

CHAPTER III

METHODOLOGY

A. Research Design

This study employed a quantitative method. The quantitative method is used to objectively test theories by measuring variables and producing numerical data that are analyzed through statistical procedures. This is consistent with the explanation of Creswell and Creswell (2017), who state that quantitative research involves testing objective theories by examining relationships among variables and analyzing numeric data using statistical procedures. Therefore, this study is designed to generate objective and measurable findings that can be used to test previously formulated hypotheses or theories (Creswell & Creswell, 2017).

The type of research conducted was experimental research. Experimental research is carried out through a systematic manipulation of an independent variable to observe its effect on a dependent variable under controlled conditions. This study falls under the category of quantitative research because it aims to examine the influence of an independent variable (treatment) on a dependent variable (outcome) under controlled conditions. This aligns with the explanation of Tanner (2018), who state that experimental research involves manipulating one or more variables while controlling other

influencing factors to determine causal relationships. They emphasize that experimental methods provide a structured and controlled approach that enables researchers to assess cause-and-effect connections accurately.

The research design applied in this study was a Quasi-Experimental design. In a quasi-experimental design, a control group is included; however, the researcher cannot fully control external variables that may influence the research outcomes. Abraham, I., & Supriyati (2022) highlight that quasi-experiments are employed when full randomization is not feasible, resulting in limited control over external factors that may affect the study. The specific design used in this study was the nonequivalent control group design, which involves an experimental group and a control group that are not randomly assigned. Both groups receive a pretest and a posttest, enabling the researcher to compare differences after the treatment is administered (Abraham, I., & Supriyati, 2022). This design allows the researcher to analyze changes in each group to determine the effectiveness of the treatment.

Table 3. 1 Pre-Test Post-Test Nonequivalent Control Group Design

Experiment	O1	X	O2
Control	O3	-	O4

Description:

E : Experimental group

K : Control group

X : Application Media

O1 : Pretest in experimental class

O2 : Posttest on the experimental class

O3 : Pretest in control class

O4 : Posttest on control class

B. Location and Time of Research

This study was conducted on eighth grade students at SMP Negeri 5 Kota Bengkulu, located at Jalan RE. Martadinata II, Pagar Dewa, Kecamatan Selebar, Kota Bengkulu. The school has various facilities to support the teaching and learning process, including: 11 canteens, 1 cooperative, 26 toilets, 1 guard room, 31 study rooms, one office room, one library room, one sports room, one student council room, two laboratory rooms, one art room, one administration room, one principal's office, one school health unit room, one prayer room, one hall, one field, and one warehouse.

C. Population and Sample**1. Population**

The next step in the research procedure was identifying the students who would serve as the population of the study. According to Amin (2023), a population refers to the entire group of individuals or subjects that share

specific characteristics relevant to the objectives of the research. They explain that the population serves as the primary source of data in a study and may include individuals, groups, or objects that meet predetermined criteria.

Similarly, Suriani, N., & Jailani (2023) describe the population as all subjects within a defined scope, setting, and period who become the target of the research. They emphasize that determining the population is important because it influences the extent to which research findings can be generalized. They also note that a population consists of individuals whose characteristics align with the focus and boundaries of the study.

Based on these expert opinions, population selection becomes a crucial step in identifying the appropriate group to be examined. A population includes individuals with certain characteristics located within a specific institution or environment relevant to the research focus. In this study, the population consisted of all eighth-grade students at SMP Negeri 5 Bengkulu City in the 2025/2026 academic year.

This population comprised 364 students, divided into eleventh classes: class VIII A with 33 students, class VIII B with 32 students, class VIII C with 33 students, class VIII D with 34 students, class VIII E with 33 students, class

VIII F with 34 students, class VIII G with 33 students, class VIII H with 33 students, class VIII I with 33 students, class VIII J with 34 students and class VIII K with 32 students. In conducting the research, the researcher designated class VIII G as the experimental class and class VIII H as the control class. This classification was carried out to compare learning outcomes between the group receiving the treatment and the group that did not receive the treatment.

Table 3. 2 Population of The Research

No	Class	Total
1.	VIII A	33
2.	VIII B	32
3.	VIII C	33
4.	VIII D	34
5.	VIII E	33
6.	VIII F	34
7.	VIII G	33
8.	VIII H	33
9.	VIII I	33
10.	VIII J	34
11.	VIII K	32
Total		364

Source: SMP N 5 Kota Bengkulu, 2025

2. Sample

In this research, a portion of the population was selected to become the focus of the study and used as the research sample. A sample is a part of the population that possesses certain characteristics and is considered representative of the entire population (Suriani, N., &

Jailani, 2023). By studying this sample, the researcher is able to draw conclusions about the population as a whole.

This study employed a purposive sampling technique. Purposive sampling is a non-probability sampling method in which participants are selected based on specific characteristics or criteria relevant to the research objectives (Campbell, S., et. al., 2020). This method allows researchers to select participants who are most likely to provide useful and relevant data. Additionally, purposive sampling includes several techniques, such as extreme case sampling, typical case sampling, and homogeneous sampling, which can be applied depending on the research context (Nyimbili, F., & Nyimbili, 2024).

The use of purposive sampling in this study was guided by the researcher's considerations, taking into account the students' conditions and learning needs. The researcher selected classes that were deemed appropriate to serve as the experimental and control classes, with the expectation that the results would accurately reflect the effectiveness of the treatment provided.

The sample in this study consisted of 66 eighth-grade students from classes VIII G and VIII H at SMP Negeri 5 Bengkulu City. Class VIII G was designated as

the experimental class, while class VIII H served as the control class, as shown in the table below.

Table 3. 3 Total Sample of The Research

Class	Subject Of Student	Male	Female	Number Of Student
VIII G	Experimental Group	15	18	33
VIII H	Control Group	14	19	33
Total				66

Source: SMP N 5 Kota Bengkulu, 2025

D. Operational Definition of Variables

Variables are components that can be measured, observed, or manipulated to gain an understanding of a particular phenomenon. A proper understanding of variables is essential so that research can be designed validly and produce accurate conclusions. In this study, there are two types of variables: independent variables and dependent variables. The independent variable influences or causes changes in other variables, whereas the dependent variable is affected by the independent variable. The variables used in this study are described as follows:

1. Students' Vocabulary Mastery (Y)

Students' vocabulary mastery is defined as the ability of students to recognize, understand, and retain English vocabulary that has been taught. Operationally, this variable is measured through vocabulary tests (pre-test and post-test) before and after the game-based intervention, with test scores serving as an indicator of vocabulary

mastery. Research results show that after students learned using the Blooket game, there was a significant improvement in their vocabulary mastery (Isyamirahim, R., et. al., 2024).

2. Use of Blooket Game (X)

The use of the Blooket game is defined as the implementation of an online game-based learning media in the process of learning English vocabulary. Operationally, this variable is measured based on students' participation in Blooket sessions as part of the learning activities (e.g., quizzes or vocabulary exercises through Blooket). The use of Blooket makes the vocabulary learning process more enjoyable, competitive, and helps students understand vocabulary more easily compared to conventional methods (Sartika, K. D., et. al., 2023).

E. Data Collection Technique

Data collection is an essential stage in research, in which the researcher directly interacts with the subjects to gather the information or data required for the study. In this research, data were obtained through a vocabulary test administered to both the experimental and control groups. This test consisted of a pre-test and post-test, which generated the quantitative data needed to determine whether the use of Blooket games had an effect on students' vocabulary mastery. The data collection process involved several steps: the administration of the pre-

test to measure initial vocabulary knowledge, the implementation of the learning intervention, and the administration of the post-test to identify improvement after treatment. The primary goal of data collection was to discover whether the integration of Blooket into the learning process contributed to enhancing the vocabulary mastery of students at SMP Negeri 5 Kota Bengkulu. Differences between the pre-test and post-test results were then analyzed to measure the level of improvement.

1. Test

A test is a tool or procedure designed to measure specific variables or constructs, in this case, students' vocabulary mastery. Tests must have validity, reliability, and clear administration standards. In other words, tests are formal instruments used to scientifically assess participants' abilities (Harianingsih, F., & Hastini, 2023).

a) Pre-test

A pre-test is a test given before learning activities begin. This initial test aims to determine the initial knowledge of students in the experimental class and serves as a benchmark for assessing participants' improvement after the treatment is given (Alam, M. I., & Susanti, 2025).

In this study, the researchers gave 20 multiple-choice pre-test questions to students in the

experimental and control classes. This pre-test served to determine the students' initial ability in vocabulary mastery before the treatment was given. The aim was to measure the students' initial level of knowledge so that it could later be compared with the post-test results. Thus, the researchers could assess the extent to which the use of Blooket could affect the improvement in students' vocabulary mastery (Yanti, E., & Isna, 2025).

b) Treatment

Treatment is an action, method, strategy, or intervention given to the experimental group to determine whether the treatment has an effect on the variable being studied (Yanti, E., & Isna, 2025).

In this study, the treatment was given to students in the experimental class using Blooket as a digital game-based learning medium to improve their English vocabulary mastery. Students were invited to learn vocabulary through interactive games, quizzes, and competitive activities available on the platform. Meanwhile, the control class continued to use conventional learning methods. The treatment was carried out over 4 sessions. During this process, the researcher guided students in navigating Blooket, understanding the meanings of words, and participating in games. The purpose of this treatment was to

introduce an interesting, interactive, and technology-based vocabulary learning method and to assess whether this method could improve students' vocabulary mastery more effectively than conventional methods.

c) Post-test

The post-test is administered after all learning sessions or treatments are completed to measure student learning outcomes, changes in abilities, and the effectiveness of the learning methods used. The main purpose of the post-test is to assess the extent to which the treatment, in this case the use of Blooket as a digital game-based learning medium, can improve students' vocabulary mastery compared to conventional methods. By comparing pre-test and post-test scores, researchers can objectively assess changes in student abilities and determine the significance of the treatment's influence (Yanti, E., & Isna, 2025).

In this study, the researchers compiled 20 multiple-choice questions that were given to students in the experimental and control classes as a post-test. The questions were designed to measure the vocabulary skills that had been learned during the treatment. Post-test scores were given with a criterion of 1 for correct answers and 0 for incorrect answers, so that the results

could be calculated quantitatively and compared directly with the pre-test scores.

This procedure was carried out to ensure that the data obtained could be used to evaluate the effectiveness of Blooket in improving students' vocabulary mastery. The post-test results will show the level of improvement in vocabulary mastery in the experimental class compared to the control class, allowing researchers to conclude whether the use of Blooket has a significant positive effect on students' vocabulary mastery. Thus, the post-test serves not only as a tool for measuring learning outcomes but also as an indicator of the success of treatment in the context of technology-based vocabulary learning.

2. Vocabulary

A vocabulary test is a tool used to measure the extent to which students understand and master vocabulary. It assesses their ability to recognize, comprehend, and use English words appropriately in various contexts (Susanto & Febrianto, 2022). In this study, the vocabulary test consisted of 20 multiple-choice questions in the pre-test and 20 multiple-choice questions in the post-test. These items were constructed based on indicators such as identifying word meanings, classifying word types (verbs, nouns, adjectives), and using words appropriately in

context. Each item had one correct answer, with a scoring system of one point for correct responses and zero for incorrect responses. After the test results were collected, the researcher analyzed the improvement shown between the pre-test and post-test scores using statistical calculations to determine whether the use of Blooket was effective in improving vocabulary mastery. The interpretation of students' performance followed the scoring classification proposed by J. B. Heaton (1988).

3. Documentation

Documentation is a data collection technique involving written records or documents that are relevant and supportive of the research process (Suriani, N., & Jailani, 2023). This technique allows researchers to collect factual data and evidence without directly interacting with participants, making it a reliable way to complement other data collection methods such as tests and observations. In this study, documentation included several components: the names of students in both the experimental and control groups, lists of pre-test and post-test scores, photographs taken during teaching and treatment sessions, as well as research permits, lesson plans, and other supporting documents.

These documented materials served as objective evidence to demonstrate that the research procedures,

including the implementation of Blooket in vocabulary instruction, were conducted according to the planned design. Moreover, documentation helped the researcher verify the consistency of the treatment, track student participation, and provide additional information to support the analysis of research findings. By systematically collecting and organizing these records, the researcher ensured that all stages of the study were transparent, traceable, and well-supported, which strengthens the validity and reliability of the research results (Suriani, N., & Jailani, 2023)

F. Research Instrument

This study used a vocabulary test as an instrument to collect data on students' abilities. The test consisted of 20 multiple-choice questions and was administered in two stages: pre-test and post-test. The pre-test served to assess students' initial abilities before they received treatment in the form of using the Blooket game, while the post-test was administered after treatment to see the extent of improvement in vocabulary mastery. Thus, these two tests allow researchers to compare the initial and final results, as well as evaluate the effectiveness of using Blooket in English vocabulary learning.

The validity and reliability of this vocabulary test were tested in advance to ensure that the instrument was able to measure students' abilities accurately and consistently. Only

questions that met the eligibility standards were used in the pre-test and post-test, both for the experimental group and the control group. The test scores were then categorized based on J.B. Heaton's (1988) vocabulary ability classification, so that the results could be interpreted systematically.

The determination of score intervals in this study is based on the calculation of students' vocabulary test scores by referring to the scoring criteria proposed by J. B. Heaton (1988). The score is converted into a 20-point scale to classify the students' vocabulary ability level. The formula used to determine the score is as follows:

$$\text{Score} = \frac{\text{The maximum score}}{\text{The total number of items}} \times 20$$

In this study, the maximum score obtained by the students is 58, while the total number of test items is 60. Therefore, the score calculation is presented below:

$$\begin{aligned} \text{Score} &= \frac{58}{60} \times 20 \\ &= 19.3 \end{aligned}$$

Table 3. 4 Comparison for Evaluation

The Maximum Score	Score (20-point scale)	Heaton's Scoring Categories
48	16	Good to Average
58	19.3	Excellent to Very Good
60	20	Excellent to Very Good

Based on the results of the calculations conducted, the students obtained an average score of 19.3 on a 20-point scoring scale. According to the scoring categories proposed by Heaton (1988), a score ranging from 18 to 20 is classified as *Excellent to Very Good*. This finding indicates that the students demonstrate a wide range of vocabulary with the ability to select and use words and idioms accurately and effectively. Errors in vocabulary usage are relatively minimal and do not interfere with the understanding of the intended meaning.

Therefore, it can be concluded that the level of students' vocabulary mastery in this study is categorized as high. This condition indicates that the data obtained have met the required criteria, and thus, the research is considered appropriate to proceed to the next stage of analysis or subsequent research procedures.

Table 3. 5 Score Interval for English Teaching Vocabulary

The Range of Score	Scores Category	
20-18	Excellent to Very good	Sophisticated range-effective word/idiom choice and usage etc.
17-14	Good to Average	Adequate range-occasional errors of word/idiom form, choice, usage but meaning not obscured.
13-10	Fair to Poor	Limited range-frequent errors of word/idiom form, choice, usage etc.
9-7	Very Poor	Essentially translation-little knowledge of English vocabulary

Source: J. B. Heaton (1988)

G. Data Analysis Technique

1. Validity of The Test

Before conducting the research, the researcher first examined the validity of the instruments to be used in this study. This step was conducted to ensure that each instrument was capable of measuring the aspects that were the focus of the research. Validity refers to the extent to which a research instrument actually measures what it is supposed to measure. The validity test aims to determine the extent to which the questionnaire used by the researcher is able to accurately measure and obtain research data from respondents (Susanto, F., 2025). Therefore, this process is important to ensure the accuracy and meaningfulness of the research results.

Table 3. 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: Flake et al. (2022)

In addition, researcher also analyzed the test results used computer statistics such as SPSS version 27. 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:

Table 3. 7 Instrument Validity

Item	Sig. (2-Tailed)	Compare Sig.	r table	r count	Result
D1	0,008	<0,05	0,374	0,475	Valid
D2	<0,001	<0,05	0,374	0,647	Valid
D3	0,189	<0,05	0,374	0,247	Unvalid
D4	0,093	<0,05	0,374	0,312	Unvalid
D5	<0,001	<0,05	0,374	0,649	Valid
D6	0,003	<0,05	0,374	0,520	Valid
D7	<0,001	<0,05	0,374	0,690	Valid
D8	0,021	<0,05	0,374	0,420	Valid
D9	0,013	<0,05	0,374	0,450	Valid
D10	0,065	<0,05	0,374	0,341	Unvalid
D11	<0,001	<0,05	0,374	0,655	Valid
D12	0,062	<0,05	0,374	0,345	Unvalid
D13	0,384	<0,05	0,374	0,165	Unvalid
D14	0,024	<0,05	0,374	0,410	Valid
D15	0,627	<0,05	0,374	0,092	Unvalid
D16	0,098	<0,05	0,374	0,308	Unvalid
D17	<0,001	<0,05	0,374	0,632	Valid
D18	0,146	<0,05	0,374	0,272	Unvalid
D19	0,060	<0,05	0,374	0,347	Unvalid
D20	<0,001	<0,05	0,374	0,778	Valid
D21	0,030	<0,05	0,374	0,397	Valid
D22	0,079	<0,05	0,374	0,325	Unvalid
D23	0,025	<0,05	0,374	0,408	Valid
D24	0,023	<0,05	0,374	0,414	Valid
D25	0,052	<0,05	0,374	0,357	Unvalid
D26	0,076	<0,05	0,374	0,329	Unvalid
D27	0,040	<0,05	0,374	0,377	Valid
D28	0,076	<0,05	0,374	0,329	Unvalid
D29	0,001	<0,05	0,374	0,560	Valid
D30	0,200	<0,05	0,374	0,241	Unvalid
D31	<0,001	<0,05	0,374	0,727	Valid
D32	0,003	<0,05	0,374	0,526	Valid
D33	0,378	<0,05	0,374	0,167	Unvalid
D34	<0,001	<0,05	0,374	0,586	Valid
D35	0,052	<0,05	0,374	0,358	Unvalid
D36	<0,001	<0,05	0,374	0,665	Valid

D37	0,304	<0,05	0,374	0,194	Unvalid
D38	<0,001	<0,05	0,374	0,674	Valid
D39	0,016	<0,05	0,374	0,436	Valid
D40	<0,001	<0,05	0,374	0,597	Valid
D41	0,002	<0,05	0,374	0,549	Valid
D42	0,840	<0,05	0,374	-0,039	Unvalid
D43	0,006	<0,05	0,374	0,492	Valid
D44	0,009	<0,05	0,374	0,471	Valid
D45	0,004	<0,05	0,374	0,512	Valid
D46	0,458	<0,05	0,374	0,141	Unvalid
D47	0,005	<0,05	0,374	0,499	Valid
D48	0,007	<0,05	0,374	0,480	Valid
D49	0,033	<0,05	0,374	0,390	Valid
D50	0,236	<0,05	0,374	0,223	Unvalid
D51	<0,001	<0,05	0,374	0,633	Valid
D52	0,004	<0,05	0,374	0,514	Valid
D53	<0,001	<0,05	0,374	0,672	Valid
D54	0,650	<0,05	0,374	-0,086	Unvalid
D55	0,739	<0,05	0,374	0,063	Unvalid
D56	0,236	<0,05	0,374	0,223	Unvalid
D57	0,762	<0,05	0,374	0,058	Unvalid
D58	0,193	<0,05	0,374	0,244	Unvalid
D59	<0,001	<0,05	0,374	0,584	Valid
D60	0,004	<0,05	0,374	0,512	Valid

Source: Sugiono (2013)

In this validity test, researcher used 60 multiple choice questions with predetermined indicators, and the test of this question instrument was carried out at SMP Negeri 20 Kota Bengkulu which of course has an accreditation level equivalent to the school that will be the main research location, namely at SMP Negeri 5 Kota Bengkulu. 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 validity of each question. From the results of the analysis, it was found that

30 of the 60 questions tested were declared valid with the following indicators:

Table 3. 8 Indicators of Vocabulary Validity Test

No.	Test Item	Test Indicator	Type
1.	The word “garbage” means ... A. Air B. Waste C. River D. Leaf	Students can define nouns related to environmental waste.	Noun
2.	The act of using waste again is ... A. Pollution B. Flood C. Recycle D. Litter	Students can define verbs related to recycling activities.	Verb
3.	Too much water on land causes ... A. Drought B. Litter C. Flood D. Smoke	Students can identify environmental problems caused by natural events.	Noun
4.	Pollution is dangerous because it makes the environment ... A. Healthy B. Dirty C. Beautiful D. Safe	Students can define the meaning of pollution in English.	Adjective
5.	Smoke from factories causes air ... A. Waste B. Recycle C. Pollution D. Flood	Students can identify types of pollution based on their sources.	Noun
6.	The correct action for trash is ... A. Throwing B. Burning	Students can identify correct actions in managing trash.	Verb

	C. Recycling D. Littering		
7.	In environmental problems, cause means ... A. Result B. Reason C. Effect D. Solution	Students can define the meaning of cause in environmental context.	Noun
8.	Which is an environmental event... A. Flood B. Thousand C. Plastic D. Trash	Students can identify environmental events in English.	Noun
9.	Which word best describes a place that is not clean... A. Fresh B. Dirty C. Healthy D. Safe	Students can define adjectives describing environmental conditions.	Adjective
10.	Dirty rivers show water ... A. Recycle B. Pollution C. Flood D. Clean	Students can identify water pollution based on its condition.	Noun
11.	Pollution makes air and water become _____. A. Dirty B. Safe C. Fresh D. Clean	Students can identify the effects of pollution on air and water.	Adjective
12.	Using bottles again is called _____. A. Reduce B. Protect C. Recycle D. Save	Students can define verbs related to recycling activities.	Noun

13.	Too much plastic is _____ for the environment. A. Helpful B. Useful C. Safe D. Harmful	Students can identify harmful materials for the environment.	Adjective
14.	Littering is a bad _____. A. Solution B. Habit C. Event D. Action	Students can define negative habits related to the environment.	Noun
15.	A clean environment supports human _____. A. Pollution B. Flood C. Health D. Trash	Students can identify the benefits of a clean environment for humans.	Noun
16.	Dirty water can cause many _____. A. Benefits B. Actions C. Solutions D. Problems	Students can identify problems caused by dirty water.	Noun
17.	Clean-up activities are done to _____ the area. A. Damage B. Clean C. Destroy D. Dirty	Students can identify the purpose of clean-up activities.	Verb
18.	Every environmental problem needs a _____. A. Cause B. Solution C. Event D. Action	Students can define the meaning of solution in environmental problems.	Noun

19.	The correct word for picture is ... A. Recycle B. Trash C. Environment D. Campaign		Students can identify vocabulary related to environmental pictures.	Noun
20.	Picture is related to ... A. Littering B. Recycling C. Pollution D. Flooding		Students can identify environmental activities shown in pictures.	Verb
21.	Picture shows a dirty ... A. Classroom B. Gutter C. River D. Bin		Students can identify polluted places shown in pictures.	Noun
22.	Pollution makes the environment ... A. Unhealthy B. Fresh C. Safe D. Clean		Students can identify the effects of pollution on environmental health.	Adjective
23.	Bin is associated with... A. Water B. Trees C. Trash D. Animals		Students can identify objects related to waste management.	Noun
24.	Solution is needed to fix a ... A. Problem B. Trash C. Thousand D. Event		Students can define the meaning of problem in environmental context.	Noun
25.	Garbage has the same meaning as ... A. Water B. Soil C. Trash		Students can identify synonyms of environmental vocabulary.	Noun

	D. Air		
26.	Conserve is closest in meaning to ... A. Save B. Destroy C. Pollute D. Waste	Students can define verbs related to environmental conservation.	Verb
27.	Environment includes air, water, and A. Books B. Land C. Desks D. Pens	Students can identify elements of the environment in English.	Noun
28.	Burning plastic should be avoided because it causes _____. A. Clean air B. Pollution C. Solution D. Protection	Students can identify the effects of burning plastic waste.	Noun
29.	Using reusable bags helps to _____ plastic waste. A. Reduce B. Increase C. Throw D. Block	Students can identify actions that reduce plastic waste.	Verb
30.	Flood often happens during _____ rain. A. Light B. Dry C. Short D. Heavy	Students can identify the causes of flood.	Adjective
31.	Environmental problems affect human _____. A. Pollution B. Health C. Trash D. Thousand	Students can identify the effects of environmental problems on humans.	Noun

32.	Recycling is a good _____ for waste problems. A. Cause B. Solution C. Effect D. Event	Students can identify solutions to waste problems.	Noun
33.	Students join clean-up activities to keep school _____. A. Clean B. Dirty C. Dangerous D. Polluted	Students can identify the purpose of school clean-up activities.	Adjective
34.	Students recycle paper to _____ trees. A. Pollute B. Protect C. Damage D. Flood	Students can identify the benefits of recycling for the environment.	Verb
35.	Trash in rivers can cause _____. A. Flood B. Clean C. Recycling D. Campaign	Students can identify the effects of trash in rivers.	Noun

2. Reliability of The Test

After the questions have been tested and validated, the next step in this research is measuring the reliability of the instrument to ensure consistent measurement results. reliability refers to how consistent the results of the research are when repeated in the same way. Reliability test aims to determine the level of consistency of the questionnaire used by researchers. So that the questionnaire can be relied upon even though the research was carried out

repeatedly with the same questionnaire at different times. The reliability test was carried out after the item in the questionnaire was declared valid (Susanto, F., 2025).

To assess reliability, Cronbach's Alpha value is compared with the r-table. If $\alpha > r\text{-table}$, the instrument is considered reliable. This test ensures that the questions are consistent with each other and do not produce random data, so that the research data can be used for further analysis (Setyaedhi, 2024).

Table 3.9 Test Instrument Reliability Criteria

No	Cronbach's Alpha	Reliability Level
1.	$-1,00 < 0,50$	Poor Reliability
2.	$0,50 - 0,75$	Medium Reliability
3.	$0,75 - 0,90$	Good Reliability
4.	$1,00 > 0,90$	Very High Reliability

Source : Setyaedhi (2024)

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 3. 10 Instrument Reliability

Reliability Statistics	
Cronbach's Alpha	N of items
0,924	60
Very Highly Reliable	

In this reliability test, researcher used 60 multiple choice questions with predetermined indicators, and the test of this question instrument was carried out at SMP Negeri 20 Kota Bengkulu which certainly had an accreditation level equivalent to the school that would be the main research location, namely at SMP Negeri 5 Kota Bengkulu. 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 0,924 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 (Setyaedhi, 2024). Therefore, it can be assumed that this test has a good level of reliability.

3. The Level Difficulty Test

The level difficulty test is a measure used to assess the level of difficulty of a question based on how many participants can answer correctly. The higher the proportion of participants who answer correctly, the easier the question is; conversely, the lower the proportion of correct answers, the more difficult the question is. This assessment is important to ensure that the questions given through Blooket have a balanced and fair level of difficulty,

so that students' abilities can be measured accurately and objectively (Fauzie, M., et.al., 2021).

In this study, a level difficulty test was conducted to evaluate students' vocabulary mastery through the Blooket game. This analysis aimed to determine whether each question was appropriate for the students' abilities and could differentiate the level of vocabulary mastery among participants. Data were obtained from the results of student participation in the Blooket game, then analyzed to identify questions that fell into the easy, medium, or difficult categories. To calculate the level of difficulty, the following formula was used:

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

Description:

P : Difficulty level in percent

Nh : Number of test takers in high score group

Rh : Number of correct answers in high score group

Ni : Number of test takers in low score group

Ri : Number of correct answers in low score group

$$\begin{aligned} P &= \frac{442 + 92}{8 + 8} \times 100\% \\ &= 66,75\% \end{aligned}$$

Table 3. 11 Interpretation of Difficult Question

Categories

Question category	Interpretation of result
Classified as easy	0.71% ----- 100%
Classified as medium	0.31% ----- 0.70%
Classified as difficult	0% ----- 0.30%

Source: Fauzie et al. (2021)

In testing research questions, it was important to 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 66,75%. 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 Discrimination Test

Item discrimination analysis is important to assess whether test items effectively differentiate students based on their abilities. According to Karim (2021), by analyzing test items including examining difficulty level and discriminating power it is possible to determine which items are of high quality and which items need to be revised or removed.

Item discriminating power (DP) measures how well an item can distinguish between students with high ability and

those with low ability. Items with high DP are typically answered correctly by high-performing students and incorrectly by low-performing students, indicating that the item is sensitive to differences in ability (Karim, S. A., et. al., 2021)

Zulaiha (2021) noted that the format of the test also affects discriminating power. Their analysis of multiple-choice and essay items in a mathematics test showed that multiple-choice items tend to have higher DP compared to essay items. Therefore, the choice of item format is important in test design. The general formula to calculate item discriminating power is:

$$DP = \frac{U-L}{N} \times 100\%$$

Description:

DP : Question discriminating 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 student in each group

$$DP = \frac{442-92}{16} \times 100\%$$

$$= 21,875\%$$

Table 3. 12 Interpretation of The Power Difference Test

T-test	Interpretation of results
$0.71 \leq DP \leq 1.00$	Very Good
$0.41 \leq DP \leq 0.70$	Good
If $+ 0.21 < DP < 0.40$	Sufficiently qualified
If $0 < DP \leq 0.20$	Weak

Source : Rusijono (2026)

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 21,875%. 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 test is a statistical procedure conducted to assess the distribution of data within a dataset or variable, specifically to determine whether the data follow a normal distribution. The criteria for interpreting the results of a normality test are as follows:

- If the significance value (P Value) is less than 0.05, it indicates that the data are not normally distributed.
- If the significance value (P Value) is greater than 0.05, it indicates that the data are normally distributed.

6. Homogeneity Test

A statistical test procedure that aims to show that two or more sample data groups are taken from populations that have the same variance. In other words, this test is conducted to determine whether the data sets being studied have the same characteristics or not. Use the following principles to determine homogeneity:

- a) Significance test (α) = 0.05
- b) If $\text{sig} > \alpha$, then the variance of each sample is the same (homogeneity).
- c) If $\text{sig} < \alpha$, then the variance of each sample is not the same (not homogeneity). To determine this, the researcher uses SPSS.

7. T-Test

In this study, the T-test was used to determine whether there was a significant difference in students' vocabulary comprehension between the experimental class that used the Blooket application and the control class that used conventional methods. Data analysis was performed using the T-test in the SPSS program. This test aims to assess the magnitude of the partial effect of the independent variable on the dependent variable, assuming that other independent variables remain constant.

The testing criteria used are as follows:

- a) If the significance value is $< \alpha$ (0.05), then the independent variables individually have an effect on the dependent variable.
- b) If the significance value is $> \alpha$ (0.05), then the independent variables individually have no effect on the dependent variable.

8. Statistical Hypothesis

Hypothesis testing serves to test whether the null hypothesis (H_0) is accepted or rejected and the alternative hypothesis (H_a) is accepted or rejected. Hypothesis testing can be carried out with the following basic guidelines for obtaining test results:

- a) If the Sig. (2-tailed) value is < 0.05 , then the hypothesis is accepted.
- b) If the Sig. (2-tailed) value > 0.05 , then the hypothesis is rejected.