CHAPTER III
RESEARCH METHOD

A. Research Design

This research employs a quantitative approach, which focuses on the thorough collection of data and involves a wide population scope. Quantitative research is known for its systematic, organized, and carefully structured process, beginning from the initial planning stage through to the research design. The approach used in this research is true-experimental (Andi, 2003).

True experimental is a type of experiment that is completely controlled, where the researcher can control all additional variables that could potentially affect the results of the research. This ensures high internal validity, so that the quality of the research implementation can be well maintained. The main feature of true experimental is the random selection of samples from a specific population, both for the experimental and control groups (Hikmawati, 2017). As an alternative to this experimental design, there is Solomon's four-group design. This design offers greater advantages in certain aspects, such as effectiveness in controlling for pre-test effects (Phan and Ngu, 2017).

This design has four groups where two of them are given a pre-test and post-test (O) while the remaining two groups each only get a pretest or posttest. The procedure is to divide the subjects into four groups, where from the four groups are divided into two groups where each of the groups will receive treatment and there is only a control group. The experimental group receives a certain treatment, while the control group does not receive any treatment. This experiment uses a randomized solomon four group design model (Şahin and Kılıç, 2024). The experimental design of the feasibility test for the application of the flipped classroom is as follows:

Table. 1. Design Method of Solomon Four Group

Group	Pre-Test	Treatment	Post-Test
A	O1	X	O2
B	O3	-	O4
C	O	X	O5
D	O	-	O6

Description:

X: Applying the Flipped Classroom Model

O: Tidak diberikan Pre-Test

-: Tidak diberikan Treatment

O1: Pre-Test on Experimental Group

O2: Post-Test on Experimental Group

O3: Pre-Test on Control Group

O4: Post-Test on Control Group

O5: Post Test on Experimental Group

O6: Post Test on Control Group

B. Place and Time of The Research

This research was carried out at SMAN 1 Bengkulu City from February to March 2025. The research involved students from class XI 2 and class XI 4, each comprising 36 students

C. Population and Sample

1. Population

Population refers to the general scope that consists of objects or subjects with certain qualities and characteristics defined by the researcher for analysis, from which the researcher can draw conclusions based on the research findings (Azizah, 2018). In this research, the population consists of 11th-grade students at SMAN 1 Kota Bengkulu for the 2024/2025 academic year, divided into 12 classes.

Table 2. Population of the research

No	Class	Female	Male	Total
1.	XI 1	17	18	35
2.	XI 2	18	18	36
3.	XI 3	22	14	36
4.	XI 4	21	15	36
5.	XI 5	22	14	36
6.	XI 6	17	19	36
7.	XI 7	25	11	36
8.	XI 8	20	16	36
9.	XI 9	18	18	36
10.	XI 10	16	19	35

11.	XI 11	16	20	36
12.	XI 12	17	18	35
Total				429

2. Sample

A sample represents a portion of the population's number and characteristics. In this true-experimental research, the technique used is randomize sampling (Azizah, 2018). The researcher selected two classes of 11th-grade students from SMAN 1 Kota Bengkulu as the sample for this research. After considering several factors, the researcher chose representatives from each class as participants for the research. Class XI 4, with 36 students, will serve as the experimental group, while Class XI 2, also with 36 students, will be the control group.

Table 3. Sample of the research

No	Class	Classroom	Female	Male	Total
1.	Experimental group	XI 4	21	15	36
2.	Control group	XI 2	18	18	36
Total			**39**	**33**	**72**

D. Research Variables

This research involves two types of variables: the dependent variable, which is affected by other variables, and the independent variable, which impacts other variables. The variables in this research are as follows:

1. X (Independent Variable): Flipped Classroom Model (a teaching model where students are introduced to content outside the classroom, typically through online videos, and class time is used for active learning activities such as discussions and problem-solving).
2. Y (Dependent Variable): Reading Ability (Reading ability refers to the capacity to understand, interpret, and engage with written text. It involves skills such as comprehension, vocabulary, and the ability to critically analyze and extract meaning from different types of texts).

E. Data Collection Technique

The researcher uses pre-test and post-test methods to collect data for this research. Based on (Ganesha and Aithal, 2022) the following are the steps involved in the data collection process:

1. Pre-test

The pre-test is designed to assess students' initial english language skills before they participate in learning activities using the flipped classroom model. The researcher uses the pre-test to evaluate the students' level of english language skills, specifically in reading abilitys. This stage sets a baseline for understanding their cognitive capacity before receiving any intervention (Hikmawati, 2017).

2. Treatment

Hikmawati, (2017) the treatment consists of implementing the flipped classroom model to support students in improving their reading abilities. This is done over five sessions, with the researcher leading learning activities both in class and at home. Collaborative tasks challenge students to understand reading abilitys. The goal of this treatment is to implement the flipped classroom model in the classroom and apply reading abilitys to improve students' reading abilitys.

3. Post-test

After the treatment, a post-test is conducted to assess the improvement in students' reading abilities. The post-test tasks follow a similar format to the pre-test but address different themes to avoid repetition and ensure independence (Hikmawati, 2017). Students are asked to understand and apply the learning provided by the researcher to measure how much their skills have improved through the application of the flipped classroom model in integrated language learning. The post-test results are compared with the pre-test scores to determine whether the flipped classroom model strategy can improve students' reading abilitys.

To assess reading abilitys based on proficiency, the data were categorized using the scoring system aligned with the CEFR framework

(Council of Europe) and adapted for holistic language evaluation (Strugielska et al., 2022). The classification is as follows:

a. A score of 80-90 is classified as an Expert Reader.
b. A score of 70 is classified as a Proficient Reader.
c. A score of 60 is classified as a Competent Reader.
d. A score of 50 is classified as a Developing Reader.
e. A score of 40 is classified as a Basic Reader.
f. A score of 30 is classified as a Limited Reader.
g. A score of 10-20 is classified as a Very Limited Reader.

Table 4. Rubric Reading Assesment

Score	Classification	Indicator
Score 80-90	Expert Reader	Demonstrates excellent comprehension of complex texts, identifies key arguments and subtle details, and analyzes information critically with fluency and precision.
Score 70	Adept Reader	Reads and understands academic and non-academic texts effectively; interprets complex ideas and connects them meaningfully, showing growing analytical ability.
Score 60	Competent Reader	Understands main ideas and supporting details in various texts; makes moderate inferences and summarizes effectively with occasional errors.
Score 50	Limited but Developing	Understands general ideas in straightforward texts; struggles with detailed comprehension and advanced vocabulary; demonstrates limited inferencing skills.
Score 40	Marginally Competent	Faces challenges in understanding complex sentence structures and vocabulary; comprehension is slow and often incomplete.
Score 30	Limited Reader	Reads with difficulty, comprehending only very basic information; struggles to identify main ideas and relationships within texts.
Score 10-20	Very Limited Reader	Relies heavily on pictures or context clues for understanding; struggles with even simple texts and demonstrates inefficient and inconsistent comprehension.

F. Research Instrument

This research aims to measure a phenomenon in order to obtain the necessary data. The number of instruments used is adjusted according to the number of variables being studied (Sulatri, 2022). In this research, the researchers will create 60 multiple-choice questions. Before administering the pre-test and post-test, the questions will undergo testing for validity, reliability, discrimination, and difficulty to determine which ones will be included in the pre-test and post-test.

The question instrument used in this research was prepared based on J. B. Heaton's (1975) theory which states that reading ability includes understanding the main idea, detailed information, meaning of vocabulary in context, inference, and the purpose or attitude of the researcher. Based on this theory, the researcher designed multiple-choice questions consisting of a reading text followed by questions that measure these aspects as a whole. Each question was designed to represent indicators of reading ability according to the cognitive level of high school students, especially grade XI. The question design also considers content validity and language clarity to ensure accurate measurement of students' reading abilities within the context of this research.

G. Data Analysis Technique

The researcher analyzed the data using the pre-test and post-test results from the experimental group and the control group. The goal is to see if the application of this flipped classroom learning model can affect the level of students' understanding in English skills, especially in Reading abilitys (Nurhidayah, 2020).

1. Validity test

Following (Listiorini, 2022) validity measures how well an instrument captures what it wants to assess. If the significance value < 0.05, the item is considered valid. The score of each question will be correlated with the total score. An item is valid if the r-count > the r-table; if the r-count ≤ the r-table,

the item is considered invalid. Over all the data computing out with aid of computer facilities SPSS ver 28.

In this validity test, the researcher used multiple-choice questions with a total of 60 questions. Then the researcher conducted a trial in one of the high schools that had the same grade as the school that the researcher would use for the place where the researcher would conduct research later, the trial was carried out with a total of 30 students as the subject of the trial, after carrying out the trial the researcher processed the data on the test questions by calculating the validity of each question. Therefore, the researcher found a total of 30 valid questions out of the 60 questions. With the following indicators:

Table 5. Indicators of Reading Test Validity

No	Indicators	Number of Items	Items
1	Identifying explicit information	8	1, 2, 3, 6, 7, 10, 11, 15
2	Understanding main ideas	5	4, 8, 14, 16, 25
3	Making inferences	6	9, 18, 26, 28, 31, 33
4	Interpreting vocabulary in context	4	27, 30, 35, 46
5	Interpreting lessons or morals	7	49, 51, 52, 54, 56, 59, 60
Total			30

Based on the results of the validity test using SPSS, which was carried out on the question instrument at SMAN 03 Bengkulu City which is a senior high school with the same grade level as the school where the researchers will conduct the research, the following table presents the results of the validity test.

Table 6. Instrument Validity

Item	r Table	Significance	r Count	Results
1	0.361	0.000	0.642	Valid
2	0.361	0.003	0.529	Valid
3	0.361	0.038	0.380	Valid
4	0.361	0.042	0.374	Valid
5	0.361	0.515	0.124	Unvalid
6	0.361	0.018	0.429	Valid
7	0.361	0.000	0.809	Valid

8	0.361	0.001	0.562	Valid
9	0.361	0.047	0366	Valid
10	0.361	0.002	0.534	Valid
11	0.361	0.000	0.645	Valid
12	0.361	0.278	-0.205	Unvalid
13	0.361	0.086	0.319	Unvalid
14	0.361	0.002	0.543	Valid
15	0.361	0.000	0.691	Valid
16	0.361	0.000	0.670	Valid
17	0.361	0.669	-0.081	Unvalid
18	0.361	0.001	0.574	Valid
19	0.361	0.340	0.239	Unvalid
20	0.361	0.617	0.095	Unvalid
21	0.361	0.766	-0.057	Unvalid
22	0.361	0.108	0.299	Unvalid
23	0.361	0.309	0.192	Unvalid
24	0.361	0.587	0.310	Unvalid
25	0.361	0.000	0.761	Valid
26	0.361	0.000	0.765	Valid
27	0.361	0.000	0.610	Valid
28	0.361	0.005	0.497	Valid
29	0.361	0.315	0.190	Unvalid
30	0.361	0.000	0.722	Valid
31	0.361	0.000	0.606	Valid
32	0.361	0.894	0.025	Unvalid
33	0.361	0.000	0.758	Valid
34	0.361	0.073	0.332	Unvalid
35	0.361	0.007	0.458	Valid
36	0.361	0.287	0.201	Unvalid
37	0.361	0.284	0.206	Unvalid
38	0.361	0.107	0.300	Unvalid
39	0.361	0.598	0.002	Unvalid
40	0.361	0.178	0.253	Unvalid
41	0.361	0.515	0.124	Unvalid
42	0.361	0.294	0.198	Unvalid
43	0.361	0.691	0.076	Unvalid
44	0.361	0.134	0.280	Unvalid
45	0.361	0.020	0.422	Valid
46	0.361	0.024	0.410	Valid
47	0.361	0.509	0.125	Unvalid
48	0.361	0.136	0.279	Unvalid
49	0.361	0.024	0.411	Valid
50	0.361	0.091	0.314	Unvalid
51	0.361	0.005	0.501	Valid

52	0.361	0.003	0.527	Valid
53	0.361	0.690	-0.076	Unvalid
54	0.361	0.000	0.642	Valid
55	0.361	0.355	-0.175	Unvalid
56	0.361	0.416	0.022	Unvalid
57	0.361	0.564	0.110	Unvalid
58	0.361	0.103	0.303	Unvalid
59	0.361	0.008	0.478	Valid
60	0.361	0.005	0.502	Valid

2. Reliability test

Reliability testing is a tool used to evaluate the consistency of a questionnaire, which serves as an indicator of a variable. If a measuring instrument consistently produces stable results, it is considered reliable. Essentially, reliability measures the consistency of a questionnaire as an indicator of a variable or construct. Various methods can be used for reliability testing, including the Alpha coefficient method. In this approach, scores are divided into halves based on the number of questionnaire items and analyzed using Reliability Analysis. Ghozali in Iis Noviyanti *et al.,* (2023) states that a question item is deemed reliable if it achieves a value of 0.600 or higher, indicating that only 40% of the test scores reflect error variations. Over all the data computing out with aid of computer facilities SPSS ver 28.

In this reliability test, the researcher used multiple-choice questions with a total of 60 questions. Then the researcher conducted a trial in one of the high schools that had the same grade as the school that the researcher would use for the place where the researcher would conduct research later, the trial was carried out with a total of 30 students as the subject of the trial, after carrying out the trial the researcher processed the data on the test questions by calculating the reliability. After processing the reliability data on the test questions, the coefficient of the reliability test is 0.904 which was explained earlier that the question item is considered reliable if it reaches a value of 0.600 or higher therefore it can be assumed that this test has a high level of reliability.

Table 7. Case Processing Summary

Case Processing Summary			
		N	%
Cases	Valid	29	96,7
	Excluded[a]	1	3,3
	Total	30	100,0

Table 8. The Result of Reliability

Reliability Statistics	
Cronbach's Alpha	N of Items
0,904	60

Based on the table above, it can be concluded that the 60 questions, tested on a total of 30 students, are highly suitable for use in the pre-test and post-test to be conducted by the researcher. This is supported by the fact that the Cronbach's Alpha value reached 0.904. The value of Cronbach's Alpha can be interpreted as follows, according to Taber (2018):

Tabel:9. 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
0.81	Robust
0.84-0.90	Reliable
0.91-0.93	Strong
0.93-0.94	Excellent

3. Item Difficulty Test

Suharsimi Arikunto in Nasikhudin *et al.*, (2024) states that the item difficulty test is carried out to calculate how many respondents answered the item correctly. The item difficulty test is an important aspect in the development of research instruments. The process of testing the level of difficulty of this item is done by giving students questions with a total of 60 questions, The questions were administered to a total of 30 students in class XI at the school where the researcher tested the instrument questions. The level of difficulty of the items is done by dividing between the number of students who

answer correctly on certain items and the number of students who test the test questions, this calculation uses the following formula:

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

Description:

P: Difficulty level in percent

Nh: Number of test takers in the highest score group

Rh: Number of correct answers in the highest score group

Ni: Number of test takers in the low score group

Ri: Number of correct answers in the low score group

$$p = \frac{562}{8} + \frac{268}{8} \times 100\%$$

$$p = 51{,}9\%$$

Based on the results of the above calculations, it can be concluded that the level of difficulty of each item that has been tested on 30 students with a total of 60 questions can be called medium based on the interpretation of the level of test difficulty based on (Susanto and Rinaldi, 2015).

Table:10. Interpretation of Item Difficulty Test

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

4. Differentiability Test

Test differentiation is a test conducted to measure the difference between students who have high abilities and students who have low abilities. The number that indicates the magnitude of the power difference is called the discrimination index. (Solichin, 2017). The process of testing the differential

power of this question was carried out by giving questions to students with a total of 60 questions, the questions were given to students with a total of 30 students in class XI who were at the school where the researchers tested the instrument questions. Items that have a differentiation index> 0.30 then the item is declared good, otherwise if the differentiation index <0.30 then the item is declared bad. In the calculation of this differential power, the researcher uses the help of Microsoft Excel, to get the index of differential power in this differential power test, we use the following formula:

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

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: Scor Maximum

$$DP = \frac{562}{8} - \frac{268}{8} : 60$$

DP: 0,58

Based on the results of the calculations above, it can be concluded that the discrimination power of each item, tested on 30 students with a total of 60 items, is considered good according to the interpretation of the test's difficulty level. (Solichin, 2017).

Table:11. Interpretation of Item Differentiability Test

Differentiability Test	Interpretation
0,00 ---------- 0,20	Poor
0,21 ---------- 0,40	Enough
0,41 ---------- 0,70	Good
0,71 ---------- 1,00	Very Good
Negative, all of them	Not good, so all items that have a negative discriminating power value should be discarded.

5. Normality Test

The normality test determines whether the collected data follows a normal distribution, ensuring that statistical assumptions are met. SPSS is used to conduct this test to check the normality of the pre-test and post-test data. If the data is normally distributed, parametric statistical tests, such as the T-test, can be used. Over all the data computing out with aid of computer facilities SPSS ver 28.

6. Homogeneity Test

The homogeneity test examines whether the sample population has the same variance, which is crucial for making accurate comparisons between the experimental and control groups. 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).

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

7. Paired T-Test

The paired t-test is used to compare two means of the same group in two different conditions. In this research, the paired t-test will be used to see the significant differences of groups A and B before and after treatment. The following criteria are applied for 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.

8. T-Test Independent

The T-test independent is used to assess the statistical significance of the differences between the post-test group C and post-test group D scores of the experimental and control groups. This test identifies whether flipped classroom model has a significant effect on students' reading abilitys. The following criteria are applied for 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.

9. Two Way-ANOVA Test

Two-way ANOVA is used to examine the impact of two independent variables on a single dependent variable simultaneously. This test helps determine if each independent variable significantly affects the dependent variable and whether there is an interaction between the two independent variables. In this research, two-way ANOVA is used to measure the effect of using flipped classroom in improving students' reading ability, therefore the factors that will be used in this test are pretest, treatment, and the interaction between the two whether it will affect the value of the parameter being tested, namely the posttest. The following criteria are applied for the two way-ANOVA test:

a. If F-count is greater than F-table, then H_0 is rejected, indicating a significant effect.
b. If F-count is less than or equal to F-table, then H_0 fails to be rejected, indicating no significant effect.