CHAPTER III

METHODOLOGY

A. Research Design

This study used an experimental methodology. The purpose of the study was to find out how well Blooket functioned as a learning media website and whether it had any influence on students' vocabulary mastery. Additionally, the quantitative research had the following traits Sawilowsky, Kelley, Blair & Markman 1994).

This study employed a true-experimental design with a pretest-posttest control group approach to assess the effect of the Blooket application on vocabulary mastery among eighth-grade students at SMP Islam Al-Azhar 52 Bengkulu City. The research utilized the Solomon Four Group Design, incorporating both pretests and posttests for the experimental and control groups to evaluate changes. Focused on Solomon's model, the study extended the traditional control group design into a four-group (and occasionally three-group) format, introducing additional randomization to assign participants within both the experimental and control groups to either undergo pre-testing or not (Solomon, 1949; McCambridge, Butor-Bhavsar, Witton, & Elbourne,

2011). Afterwards, the researcher observed the result through test. The study design could be seen as follow:

Table 2. Solomon Four group design

Group	Pre-test	Treatment	Post-test
A	O_1	X	O_2
B	O_3	X_0	O_4
C	O	X	O_5
D	O	X_0	O_6

Note:

O : No Pre Test Experimental Group

O_1 : Pre Test Experimental Group

O_2 : Post Test of Experimental Group

O_3 : Pre Test of Control Group

O_4 : Post Test of Control Group

O_5 : Post Test of Control Group

O_6 : Post Test of Control Group

X : Treatment

X_0 : No Treatment

B. Place and Time of Research

This study was conducted at Jl. Pariwisata, Timur Indah, Kec. Singaran Pati, Kota Bengkulu, Bengkulu. The researcher carried out the study to examine the impact of Blooket on students' vocabulary mastery between December 2024 and January 2025..

C. Research Variable

This study involved two types of variables: the dependent variable, which was influenced by other variables, and the independent variable, which did not depend on other variables but instead influenced them. The variables examined in this study included:

1. Vocabulary Mastery (Y): A strong vocabulary was fundamental for academic success and held even greater importance in the development of language and literacy skills (Language et al., 2024).
2. Blooket Application (X): A gamified learning platform was used, enabling teachers to design interactive games with sets of questions. Students engaged in these games by answering questions on their personal devices. The platform was mainly geared towards formative assessment, offering real-time feedback to students as they progressed. Correct answers earned students points, which could be spent on acquiring and exchanging in-game items known as Blooks (Barokah et al., 2024).

D. Population and Sample

The study selected eighth-grade students at SMP Islam Al-Azhar 52 Bengkulu City as the population. For the sample, the researcher chose 48 students, dividing them into four groups A, B, C, and D with 12 students in

each group. These groups were chosen after the researcher had obtained permission from the school to conduct the study and had discussed the research with Mr. Eki Syaputra, S.Pd., the English teacher for the eighth grade.

1. Population

The population referred to the general scope, consisting of objects or subjects with specific qualities and characteristics defined by the researcher for analysis, from which conclusions about the study's findings were drawn. In this study, the population included all eighth-grade students at SMP Islam Al-Azhar 52 Bengkulu City for the 2024/2025 academic year, divided into four classes: A, B, C, and D. The total population is presented in the table below.

Table 3. Total Students

No	Class	Gender		Total
		Female	Male	
1.	VIII. A	1O	21	31
2.	VIII B	16	13	29
3.	VIII. C	12	16	28
4.	VIII. D	9	19	28
	Total			116

2. Sample

The sample represented a subset of the population in terms of both number and characteristics. In this true-experimental study, the researcher employed the Solomon Four Group

Design. Students from the eighth-grade classes at SMP Islam Al-Azhar 52 Bengkulu City were selected as the research sample. After considering several factors, the researcher chose representatives from each class to participate in the study. A total of 48 students from grade VIII were selected and divided into four groups A, B, C, and D with 12 students in each group.

Table 4. total students in the Solomon Four group

No.	Group	Gender		Total
		Female	Male	
1.	A	9	3	12
2.	B	2	10	12
3.	C	3	9	12
4.	D	12	-	12
Total				48

E. Research instrument

The instrument used in this study was designed according to the selected data ng of 60 questions on adverbs, synonyms, antonyms, nouns, adjectives, and conjunctions at the eighth-grade junior high school level to assess students' vocabulary mastery. After conducting a trial test and calculating its validity, 21 questions were found to be valid.

F. Data Collection Technique

To analyzed the data the researcher employed the formula as follows:

1. Pre- Test

Prior to the treatment, a pre-test was administered to assess the students' initial vocabulary proficiency The test consisted of multiple-choice questions (Significance of Solomon Four Group Pretest-Posttest Method in True Experimental Research- A Study, 2013).

2. Treatment (Teaching Research Activities)

After the pre-test was administered, the treatment was conducted using a game-based strategy during the teaching and learning process. The procedure for utilizing the game to teach vocabulary was carried out through the following activities:

In the pre-teaching activities, the teacher began the lesson by greeting the students and checking attendance. To activate prior knowledge, the teacher posed several questions related to the upcoming vocabulary topic (Sari, 2021). Then, the teacher introduced the Blooket game and explained how it would be used to support vocabulary

learning.Students were instructed to download the Blooket application on their devices. Once the app was installed, the teacher provided clear instructions on how to play the game. The students listened attentively and analyzed the gameplay. With great enthusiasm, they began participating in the Blooket game.

3. **Post-test**

The post-test was administered to evaluate the extent of changes and improvements in students' vocabulary after the implementation of the Blooket application as an instructional tool (Chu, PH. and Chang, 2017)s. The main purpose of the post-test was to determine whether there was a significant improvement in vocabulary scores compared to the pre-test results. After the post-test was administered, the teacher conducted a thorough review of the Blooket game to assess its effectiveness in enhancing student learning. Following the review, individual evaluations were provided for each student. The students' correct responses from both the pre-test and post-test were carefully recorded and analyzed. The results were then categorized based on the students' score ranges in both assessments.:

Score 96 – 100 : Excellent

Score 86 – 95 : Very Good

Score 76 – 85 : Good

Score 66 – 75 : Fairly Good

Score 56 – 65 : Fairly

Score 36 – 55 : Poor

Score 0 – 35 : Very Poor

(Depdikbud, 1985:6)

Table 5. Score Classification

Score	Classification	Indicator
96-100	Excellent	Natural English with minimal errors means using English that sounds natural and follows proper grammar. Complete realization of the task set refers to completing the task thoroughly without leaving anything out. The goal is to produce clear, accurate, and context-appropriate communication.
86-95	Very Good	Using good vocabulary and structure means employing a range of varied and precise words, along with well organized sentences that go beyond basic simplicity. This involves crafting sentences with complexity, such as compound or complex structures, while ensuring that errors are non basic that is, not fundamental mistakes in grammar, spelling, or usage. The aim is to communicate ideas effectively and professionally.
76-85	Good	A simple but accurate realization of the task focuses on clear and straightforward execution without unnecessary complexity.
66-75	Fairly Good	Naturalness in expression is achieved with minimal errors, ensuring the communication is effective and easy to

			understand
	56-65	Fairly	Reasonably correct if awkward shows limited vocabulary, leading to clunky phrasing, while natural treatment with serious errors suggests stronger vocabulary but poor grammar or usage undemines clarity.
	36-55	Poor	When vocabulary and grammar are inadequate for the task, it means the word choice is too limited or imprecise, and grammatical issues hinder clear communication. This can result in incomplete or unclear responses that fail to meet the task's requirements.
	0-35	Very Poor	Incoherent with errors showing a lack of basic knowledge of English means that the message is difficult to understand due to major mistakes in grammar, vocabulary, and sentence structure. These errors suggest a fundamental misunderstanding of the language.

G. Technique of Data Analysis

This study utilized the pre-test and post-test results from the Solomon Four Group design in the data analysis. The objective was to determine whether the implementation of the Blooket application had an effect on vocabulary acquisition among eighth-grade students at SMP Islam Al-Azhar 52 Bengkulu City.

1. Documentation

Documentation was used to collect research data. One component of the documentation in this study was photography. The purpose of this documentation was to gather information from

various sources to support the research process (Lammers & Badia, 1982). Photographs of instructional activities and other relevant information were used as documentation to support the research. Additionally, the documentation included a list of the names of students from eighth grade at SMP Islam Al-Azhar 52 Bengkulu City, who were assigned to the Solomon Four Groups. A total of 48 students were divided into groups A, B, C, and D, with each group consisting of 12 students. The documentation also contained a list of students' grades prior to the research and a list of their grades after the implementation of the Blooket game.

2. Validity Test

Validity referred to the extent to which a test accurately measured the specific concept it intended to assess and allowed for meaningful interpretation of the resulting scores. To ensure the reliability of the research findings, it was essential to conduct a validity test. This evaluation determined whether the selected measurement tools were appropriate for the research objectives (sari, 2021). The outcome of this test revealed the level of validity whether it was high, moderate, or sufficient for utilizing the specific instrument in the research. The data were

processed with the help of computer facilities using SPSS version 28.

Table 6. Instrumen validy

Item	r Table	Significance	r count	Result
1	0.361	0,758	-0,059	Unvalid
2	0.361	0,001	.568**	Valid
3	0.361	0,992	0,002	Unvalid
4	0.361	0,799	-0,049	Unvalid
5	0.361	0,010	.465**	Valid
6	0.361	0,700	-0,073	Unvalid
7	0.361	0,051	0,359	Unvalid
8	0.361	0,001	.576**	Valid
9	0.361	0,291	0,199	Unvalid
10	0.361	0,003	.528**	Valid
11	0.361	0,204	0,239	Unvalid
12	0.361	0,093	0,312	Unvalid
13	0.361	0,008	.475**	Valid
14	0.361	0,000	.617**	Valid
15	0.361	0,000	.617**	Valid
16	0.361	0,499	-0,128	Unvalid
17	0.361	0,004	.505**	Valid
18	0.361	0,062	0,345	Unvalid
19	0.361	0,086	0,318	Unvalid
20	0.361	0,071	0,334	Unvalid
21	0.361	0,800	0,048	Unvalid
22	0.361	0,013	.449*	Valid
23	0.361	0,008	.474**	Valid
24	0.361	0,602	0,099	Unvalid
25	0.361	0,248	0,218	Unvalid
26	0.361	0,379	0,167	Unvalid
27	0.361	0,010	.460*	Valid
28	0.361	0,020	.423*	Valid
29	0.361	0,720	0,068	Unvalid
30	0.361	0,008	.476**	Valid
31	0.361	0,002	.553**	Valid
32	0.361	0,082	0,323	Unvalid
33	0.361	0,329	0,184	Unvalid
34	0.361	0,217	0,232	Unvalid

35	0.361	0,381	0,166	Unvalid
36	0.361	0,189	0,247	Unvalid
37	0.361	0,000	.658**	Valid
38	0.361	0,182	0,250	Unvalid
39	0.361	0,116	0,293	Unvalid
40	0.361	0,057	0,351	Unvalid
41	0.361	0,137	0,278	Unvalid
42	0.361	0,088	0,317	Unvalid
43	0.361	0,122	0,288	Unvalid
44	0.361	0,005	.501**	Valid
45	0.361	0,009	.470**	Valid
46	0.361	0,260	0,212	Unvalid
47	0.361	0,087	0,318	Unvalid
48	0.361	0,079	0,326	Unvalid
49	0.361	0,131	0,282	Unvalid
50	0.361	0,059	0,348	Unvalid
51	0.361	0,914	-0,020	Unvalid
52	0.361	0,417	-0,154	Unvalid
53	0.361	0,545	0,115	Unvalid
54	0.361	0,972	-0,007	Unvalid
55	0.361	0,186	0,248	Unvalid
56	0.361	0,045	.368*	Valid
57	0.361	0,284	0,202	Unvalid
58	0.361	0,006	.487**	Valid
59	0.361	0,000	.688**	Valid
60	0.361	0,037	.382*	Valid

In this study, the validity test was conducted using 60 multiple-choice questions. The researcher performed a trial at a junior high school of the same grade level as the school where the actual research was carried out. The trial involved 30 students as participants. After completing the trial, the test data were analyzed by calculating the validity of each question. The analysis results revealed that, out of the

60 questions tested, 21 were considered valid based on the following indicators:

Table 7. Indicators of Vocabulary Mastery Test Validity

No	Indicators	No Item	Items
1	Identifying word meanings in various contexts, particularly focusing on adverbs and their usage in sentences	4	2,5,8,10
2	Understanding synonyms by recognizing words with similar meanings and differentiating subtle differences	4	13,14,15,17
3	Recognizing antonyms by identifying words with opposite meanings and understanding their contextual usage	5	22,23,27,28,30
4	Differentiating noun forms, including singular and plural nouns, as well as distinguishing proper and common nouns	2	31,37
5	Identifying adjective characteristics, such as descriptive, comparative, and superlative forms, and their application in sentences	2	44,45
6	Using conjunctions correctly to connect clauses, phrases, or sentences while maintaining grammatical coherence	4	56,58,59,60
	Total	21	

3. Reliability Test

The tool was deemed a reliable data collection instrument based on the results of the

reliability analysis. This outcome indicated that the instrument had the necessary qualities to be utilized in the study. Cronbach's Alpha was employed in this study to assess the reliability of the measurement tool. By evaluating the correlation between different items within a single measurement, this statistical method examined the internal consistency of the instrument (Gülen Ertosun et al., 2015). The data is processed with the help of computer facilities using SPSS versionIn this reliability test, the researcher used multiple-choice questions consisting of 60 items. A trial was conducted at a junior high school of the same grade level as the school where the research was to be carried out. The trial involved 30 students as test subjects. After conducting the trial, the researcher processed the test data by calculating its reliability. The analysis yielded a reliability coefficient of 0.828. An item was considered reliable if it reached a value of 0.600 or higher. Therefore, it could be concluded that the test demonstrated a high level of reliability.

Table 8. Case Processing Summary

Case Processing Summary			
		N	%
Cases	Valid	30	100.0
	Excluded[a]	0	0.0
	Total	30	100.0

Table 9. The Result of Reliability

Reliability Statistics	
Cronbach's Alpha	N of Items
0.828	60

Based on the table above, it could be concluded that the set of 60 questions, tested on 30 students, which resulted in 21 valid questions, was appropriate for use in the pre-test and post-test conducted by the researcher. This conclusion was further supported by the Cronbach's Alpha result, which reached 0.828. This value of Cronbach's Alpha can be interpreted as follows: (Gülen Ertosun et al., 2015)

Table 10. Crobanch's Alpha Interpretation

Cronbach's Alpha	Interpretation
0.11-0.44	Low
0.45-0.65	Acceptable
0.58-0.68	Slightly Low
0.70-0.77	Fairly High
0.80	Robust
0.81-0.90	Reliable
0.91-0.93	Strong
0.93-0.94	Excellent

4. Item Difficulty Test

To assess how many respondents correctly answered each item, an item difficulty test was conducted. The item difficulty test was a crucial

component in the development of research instruments (Dare et al., 2020). The process of testing the difficulty level of the items involved administering 60 questions to a total of 30 eighth-grade students at MTSN 2 Bengkulu City, the school where the researcher carried out the instrument trial. The item difficulty level was determined by dividing the number of students who correctly answered a particular item by the total number of students who participated in the trial (Susanto, F., & Jaya, 2023). This calculation used the following formula:

$$p = \frac{Rh + Ri}{Nh + Ni} x\ 100\%$$

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 test takers in the low score group

Ri: Number of correct answers in the low score group

$$p = \frac{328}{8} + \frac{158}{8} x\ 100\%$$

$$p = 60{,}75\%$$

Based on the results of the calculations above, it could be concluded that the difficulty level of each item, which had been tested on 30 students using a total of 60 questions, was categorized as medium according to the

interpretation of test item difficulty levels (Mamarimbing et al., 2015).

Table 11. Interpretation of Item Difficulty Test

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

Source : (Mamarimbing et al., 2015)

5. Differentiability Test

The potential of a question to differentiate test-takers based on their scores was referred to as its distinguishing power. This analysis involved categorizing test-takers into high and low groups. In other words, the greater the distinguishing power of a question, the more likely it was that students from the high group could answer it correctly, while fewer students from the low group were able to do so (Hanifah, 2014).

The process of testing the differential power of the items was carried out by administering 60 questions to a total of 30 eighth-grade students at MTSN 2 Bengkulu City, the school where the researcher conducted the instrument trial. Items with a differentiation index greater than 0.30 were classified as good, while those with an index below 0.30 were

considered poor. In calculating the differential power, the researcher used Microsoft Excel to obtain the index. The following formula was used to calculate the differential power in this test:

$$DP = \frac{BA}{JA} - \frac{BB}{JB} : \text{NM}$$

$$DP = \frac{328}{8} - \frac{158}{8} : 60$$

$$DP = 0{,}355$$

Description:

DP: Question Differentiation 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 high score group test takers

NM : Score maximum

Table12. Interpretation of Item Differentiability Test

Differentiability Test	Interpretation
-1,00 ---------- 0,19	Poor
0,20 ---------- 0,29	Enough
0,30 ---------- 0,39	Good
0,40 ---------- 1,00	Very Good

Negative, all of them	Not good, so all items that have a negative discriminating power value should be discarded.

Source : (Dali, 1995)

Based the results of the Distinguishing Power (DP) calculation findings, the study's DP is 0.35 when compared to the properly qualified category. That is, if 0, 21 < DP (0, 33) then the study's DP is considered sufficiently qualified.

6. Normality Test

A normality test was necessary to assess whether the sample met the criteria for being representative, allowing the research findings to be generalized to the broader population or to accurately reflect it (Qurnia Sari et al., 2017) . his test utilized SPSS to check the normality of the pre-test and post-test data. If the data were found to be normally distributed, parametric statistical tests, such as the T-test, could then be applied. In general, the data were analyzed with the help of SPSS version 28 using computer-based tools.

7. Homogeneity Test

The homogeneity test is used to determine whether some population variants are the same or not

(Usmadi, 2020).The following criteria are used for the homogeneity analysis:

a. Significance value (α) = 0.05
b. If sig > α, the data is considered homogeneous (same variance).
c. If sig < α, the data is not homogeneous (different variance).

Overall the data computing out with aid of computer facilities SPSS ver 28.

8. Paired T-Test

The paired t-test is a hypothesis testing method used when the data are not independent, but paired (Montolalu & Langi, 2018). In this study, the paired t-test was employed to analyze the significant differences between groups A and B before and after the treatment. The following criteria were applied in the t-test analysis.:

a. If the significance value (2-tailed) > 0.05, there is no significant difference between the initial variable (pretest) and the final variable (posttest).
b. If the significance value (2-tailed) ≤ 0.05, there is a significant difference between the initial variable (pretest) and the final variable (posttest).

Over all the data computing out with aid of computer facilities SPSS ver 28.

9. T-Test Independent

The independent sample t-test is a parametric statistical method used to compare groups and determine if there is significant evidence that the population means differ statistically (Soeprajogo; Purnama & Ratnaningsih, 2020). In this study, it was used to compare the post-test scores of group C and group D, which represented the experimental and control groups, respectively. This test helped identify whether the Blooket game model had a significant impact on students' vocabulary mastery. The following criteria were applied in the t-test analysis:

a. If the significance value (2-tailed) > 0.05, then there is no significant difference between the learning outcomes using the flipped classroom model in learning reading ability in group C and D.
b. If the significance value (2-tailed) ≤ 0.05, there is a significant difference between learning outcomes using the flipped classroom model in learning reading ability in group C and D.

Over all the data computing out with aid of computer facilities SPSS ver 28.

10. Two Way-ANOVA Test

Anova or analysis of variance is a test that can be used to analyze differences in more than two independent groups (Rahmawati & Erina, 2020). The aim is to compare more than two averages and is useful for testing generalizability, meaning that sample data is considered to represent the population (Riduwan, 2010). ANOVA is a test that can be used to analyze differences in more than 2 populations of independent groups. ANOVA is actually a more generalized form of the t-test that is appropriate for use with three or more groups-it can also be used with two groups (Gu, 2014). The following criteria are applied for the two way-ANOVA test:

a. If the significance value is > 0.05, then the factors used in the research have no effect on the observed parameter.
b. If the significance value (2-tailed) ≤ 0.05, then the factors used in the research) have an effect on the observed parameters.