CHAPTER III
RESEARCH METHOD

A. Research Design

This research uses the experimental method as its primary approach. The experimental method is part of the quantitative approach, which involves administering specific treatments to observe the effect of independent variables on dependent variables under controlled conditions (Rifmasari et al., 2021). In its implementation, this research applied a true experimental design, which is characterized by the random division of research subjects into experimental and control groups (Putri & Nurmilah, 2023).

This research used a Solomon Four-Group Design, which aims to test the effectiveness of a treatment while minimizing the potential influence of pre-tests on the results of the research. This design involves four groups of participants, consisting of two experimental groups and two control groups. The first group undergoes a series of pre-tests, interventions, and post-tests; the second group receives interventions and post-tests without pre-tests; the third group serves as a control with only pre-tests and post-tests without treatment; while the fourth group only undergoes post-tests without receiving pre-tests or interventions.

This structure allows researchers to determine whether

changes in students' learning outcomes are truly caused by the intervention provided or merely due to the influence of the pre-test. Thus, the Solomon design provides higher internal validity and more reliable results (Golaki et al., 2022).

Table 2. Solomon Four Group

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

Description:

X: Applying the CIRC Method

-: No Treatment/Pre-Test given

O_1: Pre-Test on Experimental Group

O_2: Post-Test on Experimental Group

O_3: Pre-Test on Control Group

O_4: Post-Test on Control Group

O_5: Post-Test on Experimental Group

O_6: Post-Test on Control Group

B. Place and Time of Research

This research was conducted at SMP Negeri 5 Kota Bengkulu from February 2025 to March 2025. The subjects in this research were students in grades VIII.B and VIII.C, each

consisting of 34 students, for a total of 68 students.

C. Sample and Population

1. Population

The population in this research refers to the entire group of subjects who have certain characteristics relevant to the focus of the research and from which conclusions can be drawn. In this context, the research population includes all students in grade VIII of SMP Negeri 5 Kota Bengkulu in the 2024/2025 academic year.

The population consists of 11 classes, namely classes VIII A to VIII K. The details of the number of students in each class are presented in the following table to provide an overview of the population that forms the basis for sampling in this research.

Table 3. Total Students

No.	Class	Gender		Total
		Female	Male	
1.	VIII. A	23	10	**33**
2.	VIII. B	20	14	**34**
3.	VIII. C	19	15	**34**
4.	VIII. D	21	14	**35**
5.	VIII. E	16	17	**33**
6.	VIII. F	19	14	**33**
7.	VIII. G	20	15	**35**

8.	VIII. H	17	15	**32**
9.	VIII. I	18	14	**32**
10.	VIII. J	20	13	**33**
11.	VIII. K	18	15	**33**
		Total		**367**

2. Sample

The sample represents a subset of the population in terms of both number and characteristics. In this true-experimental research, the researcher utilized a true-experimental design, in which the researcher randomly assigned the control group and the experimental group (Putri & Nurmilah, 2023). The researcher selected students from classes VIII.B and VIII.C at SMP N 5 Bengkulu as the research sample.

After considering several factors, the researcher chose representatives from each class as participants in this research. Class VIII.B, which consisted of 34 students, served as the experimental group, while Class VIII.C, also consisting of 34 students, served as the control group.

Table 4. total students in the experiment and control class

No.	**Class**	**Classroom**	**Gender**		**Total**
			Female	**Male**	
1.	Experimental group	VIII.B	20	14	**34**

2.	Control group	VIII.C	19	15	**34**
		Total			**68**

D. Research Variables

In this research, there are two types of variables: dependent variables, which are influenced by other variables, and independent variables, which affect other variables. The variables in this research include:

1) Reading Ability (Y), is a crucial language skill that involves understanding and interpreting word meanings. It connects closely with writing, listening, and speaking, enhancing overall communication. Developing strong reading skills requires regular practice and integration with other language abilities to achieve comprehensive proficiency.

2) CIRC Method (X), developed by Stevans et al., is a cooperative learning approach aimed at improving reading and writing skills. It encourages active participation, collaboration, and shared responsibility in small groups, making learning more engaging. While it boosts reading comprehension, problem-solving skills, and self-confidence, it also requires significant time and effective classroom management for optimal results.

E. Research Instrument

The research instrument is a tool designed based on the selected data collection technique. The researcher used a

multiple-choice test consisting of 60 questions about narrative texts at the eighth-grade junior high school level to measure students' reading ability. After conducting a trial test and calculating its validity, 27 questions were identified as valid.

According to J.B. Heaton, multiple-choice questions are an effective form of assessment to measure students' reading ability, especially at the grade 8 junior high school level. This type of question can be used to evaluate various reading sub-skills, such as identifying the main idea, locating specific details, interpreting word meanings in context, and drawing inferences from the text. Heaton emphasizes that well-constructed multiple-choice questions should include passages appropriate for students' proficiency levels, functional and reasonable answer choices, and logical distractors to prevent guessing. In addition, the questions should be designed to assess students' critical thinking skills, not memorization. If well developed, multiple-choice tests can be an objective, efficient, and comprehensive tool for evaluating students' reading comprehension (Heaton, 1975).

F. Data Collection Technique

The researcher employed pre-test and post-test methods as part of the data collection process in this research. The data collection was carried out through the following steps:

a. Pre-Test

The pre-test was administered to assess students' initial reading abilities prior to the implementation of the learning intervention (R. F. Adri, 2020). This test served as a baseline measure to determine students' reading comprehension levels before receiving any instructional treatment.

b. Treatment

The treatment phase involved the application of the Cooperative Integrated Reading and Composition (CIRC) method. During this stage, the researcher implemented learning activities designed to enhance students' reading abilities. These activities included introducing effective reading strategies and facilitating group-based discussions, allowing students to collaborate and improve their comprehension through structured interaction.

c. Post-Test

Following the treatment, a post-test was conducted to measure any changes or improvements in students' reading skills as a result of the intervention (Manik et al., 2021). The post-test contained questions of similar structure and difficulty to those in the pre-test but focused on a different reading topic. This allowed for an objective comparison to evaluate the effectiveness of the

CIRC method in improving reading comprehension.

To assess reading ability based on accuracy, the data were categorized using the scoring system introduced by the CAEL assessment (Ismail et al., 2022). The classification was as follows:

a. A score of 80-90 is classified as an Expert Reader.
b. A score of 70 is classified as an Adept Reader.
c. A score of 60 is classified as a Competent Reader.
d. A score of 50 is classified as a Competent but Limited Reader.
e. A score of 40 is classified as a Marginally Competent 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 5. Score Classification

Score	Classification	Indicator
Score 80-90	Expert Reader	Read academic texts easily, demonstrate understanding of academic texts equivalent to experienced readers, understand main ideas and some references in supporting details easily.
Score	Adept Reader	Read academic texts with ease

70		provided quality time, demonstrate understanding of academic texts approaching experienced academic readers, interpret information with flexibility.
Score 60	Competent Reader	Understands some important reasons and can analyze relevantly, reads casually and slowly and tries better than some other readers, able to interpret some information with multiple interpretations.
Score 50	Competent but Limited Reader	Can read and understanding some of the main ideas and can identify some examples of highly source detail, reads more slowly and with better effort than most readers, can sometimes misinterpretation about the information read.
Score 40	Marginally Competent Reader	Cannot comprehending the main ideas, has little vocabulary and lacks familiarity with the type of reading, reads more slowly than most readers.
Score 30	Limited Reader	Reading with very inaccurate information of accuracy also

		fluency, reading with some comprehension of the main idea but cannot analyze the reading specifically and relevantly often unable to identify foreign terms from reading.
Score 10 to 20	Very Limited Reader	Cannot read the text efficiently, takes the interpretation meaning from source of visual through photos, themes, rand the random vocabulary, etc., can sometimes comprehend the idea of the text but almost cannot control and comprehend all supporting idea and details.

Source: (Ismail et al., 2022).

G. Data Analysis Technique

The researcher used the pre-test and post-test results from both the experimental and control groups in the data analysis. The aim was to determine whether the implementation of the CIRC method significantly improves students' reading abilities.

1. Validity Test

Validity refers to how well a test measures what it is intended to measure and allows for accurate interpretation of the test results. A validity test was necessary to ensure that the instruments used in the research are appropriate as

measurement tools. The results of this test indicated whether the instrument has high, moderate, or sufficient validity for use (Wahyuda, 2022). The data were processed with the help of computer facilities using SPSS version 28.

In this research, the validity test was conducted using 60 multiple-choice questions. The researcher administered a trial at a junior high school with the same grade level as the school where the research would be conducted. The trial involved 30 students as participants. After the trial was completed, the test data were analyzed by calculating the validity of each question. The analysis results showed that out of the 60 questions tested, 27 were deemed valid based on the following indicators:

Table 6. Indicators of Reading Test Validity

No	Indicators	Number of Items	Items
1.	Identifying explicit information	7	2, 3, 21, 23, 33, 43, 51
2.	Understanding main ideas	7	4, 6, 25, 27, 36, 47, 56
3.	Making inferences	7	8, 16, 30, 35, 40, 44, 59
4.	Interpreting vocabulary in context	2	46, 55
5.	Interpreting lesson or morals	4	19, 38, 48, 58
	Total		27

Based on the validity test results using SPSS on the instrument questions that have been tested in one of the junior high schools with the same grade level as the school where the research will be conducted, the following table presents the validity test results.

Table 7. Instrument Validity

Item	r Table	Significance	r Count	Results
1	0.361	0.685	-0.077	Unvalid
2	0.361	0.026	0.406	Valid
3	0.361	0.007	0.485	Valid
4	0.361	0.013	0.447	Valid
5	0.361	0.143	0.274	Unvalid
6	0.361	0.049	0.362	Valid
7	0.361	0.935	0.016	Unvalid
8	0.361	0.004	0.514	Valid
9	0.361	0.133	-0.281	Unvalid
10	0.361	0.133	0.281	Unvalid
11	0.361	0.209	0.236	Unvalid
12	0.361	0.598	0.100	Unvalid
13	0.361	0.386	0.164	Unvalid
14	0.361	0.129	0.283	Unvalid
15	0.361	0.290	0.200	Unvalid
16	0.361	0.015	0.440	Valid
17	0.361	0.293	-0.199	Unvalid
18	0.361	0	0	Unvalid
19	0.361	0.009	0.467	Valid
20	0.361	0.719	-0.069	Unvalid
21	0.361	0.000	0.688	Valid
22	0.361	0.247	0.218	Unvalid
23	0.361	0.011	0.459	Valid
24	0.361	0.677	0.079	Unvalid
25	0.361	0.034	0.388	Valid

26	0.361	0.889	-0.027	Unvalid
27	0.361	0.001	0.579	Valid
28	0.361	0.921	-0.019	Unvalid
29	0.361	0.441	0.146	Unvalid
30	0.361	0.010	0.461	Valid
31	0.361	0.591	0.102	Unvalid
32	0.361	0.375	0.168	Unvalid
33	0.361	0.000	0.618	Valid
34	0.361	0.098	0.308	Unvalid
35	0.361	0.000	0.645	Valid
36	0.361	0.041	0.375	Valid
37	0.361	0.413	0.155	Unvalid
38	0.361	0.003	0.524	Valid
39	0.361	0.267	-0.209	Unvalid
40	0.361	0.030	0.397	Valid
41	0.361	0.188	0.247	Unvalid
42	0.361	0.411	0.156	Unvalid
43	0.361	0.002	0.545	Valid
44	0.361	0.003	0.530	Valid
45	0.361	0.247	0.218	Unvalid
46	0.361	0.000	0.654	Valid
47	0.361	0.018	0.430	Valid
48	0.361	0.033	0.391	Valid
49	0.361	0.187	0.247	Unvalid
50	0.361	0.826	-0.042	Unvalid
51	0.361	0.025	0.408	Valid
52	0.361	0.066	0.340	Unvalid
53	0.361	0.116	0.293	Unvalid
54	0.361	0.871	0.031	Unvalid
55	0.361	0.009	0.467	Valid
56	0.361	0.007	0.480	Valid
57	0.361	0.130	0.283	Unvalid
58	0.361	0.027	0.403	Valid
59	0.361	0.005	0.498	Valid
60	0.361	0.492	0.130	Unvalid

2. Reliability Test

Reliability refers to the stability and consistency of a measuring instrument in providing similar results when used at different times. Although achieving identical results is challenging due to factors like time, population, and sample differences, a strong positive correlation indicates the instrument's reliability. Reliability was crucial to ensure valid research results, so the researcher had to ensure their measuring instruments are reliable (Ahmed & Ishtiaq, 2021). The data were processed with the help of computer facilities using SPSS version 28.

In this reliability test, the researcher used multiple-choice questions with a total of 60 questions. Then, the researcher conducted a trial at one of the junior high schools that had the same grade level as the school that would be used for the research location later. The trial involved 30 students as the test subjects. After conducting the trial, the researcher processed the test data by calculating the reliability. After processing the reliability data for the test items, the obtained reliability coefficient was 0.797. An item is considered reliable if it reaches a value of 0.600 or higher. Therefore, this test was deemed to have 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

a. Listwise deletion based on all variables in the procedure.

Table 9. The Result of Reliability

Reliability Statistics

Cronbach's Alpha	N of Items
0,797	60

Based on the data in the table above, it can be concluded that of the 60 questions tested on 30 students, 27 questions were declared valid and suitable for use in pre-tests and post-tests. The reliability analysis results showed a Cronbach's Alpha value of 0.797, which reflects a high level of internal consistency. In accordance with the opinion of Mat Nawi et al. (2020), a Cronbach's Alpha value exceeding 0.70 is generally considered adequate, thus the instrument can be deemed reliable and appropriate for measuring students' reading ability in this research.

Table 10. Strength of Association Determination

Alpha Coefficient Range	Strength of Association
< 0.6	Poor
0.6 to < 0.7	Moderate
0.7 to < 0.8	Good
0.8 to < 0.9	Very Good
0.9 >	Excellent

Source: (Mat Nawi et al., 2020)

3. Item Difficulty Test

The level of difficulty of a question refers to the extent to which a question is considered difficult, which is assessed based on the percentage of students who are able to answer it correctly. The more students who answer correctly, the easier the question is categorized, and vice versa (Rame & Kesi, 2023). In this research, difficulty level analysis was conducted by presenting 60 questions to 30 eighth-grade students at the school where the instrument was tested. Each question was analyzed by calculating the proportion of students who answered correctly compared to the total number of students who took the test. This analysis was performed using the following formula:

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

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

$$p = \frac{335}{8} + \frac{187}{8} \times 100\%$$

p = 65,25 %

Based on the calculation, the difficulty level of the given question was 65.46%. Since it fell within the range of 41% - 70%, the question was categorized as having a moderate difficulty level.

Table 11. Interpretation of Item Difficulty Test

Question Category	Interpretation
Easy	71% - 100%
Moderate	41% - 70%
Difficult	21% - 40%

Source: (Hartati & Yogi, 2019).

4. Differentiability Test

Evaluating the discrimination power of a test refers to assessing its capability to differentiate between students with low and high proficiency levels. The discrimination index for each test item can be obtained through the DP

percentage column in the Anates table. A test item with a discrimination index of 0.30 is deemed good, whereas an index below 0.30 is considered poor. The purpose of discrimination power analysis is to improve the quality of exam questions based on empirical data and to evaluate how effective each question is in distinguishing between students who have mastered the material and those who have not (Magdalena et al., 2021). In the process of calculating discriminating power, researchers used Microsoft Excel software to obtain the discriminating power index for each question. The calculation was performed using the following formula:

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

$$DP = \frac{335-187}{8}$$

$$DP = \frac{18,5}{60} = 0,3083$$

Description:

DP: Question Differentiation 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 per group

Table 12. Interpretation of Item Differentiability Test

Differentiability Test	Interpretation
0,40 ≤ D < 1,00	Very Good
0,30 ≤ D < 0,39	Good
0,20 ≤ D < 0,29	Enough
0,00 ≤ D < 0,19	Poor
Negative	No Discrimination

Source: (Nurhalimah et al., 2022)

Based on the results of the Distinguishing Power (DP) calculation, the DP value in this research was 0.3083. When compared to the qualified category, which is the "good" category with $0.30 \leq D < 0.39$, the DP in this research was considered to be qualified.

5. Normality Test

The normality test determines whether the distribution of the independent variable (X) and the dependent variable (Y) in a regression equation follows a normal distribution. The regression equation is considered optimal if the data from both variables approximate a normal distribution. This test was performed to check whether the residuals or error values from the regression model were normally distributed, which is an important assumption in linear regression analysis (Utilitarian et al., 2024).

The normality test was conducted using the Shapiro-Wilk test. The criteria for normality determination were:

a. Asymp. Sig > 0.05: Data is considered normally distributed.
b. Asymp. Sig < 0.05: Data is considered not normally distributed.

This test was conducted using SPSS 28.

.

6. Homogeneity Test

The homogeneity test is used to determine whether the variances between the control group and the experimental group are consistent. Consistent variances indicate that both groups have equal variance, which is a key assumption in many parametric statistical tests (Supena et al., 2021). The following criteria were applied in this test:

a. Significance level (α) = 0.05
b. If sig > α, the sample variances are considered equal (homogeneous).
c. If sig < α, the sample variances are considered unequal (non-homogeneous).

This test was carried out using SPSS version 28

7. Paired T-Test

The paired t test was used to compare the means of

the same group in two different conditions. In this research, the test was used to evaluate whether there was a statistically significant difference in reading ability between groups A and B before and after the treatment was given. The analysis was based on the following decision criteria:

a. If the significance value (two-tailed) > 0.05, then there is no significant difference between the pre-test and post-test results.
b. If the significance value (two-tailed) ≤ 0.05, then there is a significant difference between the scores before and after the treatment.

Data processing for this test was performed using SPSS version 28.

8. Independent T-test

Independent t-tests were used to test the statistical significance of differences in post-test scores between the experimental and control groups. In this research, the test was used to assess whether the application of the Cooperative Integrated Reading and Composition (CIRC) method had a significant effect on students' reading ability. The evaluation of the analysis results was based on the following decision criteria :

a. If the significance value (2-tailed) > 0.05, it

indicates no significant difference in learning outcomes between Group C and Group D following the use of the CIRC method.

b. If the significance value (2-tailed) ≤ 0.05, it indicates a significant difference in learning outcomes between Group C and Group D, suggesting the effectiveness of the CIRC method in enhancing reading ability.

This test was performed using SPSS version 28.

9. Two-way ANOVA Test

Two-way ANOVA was utilized to analyze the effect of two independent variables on a single dependent variable simultaneously. This statistical test determines whether each independent variable has a significant impact on the dependent variable and whether there is an interaction effect between the two independent variables.

In this research, the Two-way ANOVA was applied to assess the impact of the CIRC method on students' reading ability. The factors analyzed in this test include the pre-test, treatment, and their interaction, to determine whether they significantly influence the post-test scores as the observed parameter.

The decision criteria for interpreting the results are as follows:

a. If the significance value > 0.05, the factors do not have a significant effect on the observed parameter.
b. If the significance value ≤ 0.05, the factors have a significant effect on the observed parameter.