CHAPTER III
RESEARCH METHODS

A. Research Design

This study used quantitative research, it was a numerical data-based approach to testing theories or hypotheses through statistical analysis (Creswell and Gueterman, 2024). This research is objective and systematic, in which researchers attempt to control research variables and use validated measurement tools to collect data that can be analyzed quantitatively. Quantitative research is a method grounded in positivist rationalism, applied to study a specific population or sample. It involves collecting data through standardized instruments and analyzing the results using statistical or numerical techniques (Susanto & Jaya, 2023). There are similar opinions among other experts who state that quantitative research is a type of research in which data is collected and coded in numerical form, and its objectives include description, statistical inference, causal explanation, and prediction, using methods such as social surveys, experiments, or analysis of official data or "big data" (Williams et al., 2022).

This study using a quasi-experimental design, it is a quantitative research approach that mimics true experiments but without random assignment (randomization). Researcher can use pre-formed groups and still conduct interventions or pre-test/post-test measurements (Creswell and Gueterman,

2024). The design used is a non-equivalent control group design, where two groups are selected through purposive sampling. This is used here to select subjects based on specific considerations, such as selecting groups with certain characteristics relevant to the quasi- experiment's objectives. The two groups, called the experimental group and the control group, was given measurements before (pre-test) and after (post-test) the treatment.

Before the treatment, both groups underwent pre-test to measure the initial speaking skill. After that, the experimental group received the project- based learning method for for meetings, while the control group followed the usual learning taught by the teacher. After the period, both groups underwent a post-test to assess changes in speaking skills.

The data obtained from the pre-test and post-test was analyze using appropriate statistical tests to find out whether there was a significant difference between the experimental group and the control group. It is expected that this study can provide empirical evidence regarding the effectiveness of project-based learning method on students' speaking skills.

Table 1. Non equivalent control group design.

Class	Pretest	Treatment	Posttest
Experiment	0_1	X	0_2
Control	0_3	O	0_4

Note:

X: Project Based Learning Method

O: Not Project Based Learning Method

0_1 : Experimental group pretest result before given treatment 0_2: Experimental group pretest result after given treatment

0_3 : Experimental group post test result before given treatment 0_4: Experimental group post test result after given treatment

B. Place and Time

This research was conducted at SMPN 16 Bengkulu City. The selection of this location is based on the initial finding that students in this school still have low English language skills. This research was conducted on seventh grade students in 4 meetings, in the 2025/2026 school year, consisting of 3 stages, namely pre test (1 meeting), treatment (4 meetings), and post test (1 meeting). By placing the research in this school, researcher can directly see and measure changes in students' speaking skill before and after the application of Project-Based Learning (PjBL).

C. Population and Sample

1. Population

Population is a group of individuals who have one or more characteristics in common and is the main focus of a study. This population can be individuals, groups, organizations, or communities that are relevant to the research objectives (Creswell and Gueterman, 2024). The population in this study was all seventh-grade students at SMPN 16 Bengkulu city in the 2025/2025 school year. Class VII has local A-G.

Table 2. The Number of Population

	Classes	Population
1.	VII A	32
2.	VII B	31
3.	VII C	30
	VII D	30
	VII E	34
	VII F	33
	VII G	31
	Total	221

Source: SMPN 16 Bengkulu City, 2025

Based on the above, the total population in this study is 221 students.

2. Sample**:**

Sample is a subgroup of the population selected to represent the overall population in the study. Sample selection is done to collect generalizable data or to gain an in-depth understanding of the phenomenon under study

(Creswell and Gueterman, 2024). The sample was determined using purposive sampling technique, purposive sampling is a sample selection technique in which the researcher intentionally selects individuals or groups that were considered the most informative or relevant to the phenomenon being studied. With two classes that have similar characteristics and similar speaking skills select as the experimental and control groups. The number of students in each class was 31 students, the total sample was 62 students.

Table 3. Sample of the Study

No	Group	Students	Classes
1.	Experimental	31	B
2.	Control	31	G
	Total	62	

Source: Managed Data, SMPN 16 Bengkulu City 2025

D. Operational Definition of Variables

1. Project Based Learning Method

Project-based learning (PjBL) a learning process through working on projects within a certain period of time. Comprehensively, a project is defined as a complex task based on challenging questions or problems, involving students in design, problem solving, decision making, or investigative activities; giving students the opportunity to work independently for a sufficiently long period of time;

and culminating in a realistic product or presentation. (Stanley, 2021).

2. Speaking Skill

Speaking skills are a student's ability to speak with accuracy (grammatical accuracy and pronunciation of segments), fluency (including speaking speed and minimal pauses between sentences), and comprehensibility (ease with which listeners understand speech), where accuracy and fluency are analyzed objectively from transcripts and temporal features, while comprehensibility and perceived fluency are measured through native speaker assessments of how easily and naturally the student's speech is understood (Suzuki & Kormos, 2020).

E. Technique of Collecting Data

Data collection techniques in research can include a variety of methods, depending on the research approach used. In quantitative research, data collection techniques often involve the use of standardized instruments such as questionnaires or surveys and standardized tests to collect numerical data (Creswell and Gueterman, 2024). The data collection procedure in this study was conducted by administering a speaking test to the research sample, which consisted of seventh-grade students of SMP 16 Bengkulu City. The speaking test consists of two main stages: pre-test and post-test.

a. Pre-test

Before implementing the Project-Based Learning Method (PjBL) to enhance students' speaking skills, the researcher first administers a pre-test. This pre-test is used to assess students' initial speaking skills. The test materials include asking and giving exercises and project creation, enabling the researcher to determine the extent of students' initial skills prior to the implementation of the PjBL method.

b. Post-test

After the treatment period was completed, a post-test was administered to the students. Students were asked to create a project, give a short presentation, and participate in a role-play activity involving asking and giving. The purpose of this test was to evaluate students' final performance and measure the extent of changes that occurred after the treatment was implemented. The post-test instrument was the same or similar to the pre-test instrument to ensure the validity of data comparison. After the pre- test and post-test data are collected, the results was analyzed to answer the research objectives. Data analysis was conducted using descriptive statistical methods, presenting the results of the students' speaking tests in the form of numbers and graphs. This aims to illustrate the changes in students' speaking skills before

and after the treatment. This process is expected to produce accurate and valid data to support the research conclusions regarding the impact of the teaching method applied on improving students' speaking skills.

c. Treatment

The treatment in this study was given to the experimental group through the application of the Project-Based Learning (PjBL) method, which refers to five main syntaxes. First, teachers formulated project objectives so that students could use expressions for asking and giving directions orally in an accurate, fluent, and understandable manner. Second, students participated in vocabulary and pronunciation training through drilling techniques and simple dialogues to strengthen their language proficiency. Third, students carried out speaking projects by creating school maps and composing dialogues based on specific directions. Fourth, students present project outcomes through role-play, evaluated using a rubric based on fluency, accuracy, and comprehensibility. Fifth, after the presentation, students participated in a reflection session to evaluate the learning process and receive feedback from the teacher. Meanwhile, the control group only received conventional instruction without involving the project. After the treatment was completed, a post-test is

conducted to measure the effectiveness of PjBL in improving students' speaking skills (Ponce & Reyes, 2025).

d. Documentation

In the context of Project-Based Learning (PjBL), documentation played a central role as authentic evidence of the entire learning process, from the exploration of ideas and student collaboration to the final product and reflection. For example, the reports, slide presentations, miniature models, and videos created by students not only reflected the final outcome but also the process and their deep understanding of the material. Additionally, research in the field of high school chemistry indicated that documentation through video/audio recording, lesson plans, syllabi, and curricula (Markula & Aksela, 2022).

F. Research Instruments

The main instrument in this study was a speaking skills test, which aimed to measure students' speaking abilities before and after implementing the project-based learning method. In this case, the researcher collected data by providing a script in the form of a short dialogue about asking and giving directions. Each dialogue had gaps that the students had to fill in. The researcher gave the students time to read the script and fill in the gaps. Then, students were

asked to come to the front and have a dialogue in pairs using the script they had learned. The researcher made a video while the students were having the dialogue. Students had the dialogue in a loud voice so that the researcher and classmates could hear clearly. Then, the results of the dialogue were assessed orally according to Heaton's (1990) criteria, namely assessing accuracy, fluency, and comprehensibility. Accuracy refers to the correct use of language elements such as grammar, vocabulary, and pronunciation so that speech is linguistically correct.

Fluency describes a person's ease of speaking, which is the ability to convey ideas smoothly without too many pauses, hesitations, or distracting interruptions. Meanwhile, comprehensibility relates to the extent to which listeners can understand the speaker's meaning, whether the message is conveyed clearly and easily understood without requiring much clarification. These three aspects are used together to provide a complete picture of the quality of a person's speaking skills.

Table 4.
Scoring Criteria Speaking Skill

Score	Accuracy	Fluency	Comprehensibility
6	Pronunciation is only very slightly influenced by the mother tangue, two or three minor grammatical	Speak without too great an effort with a fairly wide range of expression.	Easy for the listener to understand the speaker's ettention and general meaning. Very few interruption or classifications require
5	Pronunciation is slightly influenced by the mother tongue. A few minor grammatical	Has to make an effort at the time to search for words, smooth delivery on the whole and only a few unnatural pauses	The speaker's attention and general meaning are fairy clear. A few interruptions by listener for the sake of classification are necessary
4	Pronunciation is moderately influenced by the mother tongue but has no serious phonological error	Although he has to make an effort and search for words, there are not too many unnatural pauses. Fairly smooth delivery mostly occasionally fragmentary but success in conveying the general meaning fair range expressions.	Most of what the speakers says is easy to follow. His attention is always clear but several interruptions are necessary to help him to convey a massage or to seek clarification

3	Pronunciation is influenced by the mother only a few serious phonological errors.	He has to make an effort for much of the time. Oflen has to research for the desired meaning. Almost give up making the effort at time. Limited range of expressions.	The listener can understand a lot of what he said,but he must constantly seek clarification. Cannot understand many of the speaker's more complex or longer sentences.
2	Pronunciation is seriously influenced by the mother tongue with an error causing a breakdown in communication	Long pauses while he searches for the desired meaning. Frequently fragmentary and halting delivery. Almost give up making the effort at times. Limited range of expressions.	Only small bits (usually short sentences and phases can someone who is listening to the speaker
1	Serious pronunciation errors as well as many basic grammatical	Full oflong unnatural pauses. Very halting and fragmentary delivery. At times gives up making the effort. Very limited range expressions.times gives up making the effort. Very limited range expressions.	Hardly anything of what is she/ he can be understood. Even, when the listener a great effort or interrupts, the speaker is unable to clarify what he seems to have said.

Source: Heaton (1990).

G. Data Analysis Technique

Before conducting the hypothesis test to determine the effect of the Project-Based Learning (PjBL) method on students' English speaking skills, the researcher conducted several preliminary statistical analyses to validate the accuracy and reliability of the data collected.

1. Validity Test

Validity refers to the extent to which an instrument accurately measures what it is intended to measure. A valid test must reflect the true score and construct it purports to assess. In this study, the validity of each item was assessed by correlating individual item scores with the total test score (Markush et al., 2024). An item is considered valid if the r-count > r-table, and invalid if the r-count ≤ r-table. The analysis was conducted using SPSS version 25, and items with a significance value less than 0.05 were considered statistically valid.

Table 5. Test Instrumen Validity Criteria

No	Cronbach's Alpha	Validity Level
1.	$\alpha \geq 0.90$	Excellent
2.	$0.80 \leq \alpha < 0.90$	Good
3.	$0.70 \leq \alpha < 0.80$	Acceptable
4.	$0.60 \leq \alpha < 0.70$	Questionable
5.	$0.50 \leq \alpha < 0.60$	Poor
6.	$\alpha < 0.50$	Unacceptable

Source: Olsen et al., 2024

Table 6. Validity Test

Indikators	Sig.	r Table	r Count	Notes
Item 1	0.280	0.532	0.287	Unvalid
Item 2	0.867	0.532	-0.046	Unvalid
Item 3	0.496	0.532	0.184	Unvalid
Item 4	0.282	0.532	0.286	Unvalid
Item 5	0.444	0.532	0.206	Unvalid
Item 6	0.331	0.532	0.210	Unvalid
Item 7	0.862	0.532	0.047	Unvalid
Item 8	0.017	0.532	0.585	Valid
Item 9	0.275	0.532	0.290	Unvalid
Item 10	0.189	0.532	0.346	Unvalid
Item 11	0.051	0.532	0.495	Unvalid
Item 12	0.110	0.532	0.415	Unvalid
Item 13	0.793	0.532	0.071	Unvalid
Item 14	0.069	0.532	0.466	Unvalid
Item 15	0.387	0.532	-0.145	Unvalid
Item 16	0.497	0.532	0.183	Unvalid
Item 17	0.027	0.532	0.550	Valid
Item 18	0.019	0.532	0.579	Valid
Item 19	0.000	0.532	0.796	Valid
Item 20	0.003	0.532	0.685	Valid
Item 21	0.000	0.532	0.820	Valid
Item 22	0.003	0.532	0.699	Valid
Item 23	0.000	0.532	0.820	Valid
Item 24	0.000	0.532	0.878	Valid
Item 25	0.000	0.532	0.878	Valid
Item 26	0.005	0.532	0.670	Valid
Item 27	0.013	0.532	0.607	Valid
Item 28	0.021	0.532	0.570	Valid
Item 29	0.001	0.532	0.749	Valid
Item 30	0.001	0.532	0.749	Valid

Table 7. Validity Test 2 Tailed

Indikators	Sig. (2 tailed)		Notes
Item 1	0.280	<005	Unvalid
Item 2	0.867	<005	Unvalid
Item 3	0.496	<005	Unvalid
Item 4	0.282	<005	Unvalid
Item 5	0.444	<005	Unvalid
Item 6	0.331	<005	Unvalid
Item 7	0.862	<005	Unvalid
Item 8	0.017	<005	Valid
Item 9	0.275	<005	Unvalid
Item 10	0.189	<005	Unvalid
Item 11	0.051	<005	Unvalid
Item 12	0.110	<005	Unvalid
Item 13	0.793	<005	Unvalid
Item 14	0.069	<005	Unvalid
Item 15	0.387	<005	Unvalid
Item 16	0.497	<005	Unvalid
Item 17	0.027	<005	Valid
Item 18	0.019	<005	Valid
Item 19	0.000	<005	Valid
Item 20	0.003	<005	Valid
Item 21	0.000	<005	Valid
Item 22	0.003	<005	Valid
Item 23	0.000	<005	Valid
Item 24	0.000	<005	Valid
Item 25	0.000	<005	Valid
Item 26	0.005	<005	Valid
Item 27	0.013	<005	Valid
Item 28	0.021	<005	Valid
Item 29	0.001	<005	Valid
Item 30	0.001	<005	Valid

The table above shows the results of the validity test of the research tool conducted during the trial at SMPN 20 Bengkulu City, with 16 students taking the speaking test with 30 dialogues. The students engaged in dialogues in pairs, where each dialogue had blank words that they had to say correctly. In accordance with Heaton's opinion that good speaking skills are assessed based on fluency, accuracy, and comprehensibility, the researcher made video recordings of each pair of students conducting the trial.

The columns displayed included Sig. (significance value or p-value), r Table (critical r value based on sample size), r Calculated (Pearson's correlation coefficient between each item's score and the total score), and Description, which indicated whether the item was valid or invalid. In this pilot test, the r Table value was set at 0.532, obtained from a sample size of N = 16, so that the degrees of freedom (df) = N − 2 = 14. An item is declared valid if the r Calculated value is greater than r Table. The test results show that of the 30 items tested, 15 items are valid, namely items 8, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, and 30. Meanwhile, the other 15 items are invalid because the calculated r value is smaller than the r table value, and some even have negative values. Values that do not meet the validity criteria indicate that these items are

not yet able to measure the same construct as the instrument as a whole. Therefore, based on the validity test that has been carried out, it can be concluded that only 15 items can be used as instruments for further research.

2. Reliability Test

Reliability is defined as the consistency of an instrument in measuring a concept across repeated trials. A test is reliable if it yields the same results under consistent conditions (Hamidgkholgh et al., 2025). In this research, Cronbach's Alpha was used to determine the internal consistency of the speaking test items. A higher alpha coefficient (≥ 0.70) indicates acceptable reliability. The test was performed using SPSS version 25.

Table 8. Test Instrumen Reliability Criteria

No	Cronbach's Alpha	Reliability Level
1.	0,00-0,20	Lack of Realible
2.	0,20-0,40	Somewhat Realible
3.	0,40-0,60	Fairly Realible
3.	0,60-0,80	Realible
5.	0,80-1,00	Highly Realible

Source: Menon et al., 2025.

Table 9. Case Processing Summary

Case Processing Summary			
		N	%
Cases	Valid	16	100,0
	Excluded[a]	0	0,0
	Total	16	100,0
a. Listwise deletion based on all variables in the procedure.			

Table 10. Reliability Statisctics

Reliability Statistics	
Cronbach's Alpha	N of Items
0,868	**30**

Based on the results of the tryout conducted at SMPN 20 Bengkulu City with 16 students and 30 questions, Cronbach's Alpha value was 0.868. Referring to the instrument reliability criteria, this value is in the range of 0.80–1.00, which is classified as Highly Reliable. This indicates that the instrument used has a high level of internal consistency, so it can be declared reliable for use in research.

3. Difficulty Level Test

   Item difficulty helps determine how challenging each test item is for the students. Item difficulty is the proportion of students who answer a particular item correctly and is crucial in test construction (Rezigalla et al., 2024).

Difficulty Index

$$P = \frac{H}{N} + \frac{L}{N} x\ 100\%$$

Note:

H = number of students in the high-achieving group who answered the question correctly

L = number of students in the low-achieving group who answered the question correctly

N = total number of students in both groups combined (H + L)

$$p = \frac{159}{4} + \frac{57}{4} \times 100\%$$

$$p = 54\%$$

Table 11. Interpretation of Difficulty Level Test

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

Source: Caneda et al,. 2024.

4. Discriminating Power Test

Discriminating power evaluates the ability of an item to distinguish between students who perform well and those who perform poorly. A good test item was produced a higher proportion of correct answers among top-performing students than among low-performing students

(Makhija et al., 2024). This analysis is vital in a study examining the effect of PjBL on speaking ability, as it helps determine which tasks effectively differentiate skill levels.

The formula used is:

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

Notes:

U is the number of correct responses from high scorers.

L is the number of correct responses from low scorers.

N is the number of students in each group (27% of the total).

$$DP = \frac{159}{4} - \frac{57}{4} : 30$$

DP: 0,42

Tabel 12. Interpretation of Discriminating Power 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

Source: Murias et al., 2024.

5. Normality Test

Normality testing is one of the stages in statistical analysis that aims to determine whether the data used in the study is normally distributed or not (Khatun, 2021). Normal distribution is one of the important assumptions in the use of parametric statistical analysis such as t-tests and analysis of variance (ANOVA). Therefore, before conducting hypothesis testing using parametric methods, a normality test must be performed on the data, particularly the pre-test and post-test data for both the experimental and control groups. In this study, the normality test was conducted using the SPSS program with the Kolmogorov-Smirnov and Shapiro-Wilk methods. Both methods produce a significance value (Sig.). If the significance value is greater than 0.05 (Sig. > 0.05), the data is considered normally distributed and can be further analyzed using parametric statistical techniques. Conversely, if the significance value is less than 0.05 (Sig. < 0.05), the data is considered not normally distributed, and the use of non-parametric techniques as an alternative should be considered.

6. Homogeneity Test

The homogeneity test aims to determine whether the variation or variance between data groups (experimental and control groups) is homogeneous or the same. This test is also an important prerequisite before using the independent sample t-test, because if the variance between groups is not homogeneous, the test results may be invalid (Zhou et al., 2023). In this study, the homogeneity test was

performed using Levene's Test for Equality of Variances, which is available in SPSS software. If the significance value (Sig.) > 0.05, the variances between groups are considered homogeneous, and parametric tests can be applied. However, if Sig. < 0.05, the assumption of homogeneity is not met, and the researcher must use an alternative test such as Welch's t-test, which does not assume equal variances between groups.

7. T-test

The t-test used to compare the means between two groups (or one group against a specific value) to see if the difference is statistically significant. Before performing a t-test, it must be checked whether the data meets the assumption of normality (whether the data distribution is close to normal) and whether the variance between the two groups is similar when using an independent t-test (Tagliaferri et al., 2024). The t-test results were produced a p-value (2-tailed) compared to the significance level α (usually 0.05).

a. If $p < \alpha$, it means that the independent variable has a unique effect on the dependent variable.
b. If $p > \alpha$, it means that the independent variable has no effect on the dependent variable.

8. F-Test / Analysis of Variance (ANOVA / ANCOVA)

The F-test is used when comparing the means of more than two groups, or when testing several variables

together, including in ANCOVA (Analysis of Covariance) models where there are control variables (covariates) that can affect the dependent variable. Assumptions that must be met include homogeneity of variance between groups, normality of residuals, independence of observations, and, in ANCOVA, also the assumption that the relationship between covariates and dependent variables is uniform across all groups (Schwarz, 2025). After the analysis is complete, the F value and p-value are obtained. If $p < \alpha$ (e.g., 0.05) → it is concluded that there is a difference in the means between groups after controlling for the control variable. If $p > \alpha$, then there is no significant difference even though the means appear to be different. It is also important to report the related effect size (e.g., eta squared or partial eta squared) and confidence intervals, so that the results are more complete and the interpretation of the magnitude of the effect between groups is clear.

9. Effect Size

Effect size is a quantitative measure that shows how large the effect or strength of the relationship between two variables in a study is not just whether or not there is an effect. For example, when comparing two groups or measuring the relationship between variables, effect size provides a clear picture of the extent to which the independent variable explains the variability of the

dependent variable. In current research, it is highly recommended to report the effect size along with the p-value and confidence interval, so that readers can understand not only that the effect is significant, but also how large the effect is and whether it is practically useful in a real-world context (Tagliaferri et al., 2024).

Table 13. Value of the effect size

No.	Value	Effect size
1.	0,00-0,1,95	Very weak effect
2.	0,20-0,395	Weak effect
3.	0,40-0,595	Modest effect
4.	0,60-0,749	Strong effect
5.	0,80-1,00	Very strong effect

Source: Tagliaferri et al,. 2024).

10. Statistical Hypotesis

A statistical hypothesis consists of two statements: H_0 (null hypothesis), which states that there is no effect or no difference, and H_1 (alternative hypothesis), which states that there is an effect or a difference (Emmert, 2024). After performing the analysis (t-test, ANOVA/ANCOVA, etc.), we observe the p-value and compare it with the α that was set before the analysis (often 0.05).

1. If $p < \alpha$, then the null hypothesis is rejected and the alternative hypothesis is accepted.
2. If $p > \alpha$, then the null hypothesis fails to be rejected (there is not enough evidence that an effect exists).