CHAPTER 3
RESEARCH METHODOLOGY

This chapter details the methodology that was employed to carry out the study. It consisted of seven components of the research process, including the research design, research variables, population and samples, data collection, data analysis, steps of research, and time and place of the research. Further details were provided below.

A. Research Design

This study used quantitative research methods and a quasi-experimental research design with a two-group design. According to Dörnyei (2007), quasi-experimental research involves manipulating an independent variable and measuring its effect on the dependent variable without randomly assigning participants to groups. In this study, two groups were utilized: the experimental group was treated by watching short YouTube videos with bimodal subtitles, while the control group used conventional teaching, which is the use of standard methods usually used by English teachers without the help of bimodal subtitles or video media. This study aims to determine the influence of using bimodal subtitles when watching short YouTube videos on students' listening comprehension.

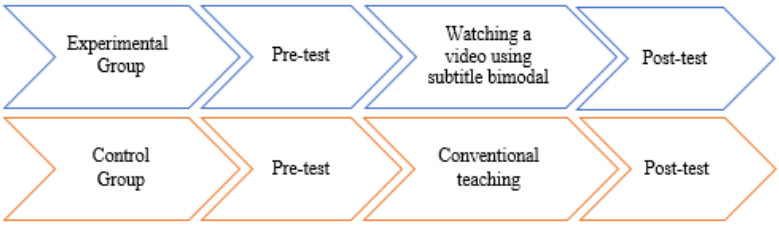

Figure 3. 1 Quasi-Experimental Research Design

B. Research Variables

According to Sugiyono (2023), independent variables are variables that affect the dependent variable, while the dependent variable is the variable that is influenced by the independent variable. This research consisted of two variables, namely the independent variable and the dependent variable.

Bimodal subtitles on short YouTube videos are independent variables because these variables will affect the dependent variable, namely, listening comprehension.

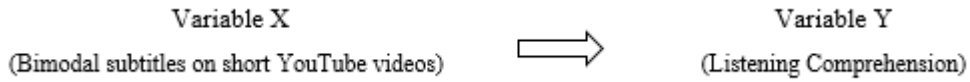

Figure 3. 2 Variables of Research

C. Population and Samples

A population is a collection of objects or subjects with unique characteristics that are highlighted and summarized by researchers in their conclusions (Sugiyono, 2023). The population in this study is all eighth-grade students at a junior high school in Tasikmalaya. The selection of this population was based on appropriate criteria, whereby eighth-grade students are at a stage of cognitive development that can support the application of innovative learning methods, namely bimodal subtitles. This population consists of 10 classes with 360 students. The specifications of this population can be seen in the following table:

Table 3. 1 Students Population

No.	Class	Number of Students
1.	VIII A	36 students
2.	VIII B	36 students
3.	VIII C	36 students
4.	VIII D	36 students
5.	VIII E	36 students
6.	VIII F	36 students
7.	VIII G	36 students
8.	VIII H	36 students
9.	VIII I	36 students
10.	VIII J	36 students
Total Population		360 students

This study used cluster random sampling to obtain samples. A sample is a

group of parts of a population that represents the characteristics of the population (Sugiyono, 2023). In this study, samples were obtained by randomly selecting two classes from the 10 classes as representative samples. Cluster random sampling is a probability sampling technique in which the population is divided into groups (clusters), then a number of clusters are selected at random, and all members in the cluster become research samples (Manikaros, N., 2023). The class selection process was conducted randomly using the lottery method. Each class was assigned an identity number according to the population size, then the numbers were written on small pieces of paper and rolled up. The pieces of paper were then placed in a container and drawn to determine the selected classes. This method ensures that each class has an equal chance of being selected, thereby avoiding sampling bias. The classes selected as samples in this study were as follows:

a. Class 8A with 33 students was selected as the experimental group.
b. Class 8B with 33 students was selected as the control group.

Therefore, the total sample in this study consisted of 66 students in the 8th grade of junior high school.

D. Data Collection

This study used a quasi-experimental design with pre-test and post-test control groups to investigate the influence of bimodal subtitles on short YouTube videos on students' listening comprehension. This design was used to identify the extent to which bimodal subtitles influenced students' listening comprehension. A pre-test was conducted before the treatment to identify students' initial level of listening comprehension. Next, the experimental group received treatment using short YouTube videos with bimodal subtitles, while the control group was taught using conventional teaching methods without videos. The learning material focused on recount texts, with an emphasis on the use of the simple past tense (both regular and irregular verbs) and time connectives. The skills that were trained focused on listening comprehension, including the ability to understand main ideas, recognize specific information, and interpret implied meaning in spoken texts.

In the experimental group, short YouTube videos with bimodal subtitles were used during learning sessions to strengthen students' listening comprehension. In this activity, text and audio were combined to help students process information visually and auditorily, thus increasing their engagement and understanding of the material. Although videos were the main source of learning, textbooks are still used as a supporting resource to reinforce understanding of language structures and vocabulary relevant to the recount text material. Meanwhile, the control group was taught using a conventional teacher-centered lecturing method. Nevertheless, students were still given the opportunity to explore the material through observation, prior knowledge, and brief discussion before the teacher gave the main explanation. The lessons were delivered orally and written on the blackboard, with audio support from textbooks, without the aid of external visual media such as YouTube videos; thus, students relied on notes, teacher explanations, and material from the books. The main difference lay in the more dynamic and authentic approach in the experimental class, which improved listening comprehension more effectively than the control class, which was more passive and limited to the material in the books.

This treatment session was conducted six times, each session lasting approximately 70 minutes, from July 21 to August 8, 2025. After completing the treatment, a post-test was given to both groups to measure changes in their listening comprehension skills. Data for the pre-test and post-test were collected using a standard listening comprehension test, as described in detail below. The data from the pre-test and post-test were analyzed using SPSS version 26, including descriptive statistics, normality tests, homogeneity tests, and independent sample t-tests to determine the effectiveness of the treatment.

The tool used for data collection in this study was a listening comprehension test consisting of 25 items that had passed validity and reliability tests, from a total of 30 items compiled previously (see the Validity and Reliability Test section). This test used a dichotomous scale

(correct/incorrect), with 1 point awarded for each correct answer and 0 points for each incorrect answer. In line with May et al. (2023), the dichotomous assessment model is a practical approach to evaluating student performance in educational assessments using binary responses (scored 0 or 1), where the total number of correct answers determines the final result (e.g., proficiency level) through a simple percentage calculation. This method increases the reliability of test results and allows for more valid comparisons between groups. The use of a binary scoring system facilitates the process of quantifying students' listening comprehension skills, particularly in identifying key details, drawing inferences, and understanding the overall meaning of the audio stimulus provided.

The test instrument is written in English to align with the language skills being measured, while the necessary instructions and explanations are also provided in English to maintain consistency and authenticity in the test context. The test is presented in print and accompanied by audio playback so that all participants have equal access during the pre-test and post-test. Before the instrument is used in the research, validity and reliability tests are conducted to ensure it can measure students' listening comprehension based on three indicators: identifying the main idea (gist), recognizing specific information (details), and interpreting implied meaning (inference).

1. Validity Test

   Validity is a fundamental principle in the development of measurement instruments, which assesses the extent to which test questions actually measure the intended abilities, such as EFL students' listening proficiency in recognizing vocabulary, understanding main ideas, making inferences, and retaining details from audio sources such as YouTube videos (Guion, 2008). According to Khotari (2004), validity is the degree of accuracy of an instrument in measuring what it is supposed to measure, so that the measurement results truly reflect the actual conditions of the object being studied.

   To test the validity of the questions, a pilot test was conducted on

a small sample (32 non-sample respondents) before the main study. Distribute a test with 30 multiple-choice questions based on listening comprehension indicators, then analyze the correlation between each question's score (0 for wrong, 1 for right) and the total test score using SPSS version 26. This process automatically calculates the item-total correlation coefficient to verify whether the questions contribute effectively to the overall measurement.

The validity of each question was determined by comparing the calculated correlation value (r-observed) from SPSS with the critical value from the table (r-table) at the 5% significance level. According to Hair et al. (2019), if r-observed is greater than r-table, the question is considered valid and retained because it aligns with the measured ability. Conversely, if r-observed is smaller, the question is invalid and must be discarded or revised because it does not correlate strongly enough. The r-table value is determined based on the degree of freedom (df = N – 2), where for N=32, df=30, and r-table=0.349. This approach ensures the final instrument. The following is a validity testing table:

Table 3. 2 The Result of the Validity Test

No.	R.Value	R.Table	Criteria
1.	0.688	0.349	Valid
2.	0.825	0.349	Valid
3.	0.126	0.349	Invalid
4.	-0.025	0.349	Invalid
5.	0.082	0.349	Invalid
6.	0.659	0.349	Valid
7.	0.710	0.349	Valid
8.	0.689	0.349	Valid
9.	-0.085	0.349	Invalid
10.	0.616	0.349	Valid
11.	0.624	0.349	Valid
12.	0.631	0.349	Valid

13.	0.641	0.349	Valid
14.	0.670	0.349	Valid
15.	0.625	0.349	Valid
16.	0.640	0.349	Valid
17.	0.569	0.349	Valid
18.	0.615	0.349	Valid
19.	0.670	0.349	Valid
20.	0.738	0.349	Valid
21.	0.672	0.349	Valid
22.	0.636	0.349	Valid
23.	0.535	0.349	Valid
24.	0.585	0.349	Valid
25.	0.175	0.349	Invalid
26.	0.605	0.349	Valid
27.	0.735	0.349	Valid
28.	0.679	0.349	Valid
29.	0.698	0.349	Valid
30.	0.645	0.349	Valid

Based on the results of the validity test of the above instrument, out of a total of 30 multiple-choice items, 25 items were declared valid because their respective r-values exceeded the r-table value of 0.349. These valid items include numbers 1, 2, 6, 7, 8, 10 to 24, and 26 to 30, with r-values ranging from 0.535 to 0.825, indicating a strong and significant positive correlation between the item scores and the total test scores. Meanwhile, 5 items (numbers 3, 4, 5, 9, and 25) were declared invalid because their r-values were below the r-table range (-0.085 to 0.175), indicating a lack of correlation with the overall construct, possibly due to ambiguity or incompatibility among the questions.

2. Reliability Test

Reliability indicates the degree to which a research instrument

consistently produces stable and replicable results when tested under the same conditions (Creswell & Creswell, 2018). In other words, a reliable instrument indicates that the measurements taken are free from random errors and can be trusted to produce similar results in repeated measurements. To determine the instrument's reliability, the researcher conducted a trial with 32 non-sample students using 30 multiple-choice questions based on Nunan's (2003) listening comprehension indicators. The analysis used Cronbach's Alpha, which measures internal consistency by assessing how well each question correlates with the total score (Siregar, 2013). The instrument's reliability was assessed by comparing Cronbach's alpha with the established internal consistency criteria described by Streiner (2003) and cited in Jugessur (2022). The following table summarizes these criteria:

Table 3. 3 Showing the Internal Consistency Value and Significance

	Cronbach's Alpha	**Internal Consistency / Reliability Test**
1.	$\alpha \geq 0.9$	Excellent (High Stakes testing)
2.	$0.7 \leq \alpha$ 0.9	Good (Low Stakes testing)
3.	$0.6 \leq \alpha < 0.7$	Acceptable
4.	$0.5 \leq \alpha < 0.6$	Poor
5.	$\alpha < 0.5$	Unacceptable

If the alpha value meets or exceeds 0.70, the instrument is considered reliable for low-risk educational assessments such as this study, indicating strong item consistency (Hair et al., 2019). Conversely, a value below 0.70 indicates the need for revision to improve stability. This approach ensures that the final instrument, consisting of 25 items, is reliable for measuring students' listening proficiency.

Table 3. 4 The Result of the Reliability Test

Cronbach's Alpha	**N of Item**
.945	25

Based on the reliability test results above, an alpha value of 0.945

for 25 multiple-choice items indicates a very high level of internal consistency and reliability. This value falls into the "excellent" category ($\alpha \geq 0.9$) according to Streiner (2003), as cited in Jugessur (2022), indicating that the items consistently measure EFL students' listening proficiency.

E. Data Analysis

In this study, data analysis was performed using SPSS version 26, including scoring the students' pre-test and post-test, descriptive statistical analysis to describe the data, and inferential statistical analysis to test hypotheses and determine the influence of bimodal subtitles on students' listening skills. For the pre-test and post-test assessments of students, the results can be seen using the following scores:

$$\text{Score} = \frac{\text{Total Correct Answers}}{\text{Total Items}} x100$$

Through this measurement, a total score is obtained that represents the student's overall listening ability, with a percentage that directly and measurably shows their level of achievement. The score is then used to objectively compare student performance before and after the intervention. The test results are then classified according to several criteria to determine the standard of student listening comprehension.

Table 3. 5 Scoring Classification

No	Score	Classification
1	90-100	Excellent
2	80-89	Very Good
3	70-79	Good
4	60-69	Fair
5	0-59	Low

(Source: Luis Villabos)

Descriptive statistical analysis involves calculating the mean, median, mode, standard deviation, range, minimum, and maximum of students' listening comprehension scores before and after the test. This analysis aims to provide a general picture of students' ability levels before and after the treatment, as well as to show the overall distribution and variation patterns of

the data between the control and experimental classes. Through this descriptive statistical analysis, researchers can determine the extent of improvement in learning outcomes after the treatment, compare differences in average abilities between groups, and examine the distribution of data to assess the consistency of student learning outcomes.

Furthermore, inferential statistical analysis. Inferential statistics are used to test hypotheses and determine whether bimodal subtitles significantly influence students' listening comprehension. In this study, the independent-samples t-test was used for inferential analysis. This test compares post-test results between the experimental and control classes. According to Chicco (2025), the independent sample t-test is a parametric statistical method used to compare the mean differences of two independent (unrelated) groups that are statistically significant. However, before conducting the t-test, normality and homogeneity tests must first be carried out as basic requirements for parametric analysis. The results of the normality and homogeneity tests are as follows:

1. Normality Test

   The normality test was conducted using the Shapiro-Wilk test to examine whether the data were normally distributed. The Shapiro-Wilk test is more suitable for small to medium sample sizes ($n < 50$) because it provides greater statistical power in detecting deviations from normality (Ghasemi & Zahediasl, 2012). The results of the test are presented in the following table:

Table 3. 6 The Result of Normality Test

Tests of Normality

	Kolmogorov-Smirnov[a]			**Shapiro-Wilk**		
	Statistic	**df**	**Sig.**	**Statistic**	**df**	**Sig.**
Pre-test Control	.176	33	.011	.949	33	.127
Post-test Control	.133	33	.147	.952	33	.151
Pre-test Experiment	.133	33	.145	.955	33	.181
Post-test Experiment	.145	33	.076	.964	33	.326

a. Lilliefors Significance Correction

To determine whether the data were normally distributed, the significance value (Sig.) of the Shapiro-Wilk test was compared to the significance level $\alpha = 0.05$. If the significance value is greater than 0.05, the data are considered normally distributed; otherwise, they are not. The result shows that all Sig. Values for both the control (0.151) and experimental (0.326) groups are higher than 0.05, indicating that the data are normally distributed**.**

2. Homogeneity Test

The homogeneity test was conducted to determine whether the data variances of the two groups were homogeneous. This is the table as a result of the homogeneity test:

Table 3. 7 The Result of the Homogeneity Test

Test of Homogeneity of Variance

		Levene Statistic	df1	df2	Sig.
Result of Listening Comprehension	Based on Mean	.576	1	64	.451
	Based on Median	.590	1	64	.445
	Based on Median and with adjusted df	.590	1	63.191	.445
	Based on trimmed mean	.561	1	64	.456

To check the homogeneity, the significance value based on mean was compared with $\alpha = 0.05$. If the sig based on mean > 0.05, it indicates that the variance of the post-test control and the post-test experiment is homogeneous. Meanwhile, if the sig is based on mean < 0.05, it indicates the data is not homogeneous. The result shows that the significance value is 0.451 > 0.05, indicating that the variances of the post-test scores between the control and experimental groups are homogeneous.

Based on the results of the normality test using the Shapiro-Wilk test, a significance value greater than 0.05 was obtained for all pre-test and post-test data, so it can be concluded that the data is normally distributed. Furthermore, the results of the homogeneity test using Levene's Test showed a p-value of 0.451 (> 0.05), indicating that the variances between the control and

experimental groups are homogeneous. With both assumptions fulfilled, the analysis can be continued using the Independent Sample T-Test to test the research hypothesis regarding the influence of the use of bimodal subtitles in short YouTube videos on students' listening comprehension skills. In this test, the difference between the two groups was considered statistically significant if the significance value (Sig. 2-tailed) was < 0.05, then the alternative hypothesis (H_a) was accepted, and the null hypothesis (H_0) was rejected.

F. Time and Place of the Research

This research was conducted in one of the junior high schools in Tasikmalaya, located at Jl. RAA. Wiratanuningrat No.10, Empangsari, Kec. Tawang, Kab. Tasikmalaya. The research time was carried out within the period of research implementation stated in the following table:

Table 3. 8 Time and Place of the Research

Description	Oct /2024	Nov /2024	Dec /2024	Jul /2025	Aug /2025	Sept /2025	Oct /2025	Nov /2025	Dec /2025
Research Proposal writing									
Research Proposal Examination									
Data Collection									
Data Analysis									
Report									
Thesis Result Seminar									
Thesis Examination									