition was the process of defining and explaining variables in a study until they became practically meaningful, allowing them to be used as measurable instruments in research (Utley et al.,2022). This study explained the variables used as the main focus, namely the Independent Variable (X) and the Dependent Variable (Y). The independent variable was defined as the variable that affected or caused changes in the dependent variable, while the dependent variable was the variable that was affected or measured in the experiment as a response to changes in the independent variable (Ira Yuniati, 2022). In this context, an explanation of the variables used in the study was provided, namely:

1. Independent Variable (Educational Video)

   Educational videos were audiovisual materials used as didactic resources to enhance teaching and learning

processes by making concepts easier to understand and more engaging for students. (Serrano-arenas, 2023).

2. Dependent Variable (Vocabulary Mastery)

Vocabulary was essential to language competency and impacted learners' performance in all language domains, including speaking, listening, reading, and writing. Vocabulary mastery was defined as the ability to acquire, understand, and effectively use words in various contexts with accuracy and confidence. According to Havwini, vocabulary mastery was considered crucial for understanding new concepts and ideas, as well as facilitating better communication.(Havwini et al., 2024). In this study, vocabulary mastery was specifically defined as students' ability to learn, retain, and apply English vocabulary effectively through the use of video-based multimedia, which was expected to enhance their language skills in a more engaging and interactive manner.

E. Data Collection Technique

Data collection was a process in research in which researchers interacted directly with the research objects to obtain the information or data needed (Indonesia, 2021). Data collection techniques were divided into tests, vocabulary tests, and documentation**.**

1. Tes

A test was a tool used to measure the extent to which

students had learned by providing them with a number of questions or tasks that had to be completed. The level of students' ability in the matter to be measured was assessed through the test. Tests should have been conducted with meticulous scoring, a sufficient number of trials, and statistical evidence that could be considered reliable (Amelia et al., 2022).

1. Pre-test

   The pre-test was an assessment conducted before the learning process began, aimed at measuring students' prior understanding of the material to be taught (Kelas et al., 2023). In this context, the pre-test served to evaluate the effectiveness of the upcoming instruction. The pre-test was administered to measure students' initial ability before the implementation of the educational video-based multimedia method. It was given to all students in both the experimental and control groups, each consisting of 25 students, totaling 50 participants. The test was conducted before the learning process began to assess the students' vocabulary mastery prior to receiving any treatment. The pre-test contained 20 multiple-choice questions covering general vocabulary appropriate for junior high school students. The test was conducted in their respective classrooms under a calm and conducive environment and was supervised by both the researcher and the English teacher to ensure that the

results truly reflected the students' original abilities without interference.

2. Post-test

The post-test was an assessment carried out after the completion of the teaching and learning process, aiming to determine how well students had mastered the material that had been taught. (Kelas et al., 2023). This type of test was also referred to as a formative test. The purpose of the post-test was to evaluate the students' level of achievement following the treatment in terms of both knowledge and skills after participating in a learning activity. It was given after students had finished learning, with the aim of comparing students' abilities before and after treatment. The post-test was conducted after the learning sessions in both the experimental and control classes were completed. The purpose of the post-test was to observe the progress and differences in students' learning outcomes after the treatment, particularly in English vocabulary mastery. The post-test consisted of 20 multiple-choice questions, equivalent to those used in the pre-test, with the same level of difficulty to allow valid comparison. The test was administered during the final meeting, under the same conditions and time duration as the pre-test, which was 30 minutes. All students were instructed to work independently and

honestly to ensure the results accurately reflected individual improvement.

1. Treatment

   The treatment in this study was the use of educational video-based multimedia learning methods to improve vocabulary acquisition during the learning process. The treatment was The participants were divided into two groups: an experimental group and a control group, with the treatment applied only to the experimental class. The videos presented various vocabulary terms used in everyday contexts such as school, home, and public places, which were adapted to the junior high school English curriculum. Each video was about 10 to 15 minutes long and was shown in four learning sessions, on the 6th, 10th, 14th, and 18th. In each session, students first watched the videos, followed by discussions and vocabulary exercises based on the video content. On the 6th and 10th, students watched and studied the first video, and on the 14th and 18th, students watched and studied the second video. Meanwhile, the control class was taught with traditional methods such as lectures, reading textbooks, and written exercises without video media support. The implementation of the treatment was scheduled according to the school schedule so as not to interfere with other lessons and was supervised by the researcher and the class teacher to ensure the learning

activities went according to plan.

3. Vocabulary Test

A vocabulary test was a tool designed to measure students' understanding and mastery of vocabulary. It evaluated their ability to recognize, understand, and use words appropriately in different contexts (Cervatiuc, 2007). In this study, the vocabulary test included both pre-test and post-test to assess the impact of video-based multimedia on students' vocabulary mastery. The test consisted of multiple-choice questions, matching items, and sentence completion tasks, ensuring a comprehensive evaluation of students' vocabulary skills.

In this research, the researcher administered a multiple-choice test to the students. The test consisted of 25 items and was given within 50 minutes. The data from the students' tests were collected and analyzed to obtain the test results. Before giving treatment to the experimental class, the researcher administered a pre-test to both the experimental and control classes, which consisted of 60 items. The items were selected from valid tryout items. First of all, the researcher conducted a tryout on a different class in the same school to obtain valid and reliable items. After analyzing the tryout results, the researcher found 28 items with good validity and reliability.

Then, the researcher gave treatment to the

experimental class. At the final stage of the research, a post-test was administered to both the experimental and control classes. The pre-test was used to determine whether the two groups were essentially the same in terms of the dependent variable at the beginning of the study. The treatment was then applied to the experimental class using educational video-based multimedia, while no treatment was given to the control class. After the treatment, a post-test was administered to both groups. This test aimed to determine the difference in scores between the two classes and to identify which treatment method was more effective.

Table 4. Score Category

The Rage of Score	Scores Category	
20-18	Excellent To Very Good	sophisticated range-effective Word idiom choice and usage etc.
17-14	Good To Average	Adaquate range casional errors of of rd/idiom form, choice, age but meaning not scured.
13-10	Fair To Poor	limited range frequent errors of word/idiom form, choice, usage etc.

9-7	Very Poor	sentially translation- little knowledge of English vocabulary.

Source : J. B. Heaton, 1997

4. Documentation

Documentation was a data collection method that involved gathering information from various types of documents or written sources. The documents could include texts, reports, notes, archives, journals, or other records relevant to the research objectives (Amelia et al., 2023). In this study, the documentation included a list of the names of students who were selected as research subjects, as well as the grades required for the study. The documentation also consisted of photographs of learning activities, pre-test and post-test results based on the material given, all of which were compiled and could be seen in the appendices.

F. Research Instruments

Research instruments were tools used to collect data from research subjects. In quantitative research, instruments were very important because they were used to measure the variables being studied by the researcher (Taqwin, 2022). In this study, the instrument used was a test. A test was an instrument designed to collect data or information regarding a person's knowledge and skills (Taqwin, 2022). The test used in this study was a vocabulary test, conducted to determine whether the use of educational video-based multimedia had an

effect on improving students' vocabulary. The test was constructed in the form of multiple-choice questions and consisted of a pre-test and a post-test, both of which were administered in two groups: the experimental class and the control class.

In addition, the test instrument was evaluated for validity and reliability. This step was conducted before the test items were used in the actual research. After the items were tested and confirmed to meet the criteria of feasibility, they were then used in the pre-test and post-test for both the control and experimental classes.

G. Data Analysis Technique

1. Validity of the test

Before starting the research, the researcher had to ensure the validity of the instrument. The instrument to be tested had to demonstrate its suitability for the elements to be tested. According to Anastasi and Urbina (2007), the validity of a measuring instrument indicated the extent to which the tool actually measured what was intended to be measured and how well the tool was used for the purpose of the measurement. In other words, validity could be defined as the correspondence between test scores or measurement results and the quality that should be measured (Kaplan & Saccuzzo, 2017). Validity could also be explained as the extent to which a test measured what it was intended to

measure (Aiken, 1980). In general, validity related to the extent to which a measuring instrument or test could provide precise results in accordance with what was intended to be measured (Setiawan et al., 2020).

Table 5. Instrumen Validity

Variable	Item	r table	r count	Result
	S1	0,432	0,310	Unvalid
	S2	0,432	0,485	Valid
	S3	0,432	0,263	Unvalid
	S4	0,432	0,514	Valid
	S5	0,432	0,626	Va;id
	S6	0,432	0,252	Unvalid
	S7	0,432	0,314	Unvalid
	S8	0,432	0,416	Unvalid
	S9	0,432	0,369	Unvalid
	S10	0,432	0,460	Valid
	S11	0,432	0,289	Unvalid
	S12	0,432	0,208	Unvalid
	S13	0,432	0,243	Unvalid

	S14	0,432	0,281	Unvalid
	S15	0,432	0,257	Unvalid
	S16	0,432	0,055	Unvalid
	S17	0,432	0,149	Unvalid
	S18	0,432	0,311	Unvalid
	S19	0,432	0,501	Valid
	S20	0,432	0,701	Valid
	S21	0,432	0,632	Valid
	S22	0,432	0,649	Valid
	S23	0,432	0,567	Valid
	S24	0,432	0,551	Valid
	S25	0,432	0,481	Valid
	S26	0,432	0,459	Valid
	S27	0,432	0,730	Valid
	S28	0,432	0,560	Valid
	S29	0,432	0,163	Unvalid
	S30	0,432	0,327	Unvalid
	S31	0,432	0,566	Valid

	S32	0,432		Unvalid
	S33	0,432	0,588	Valid
	S34	0,432	0,164	Unvalid
	S35	0,432	0,372	Unvalid
	S36	0,432	0,177	Unvalid
	S37	0,432	0,437	Valid
	S38	0,432	0,658	Valid
	S39	0,432	0,285	Unvalid
	S40	0,432	0,500	Valid
	S41	0,432	0,238	Unvalid
	S42	0,432	0,208	Unvalid
	S43	0,432	0,524	Valid
	S44	0,432	0,543	Valid
	S45	0,432	0,740	Valid
	S46	0,432	0,087	Unvalid
	S47	0,432	0,342	Unvalid
	S48	0,432	0,090	Unvalid
	S49	0,432	0,563	Valid

	S50	0,432	0,096	Unvalid
	S51	0,432	0,296	Unvalid
	S52	0,432	0,426	Unvalid
	S53	0,432	0,727	Valid
	S54	0,432	0,492	Valid
	S55	0,432	0,793	Valid
	S56	0,432	0,016	Unvalid
	S57	0,432	0,285	Unvalid
	S58	0,432	0,568	Valid
	S59	0,432	0,279	Unvalid
	S60	0,432	0,488	Valid

Table 6. Test Instrument Validity Criteria

Value of r	Interpretation
0.81 – 1.00	Very high
0.61 – 0.80	High
0.41 – 0.60	Fair
0.21 – 0.40	Low
0.00 – 0.20	Very low

Source: (Ajar & Penelitian, 2020)

In this validity test, the researchers used multiple-choice vocabulary questions with a total of 60 questions.

Then, the researcher conducted a trial in one of the junior high schools that had the same class and accreditation as the school where the researcher would later conduct the research. The trial was conducted with a total of 21 students as the test subjects. After conducting the trial, the researcher processed the data from the trial questions by calculating the validity of each question. As a result, the researchers found a total of 28 valid questions out of 60 questions, with the following indicators:

Table 7. Indicators of Vocabulary Test Validity

No	Indicators	Number of Item	Items
1	Understanding word meanings in context	5	2, 4, 5, 10, 20
2	Identifying adjectives and their meanings	1	33
3	Using verbs correctly in sentences	6	19, 22, 24, 35, 38, 49
4	Understanding common phrases and expression	5	26, 27, 28, 31, 40
5	Recognizing vocabulary related to daily activities	5	21, 23, 43, 44, 45
6	Identifying nouns in specific categories	1	37
7	Understanding auxiliary verbs and imperative expressions	5	53, 54, 55, 58, 60

	Total	**28**

2. Reability of the test

According to Kouam Arthur William, reliability is related to the consistency and stability of results (Kouam Arthur William, 2024). After testing the questions, some questions met the validity criteria, so the next step after the validity test was the reliability test. The reliability value was measured using Cronbach's Alpha; if the Cronbach's Alpha value was greater than the r-table value, then the variable could be said to be reliable. After testing the questions and obtaining those with valid criteria, the researcher proceeded to the reliability test.

Table 8. Test Instrument Reability Criteria

Alpha	Reliability Level
0,00 – 0,20	Less Reliable
0,20 – 0,40	Somewhat Reliable
0,40 – 0,60	Moderately Reliable
0,60 – 0,80	Reliable
0,80 – 1,00	Very Reliable

Source : (Iii et al., 2023)

In this reliability test, the researchers used multiple-choice questions with a total of 60 items. Then, the researchers conducted a trial in one of the junior high schools that had the same level as the school where the research would later be conducted. The trial involved 21

students as test subjects. After conducting the trial, the researchers processed the test data by calculating the reliability. The reliability coefficient obtained from the test questions was 0.897, which falls into the category of reliable items since it exceeds the threshold value of 0.600 (Hari Sugiharto Setyaedhi, 2024). Therefore, it can be assumed that this test has a good level of reliability.

Table 9. The Result of Reliability

Reliability Stetistics	
Cronbach's Alpha	N of Items
0.897	60
Reliabel	

Based on the table above, it can be concluded that the 60 questions tested on 21 students are highly feasible to be used in the pre-test and post-test conducted by the researchers. This is supported by the Cronbach's Alpha value of 0.897, as shown in Table 7, which falls into the category of Good reliability.

3. The Level Difficulty Test

The difficulty level of a test is an essential aspect in developing research instruments, including the implementation of educational video-based multimedia to enhance students' English vocabulary mastery. The difficulty level of a test item refers to how challenging a question or task is within an evaluation instrument. This test is conducted to determine how difficult a particular question is for the target population being assessed. In the

context of using educational video-based multimedia to improve students' vocabulary mastery, this test ensures that the exercises or assessments provided align with students' proficiency levels.The process involved collecting data from 8th-grade students at SMPN 52 North Bengkulu, who were given a series of vocabulary-related tasks embedded in video-based learning materials. After completing these exercises, their results were analyzed to evaluate the effectiveness and difficulty level of each item. The difficulty index of a test item is typically calculated by dividing the number of students who answered correctly by the total number of students who participated in the test. The level of difficulty is a numerical measure that shows how challenging a question item is. A question is considered good if it is neither too easy nor too difficult (Saputri et al., 2023).

The general formula is :

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

Description:

P = Difficulty level in percent

N_h = Number of test takers in high score group

R_h = Number of correct answers in high score group

N_i = Number of test takers in low score group

R_i = Number of correct answers in low score group

$P = \frac{332+75}{6+6}$ X 100%

P = 67,83%

Table 10. Interpretation of difficult question categories

Question category	Erpretation of results
Classified as easy	71% ---------- 100%
Classified as medium	41% ---------- 70%
Classified as difficult	21% ---------- 40%

(*Unhi Press 2019*, 2019)

In the development of educational video-based multimedia, it is essential to balance the difficulty level of exercises or tests provided. This ensures that the material presented is neither too easy, which could reduce students' motivation to learn, nor too difficult, which might lead to frustration. The difficulty level test helps create adaptive content that aligns with the needs of 7th-grade students at SMPN 52 North Bengkulu.

Based on the difficulty level calculation of the test items, the result reached 67.83%. According to the difficulty level classification, this percentage falls into the moderate category, as it ranges between 41% and 70%. Thus, the test items used in this study are considered appropriate for students' comprehension levels, effectively supporting the achievement of learning objectives.

4. Item Discrimination Test

The upper group consists of students with the

highest scores on all tests or exercises within the educational video-based multimedia, while the lower group includes students with the lowest scores on the same assessments. Typically, 27% of students with the highest scores and 27% of students with the lowest scores are selected as samples for discrimination index analysis.

The discrimination index of a test item is calculated by subtracting the proportion of lower-group participants who answered correctly from the proportion of upper-group participants who answered correctly. This analysis aims to measure how well a test item differentiates between high- and low-achieving students, ensuring that the questions effectively assess students' comprehension.

The general formula is :

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

Where:

DP = Question Distinguishing Power.

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

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

N = Number of students in each group (27%)

DP = $\frac{332-75}{12}$

DP = 21,41

Table 11. Interpretation item discrimination test

Item discrimination test	Interpretation of result
DP ≥0,40	Good
+0,21 < DP < 0,40	Fairly Qualified
0, < DP ≤ 0,20	Weak

Source: (Restiyawati, 2023)

In the development of educational video-based multimedia for 7th-grade students at SMPN 52 North Bengkulu, the discrimination index test is used to ensure that the exercises and assessments effectively identify students' levels of English vocabulary mastery. This test plays a crucial role in verifying that each item in the multimedia-based exercises presents an appropriate level of difficulty, allowing differentiation between students who have mastered the material and those who still face challenges.

Based on the results, the discrimination power (DP) calculation in this study reached 0.33. According to the interpretation criteria, where a DP value between +0.21 and 0.40 is considered fairly good, the discrimination power in this study is classified as quite eligible.

5. Normality Test

The normalization test is a statistical procedure used to determine whether the data in a sample are normally distributed. Data are considered normally distributed if the normal probability plot shows that the residual data form a straight line or closely follow a straight line. Additionally, the histogram results should display a bell-shaped distribution of the residuals (Ira Yuniati, 2022),

Then there are testing criteria in the Normality Test, namely:

- Sig Value (P Value) < 0,05 concluded that the data was not normally distributed.
- Sig Value (P Value) > 0,05 conclude that the data is normally distributed.

6. Homogeneity Test

The homogeneity test determines whether or not population data is homogeneous. Use the following principles to determine homogeneity:

1. Significance test (α) = 0.05
2. If sig>α, then the variance of each sample is the same (homogeneity).
3. If sig<α, then the variance of each sample is the not same (not homogeneity) todetermine it the researcher uses SPSS.

7. T-Test

The t-test is a powerful tool, but its effectiveness can be

compromised in certain situations, such as when the data is not normal or when it is used on data that is not suitable. Therefore, there is a need for alternative approaches or modifications to overcome this (Novak, 2020).

8. F-Test

The F-test was a statistical tool named by George W. Snedecor given in honor of Sir Ronald Fisher because he developed the F-test first. The F-test was used to compare two variances, as well as in analysis (ANCOVA) and multiple regression. In addition, the F-test had a function to evaluate whether linear regression used the most appropriate line for a particular data set. Therefore, this F-test was calculated as a ratio of variances and various evaluation statistics involving the F value known as the F-test (Odek & Opuodho, 2023).

9. The effect size

The effect size was a measure used to describe how much influence or relationship existed between two variables in research. This measure was important because it provided information about how large or strong the effect was, not just whether or not it existed. Two commonly used effect sizes were Cohen's d, which measured the mean difference between two groups, and Pearson's r, which measured 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 existed between two variables in research:

Table 12. Value of the effect size

No	Velue	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 was conducted using the Independent Sample t-test statistical test to compare the average post-test results between the experimental group that used multimedia-based videos and the control group that used conventional learning methods. The significance level (α) used was 0.05. The test criteria were:

- If the p-value < 0.05, then H_0 is rejected and H_a is accepted, which means there is a significant effect of using multimedia-based videos on students' vocabulary mastery.
- If the p-value ≥ 0.05, then H_0 is accepted and H_a is

rejected, which means that there is no significant effect of using multimedia-based videos on students' vocabulary mastery.