CHAPTER III

RESEARCH METHOD

A. Research Design

Quantitative research is an approach or method used to obtain data or information that can be measured using objectively operationalized measuring tools. This research employed an true-experimental method. According to Creswell in (Ramadhana & Allo, 2021), Experimental research is research that tests a concept, method, or process to see if it has an impact on a dependent or outcome variable. In a learning context, it is often not possible to separate students into experimental and control groups.

This research used a true-experimental design with a pretest-posttest control group approach to see the impact of PBL method on speaking skill among seventh grade students at Pondok Pesantren PKPPS Hidayatul Qomariyah Bengkulu City. This design has four groups where two of them are given a pre-test and post-test 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 group will get treatment and there is only one control group. The experimental group receives a certain treatment, while the control group does not receive any treatment. This design contains two additional

control groups, which serve to reduce the influence of confounding variables (P.LavanyaKumari, 2013).

Table 1. 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 PBL Method

O : No Pre-Test Given

- : No Treatment Given

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 Research

This research was conducted at Pondok Pesantren PKPPS Hidayatul Qomariyah Bengkulu City. The research took place from February 13, 2025 to March 13, 2025. The subjects of this research were female students from class VII A, consisting of 16 students and class VII B, which also consisted of 16 students.

C. Sample and Population

1. Population

Population referred to a general scope consisting of objects or subjects with certain qualities and characteristics set by the researcher to be analyzed, from which the researcher could then draw conclusions about the research findings. Population was the whole element in research including objects and subjects with certain characteristics and characteristics (Sulistiyowati, 2017). In this research, the population was all female grade VII students at Pondok Pesantren PKPPS Hidayatul Qomariyah Bengkulu City in the 2024/2025 school year which was divided into two classes: A and B. The total population could be seen in the table below:

Table 2. Total Students

No.	Class	Gender		Total
		Female	Male	
1.	VII A		19	
2.	VII B		25	
3.	VII A	16		

4.	VII B	16		
Total				**76**

2. Sample

A sample was a portion of the population to represent the entire population (Sulistiyowati, 2017). The sample represented a portion of the number and characteristics possessed by the population. In this true experimental research, the technique used was the Solomon Four Group Design. The researcher selected students from class VII at Pondok Pesantren PKPPS Hidayatul Qomariyah Bengkulu City as the research sample. After considering several factors, the researcher chose representatives from each class as participants in this study. Class VII, consisting of 32 students, served as groups A, B, C, and D, each group consisting of 8 students.

Table 3. Total Students in the Experiment and Control Group

No	Group	Gender (Female)	Total
1.	A	8	
2.	B	8	
3.	C	8	
4.	D	8	
	Total		**32**

D. Research Variables

In this research, there were two types of variables: dependent variables, which were influenced by other variables, and

independent variables, which affected other variables. The variables in this research included:

1) Speaking Skill (Y), wich was the ability to understand meaning in order to communicate clearly and concisely. Many EFL students needed to understand grammar, pronunciation, vocabulary, fluency and comprehension. Teachers used a variety of teaching strategies to help their students become more proficient speakers (Bahiroh, 2022).

2) PBL Method (X), which was teaching that centered on students as the center of their educational journey. PBL was seen as a learning method that encouraged inquiry by placing emphasis on problem solving and collaboration among students (Zakaria et al., 2024).

E. Data Collection Technique

In this research, the researcher used a quantitative research approach, so the techniques used to obtain data related to the teaching of speaking using Problem Based Learning were oral tests in the pre-test and post-test. Based on (Harianto, 2018), the following were the steps in the data collection process:

1. Pre-test

The pre-test was given to students before treatment. The procedure for implementing the pre-test is as follows:

a. Researchers distributed tests to students.

b. The researcher gave directions for working on the test

to students.

c. The researcher collected the tests from the students after they finished answering the test.

2. Treatment

a. The teacher greeted the students and checked the students' attendance list.
b. The teacher divided the students into pairs.
c. The teacher explained the Problem Based Learning (PBL) Method clearly.
d. The teacher gave brainstorming or leading questions related to the topic about students' problems in making a dialogue about the discussion (sharing) that was going to be discussed.
e. The teacher gave a topic in the form of a dialogue to be discussed or responded to.
f. The teacher asked the students to speak the dialogue in pairs in front of the class.
g. The teacher gave the students guided questions related to the topic about students' problems in making a dialogue asking for and accepting invitations.
h. The teacher asked the students to discuss with their pairs.
i. The teacher randomly asked students to retell the results of the discussion.
j. The teacher gave feedback and evaluation.
k. The teacher gave students a chance to ask questions

related to the topic about students' problems in making a dialogue asking for and accepting invitations.

3. Post-test

After giving treatment to the students, the researcher administered the post-test to them. It was administered to see the value of the treatment using PBL. The test was given slightly differently from the pre-test, but the purpose was the same. The design of the test was based on the material that had been learned. If the results of the pre-test showed good results, then the learning process that was carried out would be more easily accepted by the students, because it was proven that their initial knowledge of the material was quite good. And also, if the pre-test results showed poor results, then the students would find it difficult to accept the learning process that would be enforced by the teacher (Siregar Aisyah et al., 2023).

The test used in this research was an oral test of speaking performance. This test was given after the treatment. This was done to determine the overall speaking ability of the students in the class. The students had been given different problems that were in line with their daily life and the current learning syllabus. The students were asked to solve the problems that had been formed. To assess speaking ability based on accuracy, the data was

categorized using the scoring system introduced by Brown's assessment (Jember, 2023).

Table 4. Rubric speaking assessment

Score	Aspects			
	Grammar	**Fluency**	**Pronunciation**	**Vocabulary**
5 (81 – 100)	Grammatical andlexical accuracy are extremelyhigh	Speak fluently without hesitation or searching forwords	Very clear, stress and intonation help to make meaning clear	Effective words choice
4 (61 – 80)	Quite accurate; some errors, but meaning is always clear	Some hesitations and sometimes has to search forwords	Generally clear; reasonable control of stress and intonation	Mostly effective words choice
3 (41 – 60)	Frequent errors; meaning is not always clear	Quite hesitant, limited range of vocabulary andstructure	Frequent errors;not always clear enough to understand	Frequently errors in words choice
2 (21 – 40)	Very frequent errors; difficulty in making meaning clear	Extremely hesitant; very limited range of language available	Very frequent errors; often very difficult to understand	Ineffective words choice
1 (1 - 20)	Almost unable to communicate	Almost unable to communicate	Almost unable to communicate	Almost unable to communicate

F. Research Instrument

In this research, there were two main instruments used to collect data, namely observation sheets and speaking tests. The functions of each of them were as follows:

1. The observation sheet was used to collect data about students' participation in the teaching and learning process of speaking by using the Problem Based Learning (PBL) method.
2. The speaking test was used to measure students' English speaking skills. The author divided the score into five criteria, namely pronunciation, grammar, vocabulary, fluency, and comprehension. Each criterion was then scored on a five-point rating scale, based on the score rating scale.

Based on J.B. Heaton's theory (1975), the construction of short answer questions to assess speaking skills for seventh-grade junior high school students should be adjusted to the learners' language level and focus on communicative competence. Short answer questions are designed to encourage students to give spontaneous oral responses using simple sentence structures and familiar, everyday contexts, such as introducing themselves, providing personal information, or introducing others. Heaton emphasizes that good questions must be clear, not overly complex, and provide students with the opportunity to demonstrate their speaking ability naturally. This type of

question is effective in building students' confidence in using spoken English at the basic level.

G. Data Analysis Technique

The researcher used the pre-test and post-test results of the experimental and control groups in the data analysis. The aim was to find out whether the application of the PBL method significantly improved students' speaking ability.

1. Validity Test

Research validity referred to the extent to which scientific research methodology had been applied throughout the process to produce research findings. Validity in quantitative research referred to how well the measuring instrument captured what it intended to measure. Validity indicated that an instrument could be criticized, but not necessarily be a valid instrument (Mohajan, 2017). If the significance value was <0.05, then the item was considered valid. The score of each question was correlated with the total score. An item was valid if r-count > r-table; if r-count ≤ r-table, then the item was considered invalid. Overall data calculations were carried out with the help of computer facilities SPSS ver 25.

Table 5. Instrument Validity

Item	r Table	Significance	r count	Result
1	0,468	0	0	Unvalid
2	0,468	0,135	0.366	Unvalid
3	0,468	0,032	0,505	Valid

4	0,468	0	0	Unvalid
5	0,468	0,314	0,251	Unvalid
6	0,468	0,003	0,654	Valid
7	0,468	0,013	0,575	Valid
8	0,468	0,024	0,528	Valid
9	0,468	0	0	Unvalid
10	0,468	0,816	0,059	Unvalid
11	0,468	0,009	0,595	Valid
12	0,468	0,385	0,218	Unvalid
13	0,468	0,000	0,741	Valid
14	0,468	0,001	0,733	Valid
15	0,468	0,005	0,636	Valid
16	0,468	0,314	-0,251	Unvalid
17	0,468	0,932	0,022	Unvalid
18	0,468	0,032	0,507	Valid
19	0,468	0,002	0,668	Valid
20	0,468	0,005	0,636	Valid
21	0,468	0,005	0,636	Valid
22	0,468	0,028	0,517	Valid
23	0,468	0,019	0,546	Valid
24	0,468	0,005	0,636	Valid
25	0,468	0	-0,000	Unvalid
26	0,468	0,006	0,620	Valid
27	0,468	0,000	0,787	Valid
28	0,468	0,048	0,473	Valid
29	0,468	0,043	0,481	Valid
30	0,468	0,001	0,733	Valid

31	0,468	0	-0,000	Unvalid
32	0,468	0	0	Unvalid
33	0,468	0	0	Unvalid
34	0,468	0	0	Unvalid
35	0,468	0	0	Unvalid
36	0,468	0	0	Unvalid
37	0,468	0	0	Unvalid
38	0,468	0	0	Unvalid
39	0,468	0	0	Unvalid
40	0,468	0	0	Unvalid
41	0,468	0,000	0,791	Valid
42	0,468	0,280	0,269	Unvalid
43	0,468	0,000	0,737	Valid
44	0,468	0,057	0,456	Unvalid
45	0,468	0,003	0,653	Valid

In this validity test, the researchers used questions in the form of 45 items. Then, the researcher conducted a trial at one of the MTs that had the same class as the school where the researcher would later conduct the research. The trial was conducted with a total of 18 students as the test subjects. After conducting the trial, the researcher processed the data from the interview or question trial by calculating the validity of each question. Therefore, the researcher found a total of 23 valid questions out of the 45 existing questions, with the following indicators:

Table 6. Indicators of Speaking Test Validity

No	Indicators	Number of Items	Items
1	Personal Information	4	3, 6, 8, 41
2	Preferences and Interests	5	7, 11, 13, 26, 27
3	Family and Relationships	4	18, 19, 20, 21
4	Education and School Life	3	22, 23, 24
5	Travel and Leisure	3	14, 15, 29
6	Daily Routine and Lifestyle	4	28, 30, 43, 45
	Total		23

2. Reliability Test

Reliability was the extent to which test results were not affected by random variables or chance and the consistency of test results in assessing something. The degree of consistency in test results with respect to one or more sources of inconsistency, such as question selection, rater selection, day and time of testing, was considered. Each dependability metric included measurement error and identified specific sources of inconsistency or specific combinations of those sources (Livingston, 2018). The data was processed with the help of computer facilities using SPSS version 28.

In this reliability test, the researchers used questions with a total of 45 items. Then, the researchers conducted a trial at one of the junior high schools that had the same grade level as the school that would later be used for the research location. The trial involved 18 students as test subjects.

After conducting the trial, the researcher processed the data from the trial results by calculating the reliability. After processing the data on the reliability of the test items, a reliability coefficient of 0.896 was obtained. An item was considered reliable if it reached a value of 0.600 or higher. Therefore, it can be assumed that this test had a high level of reliability.

Table 7. Case Processing Summary

Case Processing Summary			
		N	%
Cases	Valid	18	100.0
	Excluded[a]	0	0.0
	Total	18	100.0

Table 8. The Result of Reliability

Reliability Statistics	
Cronbach's Alpha	N of Items
0,896	45

Based on the table above, it can be concluded that of the 45 questions tested on 18 students, 23 questions were valid and suitable for use in the pre-test and post-test that the researchers would carry out later, because they showed Cronbach's Alpha results reaching 0.896. The Cronbach's Alpha value could be interpreted as follows, according to (Taber, 2018)

Table 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
0.81-0.83	Robust
0.84-0.90	Reliable
0.91-0.93	Strong
0.93-0.94	Excellent

Source : (Taber, 2018)

3. Item Difficulty Test

The difficulty level of a question reflected how easy or difficult it was to solve in a test. Usually, the difficulty level was expressed as a percentage of the proportion of students who gave the correct answer. Even so, it was possible to immediately retrieve a full and accurate response to an item. The purpose of this study was to compare students who did well and those who did poorly on the overall test (Nawir et al., 2023).

The process of testing the level of difficulty of this item was carried out by interviewing students with a total of 45 questions, involving 18 seventh-grade students at the school where the researchers tested the instrument questions. The level of difficulty of the items was determined by dividing the number of students who answered correctly on a particular item by the total number of students who took the test. This calculation used the following formula:

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

Description :

P : Difficulty Item
Nh : Number of test takers in the highest score group
Rh : Number of the correct answer in the highest score group
Ni : Number of the test takers in the low score group
Ri : Number of the correct answer in the low score group

$$p = \frac{208}{9} + \frac{76}{9} \times 100\%$$

$$p = 31{,}54\%$$

Based on the results of the above calculations, it could be concluded that the level of difficulty of each item that had been tested on 18 students with a total of 45 items could be considered medium, based on the interpretation of the level of difficulty of the test.

Table 10. 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

4. Differentiability Test

Differentiability analysis was conducted to examine test questions in terms of the ability of the test to distinguish between students who fell into the low and high categories. Item discriminating power was the ability of a test item to distinguish between high-ability and low-ability testees (Magdalena et al., 2021).

The process of testing the distinguishing power of this question was carried out by interviewing students with a total of 45 questions. The researcher interviewed 18 seventh-grade students at the school where the researcher tested the question instrument. Items that had a differentiating index > 0.30 were considered good, while items with a differentiating index < 0.30 were considered bad. In calculating this differentiating power, the researchers used the help of Microsoft Excel to obtain the index of differentiating power. For this differentiating power test, the following formula was used:

$$Dp = \frac{BA}{JA} - \frac{BB}{JB}$$

Description:

DP: Question Differentiation Power

BA: Number of correct answers from high score group test

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

$$Dp = \frac{BA}{JA} - \frac{BB}{JB}$$

$$Dp = \frac{208}{9} - \frac{76}{9}$$

$$Dp = \frac{132}{9} = 14{,}67$$

$$Dp = \frac{14{,}67}{23}$$

$$Dp = 0{,}638$$

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.

Source : (Magdalena et al., 2021)

5. Normality Test

The normality test was conducted to determine whether the regression residual variables had a normal distribution or could not be analyzed using the non-parametric statistical test Kolmogorov-Smirnov (Tsagris & Pandis, 2021). If the data was normally distributed, parametric statistical tests, such as

the T-test, could be used. This test was conducted using SPSS 25.

6. Homogeneity Test

The homogeneity test aimed to determine whether the research population was homogeneous. This test evaluated whether the data from different samples had the same variance. The following criteria were used to determine homogeneity:

a. Significance level (α) = 0.05
b. If sig > α, the sample variances were considered equal (homogeneous).
c. If sig < α, the sample variances were considered not equal (non-homogeneous).

This test was conducted using SPSS 29.

7. Paired T-Test

The paired t-test was used to compare two means of the same group in two different conditions. In this research, the paired t-test was used to see the significant differences between groups A and B before and after treatment. The following criteria were applied for the T-test analysis:

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

and the final variable (posttest).

Overall, the data were computed with the aid of computer facilities using SPSS ver 28.

8. T-Test Independent

The independent T-test was 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 identified whether problem-based learning (PBL) had a significant effect on students' reading skills. The following criteria were applied for the T-test analysis:

a. If the significance value (2-tailed) > 0.05, then there was 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 was a significant difference between the learning outcomes using the flipped classroom model in learning reading ability in group C and D.

Overall, the data were computed with the aid of computer facilities using SPSS ver 28.

9. Two Way-ANOVA Test

Two-way ANOVA was used to test the effect of two independent variables on one dependent variable simultaneously. This test was conducted to determine whether each independent

variable had a significant effect on the dependent variable and to find out whether there was an interaction between the two independent variables. In this research, two-way ANOVA was used to measure the effect of using PBL in improving students' reading skills. Therefore, the factors that were used in this test were pretest, treatment, and the interaction between the two, to see whether it affected the value of the parameter being tested, namely the posttest. The following criteria were applied for the two-way ANOVA test:

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

Overall, the data were computed with the aid of computer facilities using SPSS ver 28.