

CHAPTER III

METHODOLOGY

A. Research Design

This research employed a quantitative approach used a non-equivalent control group quasi-experimental design to examine the effect of Communicative Language Teaching (CLT) on the speaking skills of seventh-grade students at SMP Negeri 4 Kota Bengkulu. Two existing classes were selected as cluster groups: one class was given and treated using CLT and for control class using conventional method. Before the intervention, both groups were given a pre-test speaking recorded to ensure initial equality and as a covariate in further analysis; after the treatment period, a post-test speaking was conducted using monologue and dialogue tasks assessed by two independent assessors using a rubric covering fluency, pronunciation, accuracy, vocabulary, and interactional competence.

CLT implementation was standardized through lesson plans and brief teacher training, as well as a fidelity checklist in the field to ensure that the difference in treatment was indeed in frequency, not content. The main quantitative analysis used ANCOVA (post-test as the dependent variable, pre-test as the covariate) to control for initial differences effect size (Cohen's d / η^2) was reported along with prerequisite tests (normality & homogeneity). This design follows methodological practices commonly used in quasi-experimental CLT studies and speech

intervention evaluations in secondary school contexts. (Al-Garni, S. A., *et al*, 2019)

Based on Sugiyono's opinion, the non-equivalent control group design scheme is described as shown in Table 3.1 (Sugiyono, 2020)

Table 3.1 Research Design

Experimental Group	O1	X	O2
Control Group	O3	-	O4

Description :

O1 = Pre-test in experimental class

O2 = Post-test on experimental class

O3 = Pre-test in control class

O4 = Post-test on control class

X = Treatment using Communicative Language Teaching

B. Research Location

This research was conducted at SMPN 4 Kota Bengkulu with seventh-grade students in the even semester of the 2025/2026 academic year, located at Jalan Cimanuk Km 6.5, Jalan Gedang, Gading Cempaka District, Kota Bengkulu. The study took place from March 5 to April 16, 2026. Based on initial observations, this school is known as one of the leading and favorite schools in Kota Bengkulu. The school is

recognized as having an A accreditation, a safe environment, and a focus on both academic and non-academic achievements, as well as creating a comfortable and conducive learning atmosphere and providing comprehensive facilities, including in literacy, digital technology, and supporting facilities for teaching and learning activities (KBM). At this school, the seventh grade consists of 10 classes, from VII.1 to VII.10.

C. Population and Sample

1. Population

The next stage in the research procedure was to determine the subjects or students who will be the population in this research (Willie, 2024). A population is a cornerstone in research methodology, encapsulating a collective of individuals who exhibit shared characteristics within a defined geographic or institutional domain. This concept serves as the foundation for research efforts, establishing the main framework for formulating research questions and hypotheses.

The population provides researchers with a comprehensive understanding of the group under study, facilitating the exploration of various phenomena and phenomena. In contrast, the target population constitutes a subset of the broader population being investigated, characterised by specific attributes or criteria directly relevant to the research inquiry. And the researcher taken

by the population in the seventh grade students at SMP Negeri 4 Kota Bengkulu on academic year 2025/2026. The population of this research consisted of 328 seventh grade students at SMP Negeri 4 Kota Bengkulu, consisting of ten classes, where the population was shown in the following table:

Table 3.2 Population of The Research

Class	Number of students
VII 1	33
VII 2	32
VII 3	33
VII 4	33
VII 5	34
VII 6	32
VII 7	32
VII 8	33
VII 9	33
VII 10	33
TOTAL	328

Source : Junior High School 4, 2025

2. Sample

In this study, part of the population was selected as the focus of the study and used as the research sample. A sample is a part of the number and characteristics possessed by the population. A sample is a number of individuals selected from the population and is a part that represents all members of the population. A good sample is

representative of the population. A sample that is not representative of every member of the population, regardless of the sample size, cannot be generalized to the population. (Suriani *et al.*, 2023).

This study used purposive sampling, which is a sampling technique based on specific considerations or criteria. Purposive sampling is generally used by researchers who want to explore a process by interviewing a predetermined and easily identifiable group of elites, selected based on specific criteria, and requiring researchers to access specific respondents. Researchers who use purposive sampling make a list of elites they want to interview before starting the research. Because the research population is very visible, it is relatively easy for researchers who use purposive sampling to find the names of respondents and their professional contact details (Bakkalbasioglu, E. 2020). The sample in this research consisted of 66 students of class VII 1 and VII 2 at Junior High School 4 Bengkulu City. Class VII 1 was designated as the experimental class, while class VII 2 as the control class, as shown in the following table:

Table 3.3 The Sample of the Research

Class	Subject of students	Male	Female	Number of students
VII 1	Experimental class	16	17	33
VII 2	Control class	15	17	32
Total				65

Source : Junior High School 4, 2025

D. Operational Definition of Variables

The research variable is a characteristic, attribute, or value of a person or object being studied and can be grouped based on several characteristics. The independent variable is the variable that influences the dependent variable, either directly or indirectly, and functions as the cause of changes. Meanwhile, the dependent variable is the variable that is influenced, becomes the main focus of the research, and depends on the independent variable. (Marliana Susianti, 2024). These variables were as follows:

1. Communicative Language Teaching (CLT)

According to Qasserras (2023), CLT is teaching method a set pedagogical principles grounded in interaction, meaning negotiation, task-based learning, and learner-centered instruction. The review emphasized that the core of CLT lies in developing communicative

competence through authentic practice and structured social interaction. It also notes that effective CLT implementation depends on adequate task frequency, appropriate scaffolding, and teacher readiness. (Qasserras, L. 2023)

2. Speaking Skills

Speaking skills refer to a person's to express ideas, feelings, information, and opinion clearly through spoken language. Speaking is a productive skill in oral communication that involves the use of appropriate fluency, accuracy (pronunciation), comprehensibility, etc. Through speaking skills, individuals can communicate meaning, participate in conversations, respond to questions, and interact effectively in social, academic, and professional situations. (Hamburabbi, *et al*, 2023)

E. Data Collection Technique

The data collection technique in this study was designed to obtain valid and reliable evidence regarding the effect of the frequency of applying the Communicative Language Teaching (CLT) on students' speaking skills. The primary instrument used was a pretest–posttest speaking test administered before and after the treatment period. Here are three data collection techniques:

1. Test

Test are instruments or procedures that are arranged in a standardized manner to assess behavior or performance systematically through a number of questions or tasks. Test aim to assess and compare the abilities, skills, qualities, or knowledge of learners based on predetermined standards (Adom, D. *et al*, 2020).

a. Pre-test

A pre-test is an evaluative tool administered before formal instruction begins, designed to assess learners' existing knowledge, skills, or attitudes related to a specific topic. A pretest is an assessment given before the implementation of a teaching approach or intervention. Pre-tests are a form of evaluation carried out before the formal learning process or administration of treatment, with the aim of assessing students' initial levels of knowledge and skills. The results of these pre-tests are used as an initial benchmark for researchers to evaluate the effectiveness of learning interventions by comparing students' achievements before and after the learning process. (Brown, 2004)

b. Treatment

In quasi-experimental research, treatment refers to the intervention, instruction, or specific condition provided by the researcher to a group of participants in order to observe, measure, and assess its effect on

outcome variables (usually the dependent). This treatment is deliberately given after the initial measurement (for example a pre-test) and before the final measurement (for example a post-test) to find out whether and to what extent the intervention affects changes in participants' performance, behavior, or knowledge. The concept of this treatment becomes a main element in research designs that aim to assess the causal relationship between independent variables (the treatment) and dependent variables under controlled conditions. (Arib, M. F, *et al.* 2024).

Table 3. 4 Treatment Plan.

Session	Activities & Focus	Objectives
1. Room in the House	- Introduction to topic “Home Sweet Home” - Learning vocabulary of rooms (living room, bedroom, kitchen, etc.) - CLT Activities: Guess the Room & Role Play (asking about rooms in a house)	To introduce students to basic vocabulary about rooms and encourage them to communicate using simple questions and answers.
2. Things in the House & Prepositions	- Review previous vocabulary - Learning	To improve students’ ability to describe objects and their

	vocabulary of things in the house (table, chair, sofa, etc.) - Learning prepositions (in, on, under) - CLT Activities: Asking and answering about object location & Group Discussion	locations using simple sentences in communicative contexts.
3. Household Chores	- Introduction to household activities (sweep the floor, wash dishes, etc.) - Practicing simple present tense in daily activities - CLT Activities: Asking about chores & Pair Work discussion	To develop students' speaking ability in expressing daily activities and interacting with peers using appropriate vocabulary.
4. Giving Instructions & Speaking Practice	- Review all vocabulary (rooms, objects, chores) - Describing messy room and giving instructions - CLT Activity: Group Discussion (Describe your	To evaluate students' speaking improvement and enable them to communicate ideas fluently and confidently in real-life contexts.

	dream house) - Speaking performance / post-test & reflection	
--	--	--

c. Post-test

After implementing the CLT in learning, a post-activity test is an assessment given after the intervention (treatment) to measure student achievement or performance after the intervention. A post-activity test is an exam given at the end of the teaching period after students receive instruction through the Communicative Language Teaching (CLT). Post-tests are a form of evaluation conducted after the learning process or instructional intervention, with the aim of measuring students' level of understanding. (Brown, 2004)

2. Speaking Test

Speaking tests in language assessment are a form of evaluation designed to measure a learner's ability to produce spoken language that reflects real communicative skills. These tests assess the extent to which learners can complete speaking tasks such as describing, narrating, or interacting in specific situations, and the scores obtained indicate the learners' level of oral proficiency in terms of pronunciation, fluency, vocabulary use, grammatical

accuracy, and communicative effectiveness. Speaking tests emphasized actual speaking ability, not just knowledge about the language, and are often used to evaluate teaching outcomes as well as the learning process in language education. (Heidari Vinchek, M, *et al.* 2024).

Table 3.5 Scoring Rubric

No	Criteria	Score	Category	Description
1	Accuracy	5	Excellent	Pronunciation is clear and natural with almost no errors.
		4	Good	Pronunciation is generally clear with a few minor errors.
		3	Pair	Pronunciation is sometimes unclear and may cause misunderstanding.
		2	Poor	Pronunciation errors frequently make understanding difficult.
		1	Very Poor	Pronunciation is very difficult to understand
2	Fluency	5	Excellent	Speaks very fluently with almost no hesitation.
		4	Good	Speaks fairly smoothly with occasional hesitation.
		3	Pair	Speech is somewhat slow and hesitant.
		2	Poor	Speaks with frequent pauses and hesitation.

		1	Very Poor	Speech is extremely slow and fragmented.
3	Comprehension	5	Excellent	Fully understands the topic and responds appropriately.
		4	Good	Understands most of the conversation and responds appropriately.
		3	Pair	Understands some parts of the conversation but needs repetition.
		2	Poor	Has difficulty understanding the conversation.
		1	Very Poor	Unable to understand or respond appropriately.

Source : Brown, H. D. (2004).

3. Documentation

Documentation such as audio or recordings is used to document students' speaking performance during learning activities and speaking tests. The instrument employed in this methodology is a voice recording device that operates to objectively capture the verbal performance of students. These auditory recordings serve as supplementary data that may be subjected to further analysis if deemed necessary, particularly to validate speaking test scores and observation results. Through audio documentation, the evaluation of students' verbal competences can be conducted with enhanced precision and accountability.

F. Research Instruments

Research will use instruments to collect data from subjects. In quantitative research, instruments are very important because they are used to measure the variables being studied. In this study, the measuring instruments used were tests, consisting of pre-tests and post-tests. These tests were designed and administered to obtain information about whether or not there was an effect on students' speaking skills when learning was changed to the CLT. Therefore, this test was given to students to evaluate their speaking skills. The pre-test and post-test were included in this test, which was given to two groups: the experimental class and the control class. Before being used in research, the validity and reliability of this vocabulary test measuring instrument must also be tested to ensure the accuracy and consistency of the measuring instrument. The main research instruments include Speaking test (pre-test & post-test)

This combination of instruments was used to measure outcomes speaking, document treatment implementation Communicative Language Teaching method, and control confounding factors. Quantitative CLT studies using a similar combination of instruments can be found in local/international literature. (Dwi Nur Hadiyansah WS & Nuril Qomariyah, 2024).

G. Data Analysis Technique

The prerequisite data analysis in this study includes normality test, homogeneity test, t-test, and f-test.

1. Validity of the test

Before conducting the study, the researcher first validated the instruments used. Those instruments must demonstrate suitability with the aspects to be measured. According to Kimberlin & Winterstein (2008), validity refers to the extent to which a study is able to accurately measure what it is intended to measure.

Table 3.6 Test Instrument Validity Criteria

No	Cronbach's Alpa	Validity Level
1	0,81 - 1,00	Highly
2	0,61 - 0,80	High
3	0,41 - 0,60	Fairly
4	0,21 - 0,40	Low
5	0,00 - 0,20	Very Low

Source : Hussey et al., (2023)

The researcher analyzed the instrument test results using SPSS version 27 statistical software. The instrument consisted of 27 items that covered three indicators of speaking skills. The results of the data processing from the instrument validity test are presented as follows :

Table 3.7 Instrument Validity

Variable	Item	r table	r count	Result
Soal	S1	0,468	0,635	Valid
	S2	0,468	0,722	Valid
	S3	0,468	-0,079	Unvalid
	S4	0,468	-0,128	Unvalid
	S5	0,468	0,397	Unvalid
	S6	0,468	-0,128	Unvalid
	S7	0,468	0,604	Valid
	S8	0,468	0,158	Unvalid
	S9	0,468	0,246	Unvalid
	S10	0,468	0,581	Valid
	S11	0,468	-0,246	Unvalid
	S12	0,468	0,531	Valid
	S13	0,468	0,494	Valid
	S14	0,468	-0,164	Unvalid
	S15	0,468	0,158	Unvalid
	S16	0,468	0,672	Valid
	S17	0,468	0,301	Unvalid
	S18	0,468	0,290	Unvalid
	S19	0,468	0,402	Unvalid
	S20	0,468	0,258	Unvalid
	S21	0,468	0,580	Valid
	S22	0,468	0,445	Valid

	S23	0,468	0,338	Invalid
	S24	0,468	0,487	Valid
	S25	0,468	0,559	Valid
	S26	0,468	0,502	Valid
	S27	0,468	0,829	Valid

In this validity test, the author used 27 test items with predetermined indicators, and this instrument test was carried out at MTS Negeri 1 Kota Bengkulu, which has the same accreditation level as the school that will be used as the research site. The instrument test was conducted on February 26, 2026, located at Jalan Nangka No. 3, Panorama, Kec. Singaran Pati, Kota Bengkulu. After analysis, 13 of the 27 items were tested and declared valid.

2. Reliability of the test

According to Kouam Arthur William (2024), reliability refers to the level of consistency and stability of a measurement result. After testing all questionnaire items, a number of items were obtained that met the validity requirements. Therefore, the next stage after the validity test is the implementation of the reliability test. According to (Creswell,.et al 2023), Cronbach's Alpha coefficient is calculated to ensure further reliability. Reliability with a coefficient above 0.70 is generally considered acceptable in

educational research. High reliability increases the trustworthiness of the study and the replicability of the results. Reliability values are measured using Cronbach's Alpha; if Cronbach's Alpha > the critical r-table value, then the variable can be declared valid.

Table 3.8 Test Instrument Reliability Criteria

No	Cronbach's Alpha	Reliability Level
1	0,00 - 0,20	Lack of Reliable
2	0,20 – 0,40	Somewhat Reliable
3	0,40 – 0,60	Fairly Reliable
4	0,60 – 0,80	Reliable
5	0,80 – 1,00	Highly Reliable

Source : Iqbal & Salsabila (2023)

In the analysis, the researcher utilized SPSS statistical software version 27. Reliability testing of the instrument was carried out using the Cronbach's Alpha method, which is used to assess internal consistency or inter-item correlation. The results of the analysis show that the instrument has high reliability, as seen from the following Cronbach's Alpha values :

Table 3.9 Instrument Reliability

Reliability Statistic	
Cronbach's Alpa	N of items
0.758	27
Reliable	

In this reliability test, the researcher was carried out on 27 test items that had been developed according to the specified indicators. Based on the reliability criteria, an instrument is categorized as highly reliable when the Cronbach's Alpha value ranges from 0.80 to 1.00. The result of the analysis showed a Cronbach's Alpha coefficient of 0.758, indicating that the instrument has a reliable level of consistency. Thus, the data obtained from the instrument can be considered reliable and suitable for further research purposes.

3. The level difficulty test

One of the crucial stages in the development of research instruments is testing the difficulty level of test items. Difficulty level analysis is carried out by examining each question based on its degree of difficulty in order to classify the questions into easy, medium, or difficult categories (Magdalena, Fauziah, et al., 2021). The determination of the difficulty level of a question is done by comparing the number of students who answered

correctly with the total number of test participants. In general, the item difficulty level can be calculated using the following formula :

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

Description :

P = Difficulty level in percentage

Nh = Number of test takers in the high score group

Rh = Number of correct answers in the high score group

Ni = Number of test takers in the low score group

Ri = Number of correct answers in the low score group

$$P = \frac{319+108}{5+5} \times 100\%$$

$$P = \frac{427}{10} \times 100\%$$

$$P = 42,7 \%$$

Table 3.10 Interpretation of difficult question categories

Question category	Interpretation of result
Classified as easy	0,71%-----100%
Classified as medium	0,31%-----0,70%
Classified as difficult	0%-----0,30%

Source : Arikunto, S. (2018)

In the instrument testing process for the research, the difficulty level of test items is an aspect that needs attention so that the items used are neither too easy nor too difficult for respondents. Based on the analysis results, a difficulty level of 42.7% was obtained. Referring to the item difficulty criteria, that value falls into the classified as medium category because it is within the 41%–70% range. Thus, the test items that have been examined can be stated to meet the criteria and are suitable for use as research instruments.

4. Item discrimination test

Discriminating power refers to an item's ability to distinguish students with high ability (upper group) from students with low ability (lower group) in responding to the material being tested (Pradita et al., 2023). This evaluation aims to determine the extent to which vocabulary test items can identify differences in English vocabulary mastery between high-ability and low-ability students. Items with high discriminating power will show a significant difference in results between the two groups. In this study, discriminating power analysis was conducted by dividing students into two groups based on the vocabulary test scores obtained, namely the upper group who obtained the highest scores and the lower group who obtained the lowest

scores. In general, the discriminating power calculation uses the following formula :

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

Description :

DP = Discrimination index

U = Number of correct answers from test takers with high scores

L = Number of correct answers from test takers with low scores number of students in each group

N = Number of students in each group

$$DP = \frac{319 - 108}{10}$$

$$DP = 21,1 \%$$

Table 3.11 Interpretation of the Power Difference Test

T-test	Interpretation of result
$0,71 \leq DP \leq 1,00$	Very good
$0,41 \leq DP \leq 0,70$	Good
If $+ 0,21 < DP < 0,40$	Sufficiently qualified
If $0 < DP < 0,20$	Weak

Source : Arikunto, S. (2018)

Discrimination power testing was conducted to determine the assessment instrument's ability to accurately distinguish students' levels of vocabulary mastery. Based on the calculations, a discrimination value of 21.1% was obtained. Referring to the applicable evaluation criteria, that value falls within the range $0.21 < DP (0.33) < 0.40$ and is therefore classified as sufficient. Thus, the instrument used in this study has adequate discrimination power and meets the established criteria.

5. Normality test

Normality testing is a statistical method aimed at determining whether research data have a normal distribution or not. The criterion of data normality is a very important consideration in parametric statistical analysis because it affects the accuracy of hypothesis testing results. Normality testing is carried out to ensure that the score data obtained from the research instrument have a normal distribution before proceeding to further analysis using parametric statistical techniques. Data are considered to be normally distributed if the significance value obtained from the normality test is higher than the predetermined significance level (usually 0.05), so that the data meet the normality requirement and are ready for deeper analysis. (Rostampour & Azarmi-Atajan, 2023). Then there are testing criteria in the Normality Test, namely:

- a. Sig Value (P Value) $< 0,05$ concluded that the data is not normally distributed
- b. Sig Value (P Value) $> 0,05$ concluded that the data is normally distributed

6. Homogeneity test

According to Sugiyono (2020), homogeneity test is a statistical method used to determine whether two or more groups of data have the same or homogeneous variance. Usually, this test is performed as an initial step before applying other statistical tests, such as the t-test or ANOVA, because some parametric tests require equal variances among data groups.

In general, homogeneity indicates that the distribution of data in each group is relatively balanced. If there are no significant differences in variance between groups, then the data can be considered homogeneous. Use the following guidelines to assess homogeneity:

- a. Significance test (α) = 0.05
- b. If $\{sig\} > \alpha$, then the variance of each sample is considered the same (homogeneity)
- c. If $\{sig\} < \alpha$, then the variance of each sample is different (not homogeneity). To assess this, the researcher uses SPSS.

7. T-test

The t-test is a statistical technique used to compare the mean values of two groups or two different conditions with the aim of determining whether the observed difference is statistically significant (Pallant, J. 2020). Data analysis was carried out using the t-test with the aid of SPSS software. This test aims to determine the significance level of the partial effect of an independent variable on the dependent variable, assuming that the other independent variables are held constant. The decision-making criteria in this test are as follows:

- a. The independent variable is said to have a significant effect on the dependent variable if the significance value is less than 0.05
- b. The independent variable is said to have no effect on the dependent variable if the significance value is greater than alpha 0.05

8. F-test

George W. Snedecor presented this statistical method as the F-test in honor of its creator, Sir Ronald Fisher, who was a leading expert in the field of statistics. The F-test became one of the vital techniques in statistical analysis used to compare two or more variances, which reflect the degree of data variation among different groups (Odek & Opuodho, 2023). One of the most common uses of the F-test is in Analysis of Covariance (ANCOVA). ANCOVA

is used to compare the means of two or more groups while controlling for one or more variables that act as covariates and may influence the study outcomes. The f-test is very useful in experimental research, especially for evaluating the effectiveness of a particular treatment or technique.

9. The Effect Size

Effect size is a tool used to describe how large an effect or the strength of the relationship between two variables is in a study. The importance of effect size lies in its ability not only to indicate the presence of an effect but also to describe the level or intensity of that effect. Some commonly used effect size measures include Cohen's d, which is used to measure the difference in means between two groups, and Pearson's r, which serves to evaluate the strength of the relationship between two variables (Funder & Ozer, 2019). The following is the category of effect size values used to assess how large the influence or association between variables is in research :

Table 3.12 Value of the effect size

No	Cronbach's Alpa	Validity Level
1	≥ 0.01	Small
2	≥ 0.06	Medium
3	≥ 0.14	Large

Source : Jacob Cohen, 1988

10. Statistical Hypothesis

Hypothesis testing is carried out to make decisions regarding the acceptance or rejection of the null hypothesis (H_0) and the alternative hypothesis (H_a). This testing process can be conducted using SPSS version 27 software as an aid in the analysis. The following are the basic criteria used to determine the results of hypothesis testing:

- a. If the Sig. (two-tailed) value is less than 0.05, then the hypothesis is considered accepted
- b. If the Sig. (two-tailed) value is greater than 0.05, then the hypothesis is considered rejected

