

## **CHAPTER III**

### **METHODOLOGY**

#### **A. Research Design**

This research uses a quantitative method with an experimental design. Creswell (2015) states that quantitative research is a scientific approach used to test hypotheses by examining the relationship between variables through systematic, objective, and measurable procedures. This approach emphasizes the use of numerical data and statistical analysis to ensure that the findings are reliable and valid. Furthermore, experimental research aims to test ideas, practices, or procedures to determine their effect on outcomes or dependent variables, as well as to identify cause-and-effect relationships between independent and dependent variables. In addition, experiments are used to determine potential cause and effect between independent and dependent variables. In this context, the independent variable refers to the treatment given to the participants, while the dependent variable represents the outcome measured after the treatment is applied. At least two groups are used in this type of design; therefore, this study applies a quasi-experimental approach, where participants are divided into two groups, namely the treatment group and the comparison group (Arib et al., 2024). The treatment group receives a specific intervention, while the comparison group is taught using conventional methods. By comparing the results

of both groups, this research aims to determine whether the given treatment has a significant effect on students' outcomes, particularly in improving their speaking ability.

**Tabel 1.**

**Pre-Test Post-Test Nonequivalent Control Group Design**

| Class      | Pretest        | Treatment | Posttest       |
|------------|----------------|-----------|----------------|
| Experiment | O <sub>1</sub> | X         | O <sub>2</sub> |
| Control    | O <sub>3</sub> | -         | O <sub>4</sub> |

Source: (Wulanda et al., 2025)

Discription:

Ei : Eixpeirimeintal groiup C : Cointroil groiup

X : Applicatioin Meidia

Oi1 : Preiteist in eixpeirimeintal class

Oi2 : Poistteist oin thei eixpeirimeintal class Oi3 : Preiteist in cointroil class

Oi4 : Poistteist oin cointroil class

**B. Location and Time of Research**

This research was conducted on the tenth grade students of SMAN 04 Seluma which is located on Jl Desa Pinju Layang, Kec Semidang Alas, Kab Seluma. The research was conducted based on a research permit issued by the university, from 15 September to 15 October, 2025.

## **C. Population and Sample**

### **1. Population**

The next step in the research procedure was to determine the subjects or students who will be the population in this study. Population is the generalization area consisting of objects or subjects that have specific qualities and characteristics determined by the researcher to be studied and then conclusions are drawn. Population refers to the entire group of objects or subjects that are the focus of a researcher's attention. It encompasses all units or elements that the research aims to study and analyze in order to draw meaningful conclusions (Suriani et al., 2023).

From the above opinion, it can be concluded that population is an important step in research because it is the basis for determining the research subject. Population includes a group of people who have certain characteristics in a particular place or institution. In educational research, the population can come from students, teachers, schools, or programs that you choose to study. This study was conducted on tenth grade students at SMAN 04 Seluma in the 2024/2025 academic year. The research population consisted of 209 students.

**Table 2. The Population of the Research**

| No    | Class     | Total |
|-------|-----------|-------|
| 1     | Class X.1 | 30    |
| 2     | Class X.2 | 30    |
| 3     | Class X.3 | 31    |
| 4     | Class X.4 | 30    |
| 5     | Class X.5 | 28    |
| 6     | Class X.6 | 31    |
| 7     | Class X.7 | 29    |
| Total |           | 209   |

Source: (SMAN 04 Seluma, 2025)

## **2. Sample**

A sample is a part of a population that is considered to represent that population through a specific selection process. Purposive sampling is a sampling technique used when researchers already have target individuals with characteristics that suit the research needs. This means that researchers deliberately select research subjects based on certain criteria that have been determined in advance, because these subjects are considered capable of providing relevant, in-depth data that is in line with the

research objectives (Susanto & Jaya, 2023). The researcher determined the classes that were deemed most appropriate to be used as the experimental and control groups so that the research results could describe the effectiveness of the treatment given. The research sample consisted of 60 students from classes X 6 and X 7 at SMAN 04 Seluma. In practice, class X 7 served as the experimental class, while class X 6 served as the control class, as shown in the following table.

**Table 3. Sample of the Research**

| No | Class          | Male | Female | The Number of Student |
|----|----------------|------|--------|-----------------------|
| 1  | Class X 6      | 12   | 19     | 31                    |
| 2  | Class X 7      | 11   | 18     | 29                    |
|    | Total Sampling |      |        | 60                    |

Source: (SMAN 04 Seluma, 2025)

The researcher took these classes because the selection was based on several considerations. Academically, both classes have relatively similar initial abilities, are taught by the same teacher, and use the same curriculum. In addition, based on preliminary observations, students in both classes still experience difficulties in using English, such as incorrect use of pronouns and low confidence in speaking.

#### **D. Operational Definition of Variables.**

According to Sugiyono (2013) the operational definition of a variable is a set of instructions that explain in detail what needs to be observed in the process of measuring or testing a variable to ensure its accuracy. The operational definition of a variable also serves to establish rules and steps that researchers must follow so that data collection and analysis activities can be carried out in a more focused, efficient, and consistent manner. The two most common types of variables are dependent variables and independent variables. Dependent variables are influenced by one or more other variables and are usually represented by the symbol Y. Conversely, independent variables influence other variables and are usually represented by the symbol X.

In this study, the independent variable (X) is the Cake application. The dependent variable (Y) is students speaking mastery. These variables are as follows:

##### **1. Speaking**

According to Brown (2001) speaking is a spoken communication process between a speaker and a listener that involves productive ability to produce utterances as well as receptive ability, namely listening with comprehension. In this process, the speaker is not only required to deliver messages verbally, but also must be able to understand the responses from the interlocutor so that

communication can run effectively.

## 2. Cake Application

Cake is a free app created by Cake Corp that was released on March 22, 2018. This app has a lot of features for learning English, such as a speaking tool to practice speaking using dialogue, watching videos to improve speaking skills, and guessing the correct answer and comparing it to help memorize each expression. One of the features is speaking practice that can be used to encourage conversation with native speaker (Redjeki & Muhajir, 2022).

## E. Data Collection Techniques

Data collection techniques are methods used by researchers to obtain information or data in a study, and they are an important part of research methodology. Through this process, researchers gather information based on facts found at the research site so that the data obtained reflects the actual conditions. In addition, data collection techniques are carefully selected and adjusted to the research method used, such as observation, interviews, questionnaires, documentation, and tests. By choosing appropriate techniques, researchers can ensure that the data collected is relevant, accurate, and able to support the objectives and results of the study (Daruhadi & Sopiati, 2024).

In this study, the researcher used pre-tests and post-tests in both the experimental class and the control class because these tests were important for determining the results of the study. The data collection process was carried out in several stages, namely pre-tests,

treatment, and post-tests. The purpose of this process was to determine whether the Cake application was effective in improving students' speaking skills at SMAN 04 Seluma. The pre-test and post-test scores were then compared to see the differences in learning outcomes, as well as to determine the students' speaking skills before the treatment was given. There are three data collection techniques:

#### 1. Test Technique

A test is a collection of questions, statements, or tasks used to obtain information about the skills, knowledge, or other abilities possessed by individuals or groups.

##### a. Pre-test

A pre-test is a test given before the learning process begins to determine the extent of students' mastery of the material to be taught. In the context of this study, the pre-test serves to assess students' initial abilities and as a reference for evaluating the effectiveness of the learning process (Siregar et al., 2023). In this study, the researcher provided 17 pre-test questions in the form of dialogues with missing sentences. Students were asked to fill in the blanks by choosing the most appropriate sentence from the options provided by the researcher. After filling in the missing sentences, students then continued the dialogue in pairs to assess their speaking skills more realistically. This test was given to students in the experimental and control classes so that the results could be compared

with the scores after the treatment was given. In this way, the researchers were able to assess the extent to which the use of the Cake application had an effect on improving students' speaking skills.

b. Post-Test

A post-test is a test given at the end of a program or teaching unit. The purpose of a post-test is to determine the extent to which students have achieved the material they have learned, both in terms of knowledge and skills, after they have undergone the learning process. Thus, post-tests serve to assess student learning outcomes and measure the effectiveness of the treatment or teaching methods that have been provided (Siregar et al., 2023). In this study, the researcher provided 17 post-test questions similar to the pre-test, namely dialogues with blank sentences that students had to fill in by choosing the most appropriate answer. After the students completed the blanks, they were also asked to continue the dialogue in pairs to assess their speaking skills after the treatment was given. The post-test was assessed using the following criteria: a score of 1 for each correct answer and a score of 0 for incorrect answers. The post-test results were then compared with the pre-test scores to see the improvement in students' speaking skills after

using the Cake application.

c. Treatment

Treatment in experimental research is a specific treatment or action given by researchers to the test group to determine its effect. Treatment can be in the form of methods, media, strategies, or other forms of intervention that are deliberately applied so that researchers can observe the changes or results that occur after the treatment is given (Arib et al., 2024). The treatment in this study was given to students in the experimental class through the use of the Cake application as a learning medium to improve speaking skills. Students in the experimental class participated in the learning process by utilizing the Cake application, namely by watching and studying videos that were relevant to the predetermined material. Meanwhile, students in the control class did not receive the same treatment and continued to participate in learning using the conventional methods normally applied by teachers in the classroom.

The treatment in this study was carried out over four meetings, with each meeting lasting 80 minutes, resulting in a total treatment time of 320 minutes. During the learning process, the researcher acted as a facilitator who guided students in using the Cake

application, provided directions on selecting and studying videos relevant to the material, helped students correct their English pronunciation, and gave additional explanations related to the material, vocabulary, and the use of expressions that the students had not yet understood. The treatment in this study functions to introduce an engaging, enjoyable, and technology-based method for teaching speaking skills through the utilization of the Cake application. The goal of this treatment is to determine whether the use of the Cake application can enhance students' speaking ability more effectively compared to the conventional learning methods generally used in the classroom.

## 2. Speaking Test

A speaking test is referred to as a productive skills test because learners are required to produce language directly through speaking activities. In this test, students are not only required to comprehend the language but must also be able to use vocabulary, grammar, and pronunciation appropriately in order to convey ideas and communicate effectively (Tresnadewi et al., 2021).

To help improve students speaking mastery, the researcher conducted a study using the Cake application as a learning medium. The study was carried out with students of classes X 6 and X 7 at SMAN 04 Seluma, who were

selected through purposive sampling. Before the treatment, the research began with a pre-test consisting of 17 questions in the form of dialogues with incomplete sentences. The purpose of this assessment was to determine how much spoken dialogue the students had mastered prior to the treatment. After the pre-test, the students learned using the Cake application for four meetings. The Cake application is a mobile-based learning medium that presents a collection of short, authentic videos originating from real content such as films, TV shows, and vlogs. It is intended to enhance students' speaking skills through pronunciation practice and contextual dialogue simulations. After the students learned using the Cake application, they were given a post-test. In addition, the post-test consisted of the same 17 questions as the pre-test. The purpose of this test was to determine whether the students speaking mastery improved after using the Cake media.

There was only one correct answer for each question. In the scoring process, each student was given a score of 1 for a correct answer and 0 for an incorrect answer. After all students completed the pre-test and post-test, the researcher collected and calculated each student's pre-test and post-test scores. Then, to determine whether there was any improvement, the researchers compared the pre-test and post-test scores. Furthermore, the data were analyzed

using statistical methods to find out whether the Cake application truly helped students improve their speaking skills.

### 3. Documentation

The data collection technique through documentation is a method of collecting data derived from documents or records of events that have occurred previously. The documents used in this study include written sources such as the list of students who became the research subjects, students' score records (tryout, pre-test, and post-test), as well as visual documents in the form of videos and photographs of learning activities. In addition, the documentation also includes the results of the pre-test and post-test on the material provided during the research. All of these documents play an important role in providing accurate and objective information to support the research process, particularly in strengthening students' learning outcome data and ensuring the validity of the research findings (Fauziyah et al., 2021).

### F. Research Instruments,

Research instruments serve to measure research variables systematically and objectively, enabling researchers to obtain accurate, consistent, and reliable data for analyzing research phenomena. Through effective instruments, researchers can ensure that their

data collection process is focused and aligned with the research objectives. In addition, the right instruments help researchers obtain information that truly describes the conditions or behaviors being studied, so that the research results are more valid and accountable (Subhaktiyasa, 2023).

The research instrument used in this study was a test in the form of dialogue with incomplete sentences that reflects the use of expressions in spoken English communication. This test was intended to measure students' speaking ability indirectly, particularly in terms of their ability to understand and complete expressions commonly used in daily conversations, both before and after the implementation of the Cake application. The test was administered in two stages, namely the pretest and the posttest.

Students' answers were scored using a dichotomous system based on the principles of the Guttman scale. The Guttman scale was first introduced by Louis Guttman in 1944, a sociologist and statistician from the United States, and is characterized by the use of two alternative response options (Faizah et al., 2024). In this scale, responses are provided in dichotomous forms such as yes-no, ever-never, or similar categories. Scoring was conducted using two score categories, with a score of 1 assigned to answers

considered correct or indicating a positive response, and a score of 0 assigned to answers considered incorrect or indicating a negative response (Lumuko et al., 2023). The test instrument was developed based on the materials available in the Cake application and was adjusted to the students' level of ability in order to support the achievement of learning objectives.

#### **G. Data Analysis Techniques**

In data collection, instruments play an important role in research as they function as measurement tools and provide information about the object being studied. Research instruments are used to measure observed phenomena as well as to collect the data needed in the study (Khan Mohmand, 2019).

##### **1. Validity**

Validity is used to ensure that measuring instruments (such as tests or questionnaires) actually measure what is intended to be studied, not something else. A test that is declared valid specifically for a particular study will not represent the validity of other studies. An instrument that is declared valid means that it can be used in studies and can add to the quality of those studies (Fadli et al., 2023).

**Table 4. 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       |

(Fadli et al., 2023)

In addition, the researchers analyzed the test results using statistical software, specifically SPSS version 25. The test instrument consisted of 30 items in the form of fill-in-the-blank dialogues. Before being used in the main study, the instrument's validity was first tested. The validity test was conducted using 30 items that had been formulated based on previously determined indicators. The pilot test of this instrument was carried out at Seluma State High School 05, which has the same accreditation level as the research site school, namely Seluma State High School 04. The pilot test involved 20 students as research subjects. After the pilot test was conducted, the data were analyzed to determine the validity level of each item. Based on the results of the analysis, it was found that 17 of the 30 items were valid and suitable for use in the study, in accordance with the predetermined indicators. The items that were not valid were numbers 3, 6, 7, 8, 13, 14, 15, 18, 21, 23, 28, 29,

and 30.

## 2. Reliability

According to Ghozali (2009) reliability is a tool for assessing the extent to which a questionnaire is able to produce consistent answers over time. An instrument is said to be reliable if the responses given by a person to the statement items remain stable when measurements are taken repeatedly under the same conditions. In other words, reliability indicates the level of stability, consistency, predictive ability, and accuracy of a measuring instrument. The higher the reliability value, the greater the confidence that the instrument can produce accurate and reliable data in describing the construct being measured. Therefore, measurements with high reliability are an important requirement in research so that the data obtained truly reflects the actual situation

**Table 5. Test Instrument Reliability Criteria**

| No | Cronbach's Alpa | Reliability Level  |
|----|-----------------|--------------------|
| 1. | 0,00 – 0,20     | Lack of Reliabel   |
| 2. | 0,20 – 0,40     | Some what Reliable |
| 3. | 0,40 – 0,60     | Fairly Reliabel    |
| 4. | 0,60 – 0,80     | Reliable           |
| 5. | 0,80 – 1,00     | Highly Reliable    |

Source: (Ramdani et al., 2023)

Researchers analyzed the reliability value using the SPSS version 25 program. This reliability test was carried out to see whether the research instrument provided consistent results. The calculation uses the Cronbach's Alpha, which measures the extent to which the questions in the instrument are related to each other. The calculation results show the reliability coefficient value (Cronbach's Alpha) as follows:

**Table 6. Instrument Reliability**

| Reliability Statistics |            |
|------------------------|------------|
| Cronbach's Alpha       | N of Items |
| 0,712                  | 30         |
| Reliabel               |            |

In this reliability test, the researchers used 30 dialogue questions with blank sentences and - indicators. The test instrument was administered at SMA Negeri 05 Seluma, which has the same accreditation level as the school that would be the main research location, namely SMA Negeri 04 Seluma. The tryout involved 20 students as test subjects for the test instrument. After the trial was completed, the researchers analyzed the data by calculating the reliability of the tryout results. After processing the reliability data of the trial questions, the tryout reliability

coefficient was 0.712, which, based on the above category, means that the items are considered reliable if they reach a value of 0.600 or higher. Therefore, it can be assumed that this test has a good level of reliability (Rosita et al., 2021).

### 3. The level difficulty test

Analyzing the difficulty level of each exercise or question is an important step in developing technology-based research instruments. Difficulty level analysis is the process of assessing test items based on their level of difficulty so that it can be determined which questions are easy, medium, or difficult (Magdalena et al., 2021). In this study, the Cake application was used as a learning medium to improve students' speaking skills. Difficulty level analysis was conducted to ensure that the speaking exercises provided were in line with the students' abilities, so that the material was neither too easy nor too difficult. Data was collected through speaking exercises given using the Cake application, and then the students' results were analyzed. The number of students who were able to complete the exercises correctly compared to the total number of participants was used to determine the difficulty level of an item. The most commonly used formula is

$$P = \frac{R_h + R_i}{N_h + N_i} \times 100\%$$

Description:

P = Difficulty level in percent

$N_h$  = Number of test takers in high score group

$R_h$  = Number of correct answers in high score group  $N_i$  =

Number of test takers in low score group

$R_i$  = Number of correct answers in low score group  $P = \frac{145}{5} + \frac{65}{5}$

$\times 100\%$

$5 \quad 5$

$P = 42\%$

**Table 7. Interpretation of difficult question categories**

| Question category       | Interpretation of results |
|-------------------------|---------------------------|
| Classified as easy      | 70% ----- 100%            |
| Classified as medium    | 0,30% ----- 0,70%         |
| Classified as difficult | 0% ----- 0,30%            |

Source: (Nurhalimah et al., 2022)

Testing research questions requires a balance in the level of difficulty of test items, so that no item is too easy or too difficult. Based on the calculation of the difficulty level, a value of 42% was obtained. According to the applicable criteria, this value falls into the medium category, because the range of 30% to 70% is classified as a medium level of difficulty. Thus, these questions are appropriate and suitable for use in research.

#### 4. Item discrimination test

Discriminant analysis is the process of assessing test items based on their ability to distinguish between students with low and high abilities. The determination of

discriminant power in the test item analysis program can be seen in the discriminant power table in the DP (%) column. Items with a discrimination index  $\geq 0.30$  are categorized as good, while items with an index below 0.30 are considered poor. Discrimination analysis is useful for improving the quality of each item based on empirical data and for determining the extent to which an item can distinguish between students who have understood the material and those who have not (Magdalena et al., 2021). The most commonly used formula is:

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

Description:

DP = Question discriminating power.

U = Number of correct answers from high score group test takers

L = Number of correct answers from low score group test takers

N = Number of students in each group

$$DP = \frac{29-13}{42} \quad DP = 38,1\%$$

**Table 8. Interpretation of the power difference test**

| T-test                   | Interpretation of results |
|--------------------------|---------------------------|
| $0,70 \leq DP \leq 1,00$ | Very good                 |
| $0,40 \leq DP \leq 0,69$ | Good                      |
| If $0,20 < DP < 0,39$    | Sufficiently qualified    |

|                       |           |
|-----------------------|-----------|
| If $0 < DP \leq 0,19$ | Weak      |
| -                     | Very weak |

Source: (Magdalena et al., 2021).

Discrimination power testing was conducted to ensure that the test administered could effectively determine each student's level of speaking proficiency. The Discrimination Index calculation in this study reached 38,1%. Compared to the qualifying category, which is in the range of  $0.20 < DP (0.38,1) < 0.39$ , it can be concluded that the Discrimination Index in this study is sufficiently qualified.

#### 5. Noirmality teist

Normality test according to Ananda & Fadhli (, as cited in Kamtini & Apriani, 2023) the normality test can be carried out to find out whether the data obtained by the researcher is normally distributed or not. This test is very important to ensure that the research tool works properly and that the research findings are acceptable. Researchers usually use programs such as SPSS to conduct normality tests. In addition, when assessing whether the data is normal.

There are several criteria that must be considered in this :

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

- b. Sig Value (P Value)  $> 0,05$  conclude that the data is normally distributed.

#### 6. Hoimogeinity teist

The homogeneity test is used to determine whether some population variants are equal. The basic assumption of analysis of variance (Anova) is that As a condition for both independent sample t test and Anova analysis, this test is performed. The population variances are equal. The test of equality of two variances checks whether the distribution of data is homogeneous by comparing the two variances. If two or more groups of data have the same level of variance, then the homogeneity test is no longer necessary (Sianturi, 2022). Use the following principles to determine homogeneity:

- a. Significance test ( $\alpha$ ) = 0.05
- b. If  $\text{sig} > \alpha$ , then the variance of each sample is the same (homogeneity).
- c. If  $\text{sig} < \alpha$ , then the variance of each sample is the not same (not homogeneity) to determine it the researcher uses SPSS.

#### 7. T-teist

T-test is a statistical test used to test the truth or falsity of the null hypothesis. To check for significant differences between two groups with small samples (is usually less than 30) and non-normally distributed data, the

t-test uses the t-statistic value resulting from the difference between the means of the two groups and the difference between the sample size and variability of each group (Syafriani et al., 2023). and the formula to be used is as follows.

- a. The independent variable affects the dependent variable uniquely if the significance level is less than 0.05.
- b. The independent variable does not affect the dependent variable if the significance level is greater than (0.05).

8. F-teist

The F test is conducted to determine whether the independent variables together, or simultaneously, affect the dependent variable. This is also done to see the total effect of the independent variables on the dependent variable. The action used is 0.5 or 5%. A significant value of F below 0.05 indicates that the independent variables affect the dependent variable simultaneously or vice versa (Ghozali, 2016).

The F test conditions are as follows:

A significant value of F below 0.05 indicates that  $H_0$  is rejected and  $H_1$  is accepted, which indicates that each independent variable has a significant effect on the dependent variable. In other words, each independent

variable has a significant impact on the dependent variable.

#### 9. The effect size

Effect size is a numerical measure that describes how much influence a variable or treatment has in a study. In quantitative research, this measure can be calculated using statistics such as Cohen's d or eta squared. The effect size is done to determine if there is a significant difference and how large the effect is in practical terms. This is very important because a p-value only indicates whether an effect exists, not how large it is (Nassaji, 2021). The following is the category of effect size values used to measure how much influence or relationship between two variables in research:

**Table 9. Value of the effect size**

| No | Value      | Effect size        |
|----|------------|--------------------|
| 1. | 0-0,1,95   | 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: (Nassaji, 2021)

#### 10. Statistical Hypothesis

One of the branches of inferential statistics is hypothesis testing, which is used to statistically test the truth of a statement and draw conclusions whether the

statement can be accepted or rejected (Maqfiro et al., 2021). Hypothesis testing can also be done using SPSS version 22 with basic guidelines for collecting the following test results:

- a. If the value of  $hig >$  or  $Sig. (2-tailed) < 0.05$  it is said that the hypothesis is accepted.
- b. If the  $hig$  value  $<$  or  $Sig. (2-tailed) > 0.05$  it is said that the hypothesis is reject.

