

CHAPTER III

RESEARCH METHOD

A. Research Design

A quantitative approach was applied in this study. Creswell & Creswell (2012) describe quantitative research as a systematic, objective way of gathering and analyzing numerical data to obtain reliable, relevant information about a specific phenomenon. This approach fit the study since it focuses on measuring and comparing students' numeric speaking scores before and after treatment, across both groups.

This study applied a quasi-experimental design, which typically requires at least two groups: one functioning as the experimental group and the other as the control group (Arib et al., 2024). In quasi-experimental research, treatment is applied to the experimental group, but researchers cannot fully control outside variables or fully randomize class assignment. Creswell & Creswell (2018) define it as a design where researchers have only partial or no control over randomly assigning participants to different levels of the variable being studied.

For this research, Class VIII A served as the experimental group, receiving instruction through Stimuler across three meetings centered on picture-description tasks, whereas Class VIII B acted as the control group, taught using the conventional method. Through this design, the researcher

could compare shifts in students' speaking ability before and after the intervention, allowing an assessment of how effective Stimuler is as a digital learning tool.

Dannels (2018) notes that research design shapes participant selection, variable manipulation, data-collection and analysis methods, and how external variability is controlled all working together to answer the research problem. This research adopts a quantitative quasi-experimental approach based on a pre-test/post-test control-group format. The choice of a quasi-experimental design was made because the classes had already been determined by the school, so researchers were unable to carry out random assignment. The design aims to reveal whether applying Stimuler as a digital learning innovation makes a real difference to students' speaking ability.

In practice, the experimental group received instruction through Stimuler, while the control group followed conventional teaching. Both groups completed pre-tests and post-tests to monitor changes in speaking ability throughout the treatment. To keep the findings valid, both classes were run under comparable conditions. The same teacher taught both classes, and content, session length, and learning goals stayed consistent across them. The one difference was the use of Stimuler in the experimental class, while the control class was taught conventionally.

This research was carried out at SMP Negeri 11 Bengkulu Tengah, a state junior high school located in Central Bengkulu, during the second semester of the 2025–2026 academic year, spanning April to May 2026

Table 3.1 Non-equivalent control group design scheme.

Class	Pre-test	Treatment	Post-test
Experiment	O_1	X	O_2
Control	O_3	-	O_4

Description:

X : Treatment using Stimuler application

O_1 / O_3 : Pre-test of speaking skills

O_2 / O_4 : Post-test of speaking skills

For the sake of validity, the experimental and control classes were held under comparable conditions.

The same teacher taught both classes, and the material, session length, and learning goals remained identical across both groups. The only difference among the two classes was the use of the Stimuler application within the experimental class, whereas the control class received instruction through conventional methods.

B. Population And Sample

All eighth graders at the school made up the population, from which the sample was drawn.

1. Population

According to Candra Susanto et al. (2024), a population encompasses the entire group of individuals, objects, or events that become the focus of a study, functioning as the complete collection from which generalizations can be made. Its elements consist of every subject to be measured the objects the study examines. Sugiyono (2019) frames population as the generalization scope covering objects or subjects with certain traits and quantities, set by the researcher, from which conclusions are eventually drawn.

The population of this study includes the entire group of grade-eight students at SMP Negeri 11 Bengkulu Tengah. This school has two eighth-grade classes, namely VIII A and VIII B, each consisting of 20 students, bringing the total population to 40 students

Table 3.2 The Research Population

No	Classes	Population
1	Class VIII A	20
2	Class VIII B	20
Total		40 Students

Source: SMPN 11 Bengkulu Tengah, 2026/2027

2. Sample

This study used purposive sampling, selecting classes with similar academic characteristics. As stated by Sugiyono (2019), a sample is a smaller group that reflects the size and traits of the entire population. A sample is a segment of the population selected for study. Working with a sample lets researchers generalize about the population more efficiently and cheaply (Candra Susanto et al., 2024).

This sampling technique fit the current study since both classes shared comparable academic performance, class size, and English lesson schedules. Class VIII A was designated as the experimental group and received treatment through Stimuler, while Class VIII B served as the control group, taught with traditional methods.

The population comprised all eighth graders at SMP Negeri 11 Bengkulu Tengah. There are two eighth-grade classes at this school: Class VIII A and Class VIII B. Class VIII A has 20 students, while Class VIII B has 20. Therefore, the population in this research totaled 40 students

Table 3.3 The Research Sample

Class	subject of the student	Men	Wom en	Number of students
1.	8A	16	4	20
2.	8B	15	5	20
Total				40

Source: SMPN 11 Bengkulu Tengah , 2026/2027

The chosen classes were fairly comparable in English proficiency, learning environment, and study conditions. Both classes were taught by the same teacher using similar materials, making them comparable prior to the treatment

C. The Research Instrument

In order to obtain precise and quantifiable information on students' speaking ability before and after the treatment, this research employed specific instruments. As explained by Donoboyeva (2025), constructing and validating research instruments is essential for producing accurate, dependable, and meaningful data within educational research. Both a speaking test and an assessment rubric were used for pre- and post-test measurements, keeping the process consistent and valid.

1. Speaking Test

The speaking test functions as the primary instrument for measuring students' speaking performance and was given twice, as a pre-test and post-test, both using the same format to ensure the comparison remained valid.

The speaking test consists of:

a. Picture description

It was chosen since it effectively measures fluency, pronunciation, accuracy, and confidence when speaking. Describing pictures pushes students to voice ideas based on visual prompts, letting them practice speaking more naturally and meaningfully.

2. Field Notes (Classroom Observation)

Another instrument used was field notes, which recorded additional information on how students behaved and participated during lessons. These notes were important because they provided richer context beyond numerical data, illustrating how the learning process unfolded in both groups.

Field notes were used to systematically capture data on different aspects of student activities during the intervention. These aspects covered students' levels of participation, attention, motivation, and self-confidence

while engaging in speaking activities. The researcher also observed how students responded to the instructional materials and teaching methods used in the classroom.

In addition, several behaviors were observed, such as student preparedness (for example, whether they brought learning materials such as dictionaries), recording their attendance at each session, their active participation in speaking exercises, and their willingness to perform speaking tasks in front of the class. The researchers also noted whether students showed enthusiasm, hesitation, or a lack of engagement during learning activities.

Observation activities were carried out in both the experimental and control classes, with the researcher acting as a non-participant observer during the entire teaching process. Each session's observations were recorded descriptively as field notes to make sure no relevant details were overlooked.

Rather than being processed statistically, the field-note data functioned to support the results obtained from the speaking tests. This method allowed the researcher to develop a deeper understanding of student learning behavior and classroom dynamics over the course of the study.

3. Speaking Assessment Rubric

These assessment guidelines are employed to assess students' speaking ability based on four main aspects, namely:

a. Fluency

The ability of students to speak fluently without many pauses or hesitations.

b. Accuracy

The appropriate application of grammar and vocabulary.

c. Comprehension

The ability to understand something, especially written or spoken information.

The rubric in this study was adapted from Brown (2004), whose three assessment aspects fluency, accuracy, and comprehension were adopted. Fluency here means being able to speak smoothly and continuously. Accuracy indicates the level of correctness, or how close speech is to the ideal standard. Comprehension covers how well students grasp and accurately convey meaning during speaking activities. These criteria were chosen because they suit measuring speaking ability at the junior high level

Table 3.4 Speaking Assessment Rubrics

No	Criteria	Score	Category	Description
1	Accuracy	5	Excellent	Pronunciation is clear and natural with almost no errors.
		4	Good	Overall, pronunciation is clear with only occasional 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: Adapted from (Brown, 2004). *Language Assessment: Principles and Classroom Practices*. Pearson Education.

4. Instrument Validity

Instrument validity refers to how precisely an instrument measures its intended target. This study applied content validity to make sure the speaking test genuinely captured students' speaking ability. Both the test and rubric were built on Brown's (2004) speaking-assessment theory, covering accuracy, fluency, and comprehension. These elements were chosen for aligning closely with the research objectives and being considered important components in assessing students' speaking performance.

To ensure the appropriateness of these instruments, the speaking test, scoring rubric, and scoring criteria were reviewed by an expert. The expert checked task relevance, scoring appropriateness, instruction clarity, and whether language level fit eighth-grade students. Based on the expert's input, the instruments were deemed appropriate for measuring students' speaking skills and are therefore

suitable for use in this study. The expert validation sheet is included in the Appendix

5. Instrument Reliability

Reliability describes how steady and consistent the results produced by an instrument are. An instrument is deemed reliable if it yields comparable scores when applied under similar conditions. Because the assessment of speaking ability involves subjective evaluation, establishing the reliability of the assessment process is crucial for minimizing assessment bias and ensuring inter-rater consistency.

For this research, two raters independently assessed students' speaking performance using an identical scoring rubric. The two raters evaluated students based on the same assessment criteria: accuracy, fluency, and comprehension, to ensure consistency in the assessment. The consistency between the scores given by the two raters was tested through inter-rater reliability.

A reliability analysis was performed utilizing Cronbach's Alpha in SPSS to assess the level of agreement between the two raters. Cronbach's Alpha was used because it measures the internal reliability of scores assigned by different raters using the same rating criteria. According to Sugiyono (2019), an instrument is considered to have good reliability when the Cronbach's Alpha value

is 0.70 or higher. In this study, a detailed SPSS analysis of the reliability is provided in the Appendix.

D. Data Collection Technique

In conducting research, data collection is essential. Data is basically the information needed to investigate research problems after they have been properly designed. Mazhar (2021) points out that without collecting certain information, research cannot be carried out.

Data collection in this study was carried out through a speaking test administered to the research sample, namely students at SMP Negeri 11 Bengkulu Tengah. This test consisted of two phase a pre-test and a post-test while only the experimental group received the Stimuler-based treatment in between.

1. Pre-test

Before applying the Stimuler application as a digital learning innovation, the researcher gave a pre-test to both the experimental and control groups. During this stage, students' were asked to take an individual speaking test in the form of describing pictures to measure their initial speaking ability. Each student was given approximately 3-5 minutes to describe the picture individually. Their speaking performances were assessed in accordance with the aspects of fluency, accuracy, and comprehension. The pre-test aims to obtain basic data on

the students' initial abilities, namely fluency, pronunciation, accuracy, and confidence, which will later be compared with the post-test results after the treatment is given.

2. Treatment

Following the pre-test, the experimental group underwent treatment using the Stimuler application across three sessions, with an emphasis on picture-description exercises. This lesson is designed to provide repeated speaking practice through tasks that involve describing pictures to improve students' fluency, pronunciation, accuracy, and confidence. This learning process was conducted under the guidance of a teacher. Meanwhile, conventional teaching methods were applied to the control class, with no involvement of the Stimuler application.

3. Post-test

Once the treatment sessions ended, a post-test was administered to both groups. During the post-test, students carried out speaking tasks similar in format to the pre-test, namely describing pictures. Each student was given approximately 3-5 minutes to describe the picture individually. Their speaking performances were evaluated according to the following aspects: fluency, accuracy, and

comprehension. The post-test was conducted to determine the extent to which students' speaking skills changed after using the Stimuler application as a learning medium. The assessment instruments and format for the post-test were identical to those used in the pre-test to ensure a valid comparison of data.

Once the pre-test and post-test data had been gathered, the researcher performed a descriptive statistical analysis, presenting students' speaking scores in numerical tables to illustrate the change in speaking ability before and after treatment. This analysis process is anticipated to provide a clear understanding of the Stimuler application in improving students' speaking skills and yield valid data to substantiate research findings related to the use of digital learning innovation

E. Data Analysis Technique

This part explains the analytical procedures used in this research, covering descriptive statistics, inferential statistics, and the assumption tests needed before hypothesis testing could be carried out. These techniques are used to determine how effective the Stimuler application is on students' speaking abilities.

1. Descriptive Statistical Analysis

Descriptive statistics were applied to summarize the pre-test and post-test results across both the experimental

and control groups, covering values such as minimum and maximum scores, mean, median, and standard deviation. Descriptive analysis can be presented in the form of tables, histograms, bar and leaf diagrams, and box diagrams. This analysis aims to provide an initial overview of the increase in students' skills following the use of the Stimuler application.

2. Inferential Statistical Analysis

Inferential statistics were applied to evaluate the research hypothesis, specifically to determine whether the implementation of Stimuler produced a meaningful effect on the speaking ability of Grade VIII students. The analysis was performed using a t-test (independent samples t-test), conducted using SPSS version 27.

The criteria for decision-making in hypothesis testing are as follows: when the significance value (Sig.) falls below 0.05, the alternative hypothesis (H_a) is supported while the null hypothesis (H_0) is rejected, indicating a meaningful effect from Stimuler's use on students' skills. On the other hand, if the significance value exceeds 0.05, the null hypothesis (H_0) is retained and (H_a) is not supported. This indicates no significant effect from the use of the Stimuler application on students' speaking abilities.

To analyze the disparity in posttest scores between the experimental class, which utilized the Stimuler application, and the control class, which was taught using conventional learning methods, the independent samples t-test statistical technique was applied

3. Normality Test

A normality test was carried out to check whether student test scores were distributed normally. In this study, the test of normality was conducted using the Shapiro–Wilk technique through SPSS software version 27.

The decision criteria were set as follows: If the significance value (Sig.) > 0.05 , the data are considered normally distributed. If the significance value (Sig.) < 0.05 , the data are considered not normally distributed. Provided the data are normally distributed, parametric statistical analyses such as the independent samples t-test can be applied.

4. Homogeneity Test

A homogeneity test was conducted to determine whether the variance was equal between the experimental and control groups, as this is one of the necessary conditions before running an independent-samples t-test. The homogeneity test in this study was carried out using Levene's test with the aid of SPSS software version 27.

The decision criteria were set as follows: If the significance value (Sig.) > 0.05 , then the variance is homogeneous. If the significance value (Sig.) < 0.05 , then the variance is not homogeneous. If the data is homogeneous, the independent samples t-test can be performed with the assumption of equal variance

5. Paired Sample T-Test

A paired-sample t-test is applied to compare two related sets of data, typically collected from the same subjects before and after a treatment. This test suits research designs that measure change over time (Kopirana et al., 2023). It compares two sets of interrelated scores collected from the same subjects before and after the treatment.

For this study, the paired-sample t-test was applied separately to both the experimental and control groups to examine whether students' speaking ability improved following instruction. This decision was made by comparing the significance (Sig., two-tailed) with the significance level (α) of 0.05. When the significance is lower than 0.05 (Sig. < 0.05), the null hypothesis (H_0) is rejected, while the alternative hypothesis (H_a) is accepted, which indicates a statistically significant difference between the pre-test and post-test scores. In contrast, when the significance value is above 0.05 (Sig. > 0.05), the null

hypothesis (H_0) is accepted, indicating the absence of a statistically significant difference between the two sets of scores

6. T-Test

An independent-samples t-test was carried out to determine whether there was a meaningful difference in speaking achievement between the experimental and control groups. This analysis was conducted with the help of SPSS software, version 27. An independent-samples t-test was employed to compare the mean post-test scores of the two groups to test the effect of using the Stimuler application on students' speaking skills.

The criteria for decision-making were determined as follows: If the significance value (Sig. 2-tailed) < 0.05 , then the alternative hypothesis (H_a) was supported and the null hypothesis (H_0) was rejected. This condition indicated a difference between the experimental group and the control group. If the significance value (Sig. 2-tailed) > 0.05 , then the null hypothesis (H_0) is retained, and the alternative hypothesis (H_a) is not supported. This condition indicates that no significant difference exists between the two groups.

F. Teaching Procedure in Experimental and Control Groups

This part explains the instructional steps used for both the experimental and control groups throughout the research.

1. Teaching Procedure in Experimental Class

Instruction in the experimental group was carried out across five meetings, which included a pre-test, treatment sessions, and a post-test. Throughout these sessions, the Stimuler application was used as the main learning medium to strengthen students' fluency, accuracy, and comprehension in speaking.

a. Session 1 (Pre-test)

The lesson began with greeting the students and taking attendance. The teacher presented the purpose of the activity. The students were shown a sheet of paper with four pictures on different themes and asked to explain them orally before the class. This activity was conducted individually to assess the students' initial speaking skills. The teacher evaluates students' performance based on fluency, accuracy, and comprehension. No intervention is provided during this session. The lesson concludes with a brief summary.

b. Session 2 (Treatment 1: Introduction and Basic Practice)

The lesson began with greeting the students and reviewing the previous activity. The teacher

explained sets out the learning objectives and provides a brief introduction to the Stimuler application. The students were asked to use the app to practice basic speaking by listening to and repeating sentences. The app provided automatic feedback, and the teacher monitored the students. After the practice session, the students were given pictures with a "family" theme and asked to describe them using basic vocabulary and simple sentences. The teacher provided feedback, and the lesson ended with a summary.

c. Session 3 (Treatment 2: Intensive Practice)

Class started with a quick recap of the earlier session, followed by the teacher's short explanation on pronunciation and how to build ideas while speaking. Students then practiced with the Stimuler application and refined their speaking based on the feedback received. The teacher guides and assists the students.

After that, the students were given a worksheet with blank columns and asked to draw pictures on the theme of "friends," then take turns describing their drawings.

They were encouraged to compose three longer, more detailed sentences. The teacher provided feedback, and the lesson concluded with a summary.

d. Session 4 (Treatment 3: Practice and Development)

Class began with a review of prior material. The students practiced speaking using the Stimuler application to improve their accuracy and fluency. Next, the students were presented with an image of a school and was asked to explain everything in the school, for example, the cafeteria, the library, the parking lot, and so on, in a more structured and confident manner. Each student took turns speaking in front of the class. The teacher evaluated their performance and provided feedback. The lesson concluded with a summary.

e. Session 5 (Treatment 4: Advanced Practice)

The lesson began with greeting the students and practicing the use of the Stimuler application to hone their speaking skills, followed by an explanation of the upcoming task to describe a picture. The students were given a sheet of paper with blank columns and asked to draw a picture on the theme "My Bedroom," that is, anything found in their room, and then asked to describe it orally. The teacher evaluated the students' speaking skills based on fluency, accuracy, and comprehension. The lesson then concluded with closing remarks.

f. Session 6 (Treatment 5: Communicative Practice)

The session opened with a revision of the previously taught material. The students practiced speaking using the Stimuler application to improve their accuracy and fluency and build their self-confidence. Next, students are shown a picture of a beach and asked to describe everything on the beach, such as the ocean, umbrellas, boats, and so on, in a more structured way, using five longer, more detailed sentences delivered with confidence. Each student takes turns speaking in front of the class. The teacher evaluates their performance and provides feedback. The lesson concludes with a summary.

g. Session 7 (Treatment 6: Describing Public Places)

The lesson began with greeting the students and reviewing the previous lesson's material. The students first practiced speaking using the Stimuler application to improve their fluency, accuracy, and comprehension through interactive speaking exercises and automatic feedback. After completing the exercises, the teacher shows a picture of a traditional market. Students are asked to describe the picture orally by naming the people, objects, and activities visible in the image. They are encouraged to produce at least five complete sentences using appropriate vocabulary and grammar. The teacher provides feedback on the students'

speaking performance, and the lesson concludes with a summary.

h. Session 8 (Treatment 7: Daily Activities)

The lesson began with greeting the students and reviewing the previous material. The students practiced speaking using the Stimuler application by completing several speaking activities and improving their performance based on the feedback provided by the application. Afterward, the teacher shows a picture depicting daily activities. The students were asked to describe the sequence of activities shown in the picture using complete and well-structured sentences. Each student presents their description orally in front of the class, while the teacher evaluates their speaking performance based on fluency, accuracy, and comprehension. The lesson concludes with a brief overview of the material.

i. Session 9 (Treatment 8: Free Speaking Practice)

The lesson began with greeting the students and reviewing the previous lesson's material. The students practiced speaking using the Stimuler application to strengthen their speaking skills and boost their confidence ahead of the final assessment. Next,

students choose one image from several provided by the teacher, such as a park, a zoo, a classroom, or a restaurant. They are asked to describe the selected image orally using at least five complete and detailed sentences. The teacher provides feedback and encourages students to speak more fluently and confidently. The lesson concludes with a summary and preparation for the final test.

j. Session 10 (Post-Test)

The lesson began with greeting the students and explaining the purpose of the final exam. The students were given a sheet of paper containing four pictures with different themes and asked to give an oral description of them in front of the class. The format of this assignment was similar to that of the initial exam to ensure a valid comparison. The teacher evaluated the students' speaking skills based on fluency, accuracy, and comprehension. The lesson ended with closing remarks.

2. Teaching Procedure in Control Class

Learning in the control group also spanned five sessions, but instruction relied on conventional teaching techniques and did not involve the Stimuler application. The learning activities were designed to align with those in

the experimental class, particularly regarding speaking tasks, but without the use of digital media.

Every meeting opened with greetings and attendance, after which the teacher stated the learning objectives and introduced the topic for the day. Content delivery relied on textbooks and the whiteboard, with the teacher demonstrating pronunciation and common spoken expressions.

After that, students practiced speaking based on the teacher's instructions. They were given speaking tasks involving describing pictures, similar to those used in the experimental class. The teacher monitored student performance and provided feedback on fluency, accuracy, and comprehension.

At the end of each session, the teacher concluded the lesson with a summary