CHAPTER III
RESEARCH METHO

A. Research Design

The research uses quantitative methods; it provides results that can be used to describe or highlight numerical changes in population characteristics, generalize to other situations that may arise, provide insight into predictions, and explain relationships between cause and effect. Quantitative research a type of research that involves collecting numerical data. This can be done by comparing various variables or evaluating how effective several solutions are. The purpose of quantitative research is to determine how two variables, known as independent variables, and another, known as a dependent variable or outcome, relate to each other in a population. Quantitative research can be designed experimentally or descriptively, with subjects being measured before and after treatment. Because quantitative methods prioritize artistry in research, they are also referred to as artistic methods because the research focuses more on different patterns, so that it can be said to be non-patterned. Quantitative methods produce all research data as interpretations of field data; therefore, this method is also referred to as an interpretive method (Susanto & Jaya, 2023).

The basic philosophy underlying quantitative research is known as positivism and is based on the scientific method of inquiry. Measurement is necessary if the scientific method is to be used. The scientific method provides an empirical or theoretical basis for the analysis of populations and samples (Kraska, 2022). Quantitative research focuses on testing the validity of hypotheses made by researchers to identify causal relationships and laws to predict the phenomenon. Therefore, this research emphasizes controlling the influence of variables that can interfere with the process of identifying causes and laws. Data in quantitative research, which is usually in the form of numerical values, is collected through tests or surveys, and used for statistical. used for analyzing statistics (Pyo et al., 2023).

This study was conducted using a quasi-experimental design with an unequal control group. The experimental and control groups were used to conduct pre- and post-tests. In this study, the method used was quasi-experimental. Quasi-experimental is a prospective or retrospective study involving patients or groups of patients, validity, and analytic techniques (Maciejewski, 2020).

Table 3.1 Non-equivalent control group

Group	Pre-test	Treatment	Post-test
Eksperimental class	O1	X	O2
Control class	O3	O	O4

Notes:

X: Task-Based Language Teaching

O: Not Task-Based Language Teaching

O1: Pre-test on eksperimental class

O2: Post-test on eksperiment class

O3: Pre-test on control class

O4: Post-Test on control class

B. Place on Time of research

To ensure the research runs smoothly and produces valid data, the selection of the research location is an important step that must be carefully considered. The research site should be easily accessible, have subjects that align with the research focus, and be supported by a strong system and environment. The research location should also facilitate researchers in obtaining permissions and supporting accurate data collection. Because this affects the validity of the research results, the chosen location must be of adequate quality. The research process can be conducted more efficiently and effectively in the right place. This research was conducted with seventh-grade students at SMPN 6 Kaur during the academic year 2025/2026, located in the Tanjung Kemuning District of Kaur Regency. Seventh-grade students were selected as the subjects

of the study, with a total population of 192 students for this research.

C. Population and Sample

1. Population

A population is a specific group of things (people, objects, events, etc.) selected by researchers so that the results of their studies or research can be generalized to that group. A population has at least one feature that distinguishes it from other groups (Susanto & Jaya, 2023). A population is a certain group of things (people, objects, events, and so on) selected by researchers whose study or research results can be generalized to that group. A population has at least one characteristic that distinguishes it from other groups. Based on this statement, the population in this study was seventh-grade students. The seventh grade consisted of 6 classes with a total of 192 students. The researcher considers the seventh-grade students as the study population.

Table 3.2 The number of the population

No	Classes	population
1	Class VII A	32
2	Class VII B	32
3	Class VII C	32
4	Class VII D	32
5	Class VII E	32
6	Class VII F	32
	Total	192

Source: SMPN 6 Kaur 2025

2. Sample

The quality of samples taken from the population also determines the quality of the research process, so when determining samples, we must strictly follow the procedures to ensure that the samples truly represent the research population (Susanto & Jaya, 2023). A sample is a smaller subset of a population that shares many characteristics, which makes it easier to manage compared to a larger group (Tanzares et al., 2024). The technique employed is purposive sampling. This study employed a purposive sampling method because this method can be employed directly and collect samples from two classes without randomization. Class A was deemed the experimental class, while Class B was considered the control class.

Table 3.3 The sample of the study

No	Group	Students	Classes
1	Experimental	32	A
2	Control	32	B
	Total	64	

D. Definition of operational

In this study, there are two types of variables: dependent variables, which are affected by other variables, and independent variables, which affect other variables. The following are the types of variables:

1. One of the most important communication skills in the world is the ability to speak. They use it to convey important information about other people. If we speak English, in particular, we must learn and use it when communicating with others in other countries (Crisianita & Mandasari, 2022).
2. Task-based language teaching (TBLT) is a language learning approach that focuses on encouraging students to use the target language in correct and significant ways (Bygate, 2020)

E. Technique of collecting data

Data was collected through guided role-playing tests in which students performed dialogues that had been completed orally.

1. Test

The data collection procedure in this study was carried out by administering speaking tests to the research sample, which consisted of 7th-grade students of SMPN 06 Kaur. The speaking tests comprised two main stages: pre-test and post-test.

a. Pre-test

Before implementing task-based language teaching (TBLT) to improve students' speaking skills, researchers gave a pre-test first. Researchers used the pre-test for students to take a speaking test with material in the form of a role-play type test, which aims to see their initial ability in speaking, and can improve their speaking skills.

b. Post-test (Final test)

After the treatment period was identified as completed, a post-test was administered to the students. Students were asked to do a short presentation, role play, or discussion on a previously given topic (similar to the tasks in the treatment). This test aims to evaluate students' final performance and assess the extent of change after the treatment is applied. The post-test was the same or a similar instrument as the pre-test to ensure the validity of data comparison. Data analysis after collecting data from the pre-test and post-test, the results

were analyzed to answer the research objectives. Data analysis was conducted using descriptive statistical methods, where the students' speaking test results were presented in the form of numbers and graphs to illustrate the changes in their speaking ability before and after the treatment. This provides a clear picture of the impact of the treatment on improving students' speaking skills. This process is expected to produce accurate and valid data to support the research conclusions regarding the effect of the applied teaching methods.

2. Try out

Instrument testing, also known as try-outs, was conducted before the actual data collection. This was done to ensure the validity and reliability of the instruments and to identify possible weaknesses in the questions or instructions. To avoid bias, the testing was conducted at SMP NEGERI 2 KAUR on July 15, 2025. The 16 respondents had characteristics comparable to the target population but were not included in the main sample. The instruments tested included English-speaking skills, and the researchers tracked and recorded the respondents' responses. After that, statistical validity analysis was performed using reliability and instrument validity analysis methods. The results showed how

many items were valid and how many were invalid and needed to be changed or removed.

3. Documentation

The documentation aims to collect direct data from the research site, including relevant books, regulations, activity reports, photographs, documentary films, and research data. This documentation includes books, regulations, activity reports, photographs, documentary films, and other relevant materials. Multimedia is typically recorded, played, displayed, or accessed by information content processing devices, such as computers and electronic devices; however, it can also be part of live performances. Multimedia includes a combination of text, audio, still images, animation, video, and forms of interactive content (Ridwan, 2015). Data was collected from various sources to support the research. This may include photos of teaching and learning activities, as well as other relevant evidence. Additionally, the documentation includes a list of student names in the 7th grade of SMPN 06 Kaur, as well as a list of grades before and after the implementation of the task-based language teaching method.

F. The Research Instrument

Speaking skills, role-playing, are a commonly used approach in the classroom. In role-playing, students were assigned roles and placed in situations they might encounter outside the classroom. Because the roles mimicked real-life situations, this helped students develop real speaking skills by imagining and acting as other people (Zyoud, 2016). Role-playing tests, conducted orally, are the instruments of this study. Then, for about four minutes, the pairs read the conversation aloud in front of the class. The way students read the dialogue aloud determined their speaking ability. Therefore, students work in groups and take turns playing roles during the role-play process. After that, students received a short dialogue containing clues with missing words for each dialogue question. They then used the clues to complete the missing words according to the context. To assess each pair's speaking performance, after the dialogue was complete, each pair took turns reading it aloud in front of the class. According to JB Heaton, comprehensibility, accuracy, and fluency were the three standards that must be met to assess a person's speaking skills. Even though the dialogues do not have missing words, students were assessed based on comprehensibility, accuracy, and fluency in speaking. Therefore, this tool did not measure writing skills, but speaking skills.

G. Technique of data analysis

Before computing the value of the test to test the hypothesis, the researcher went through several stages to explore the data.

a. Validity test

In research, the validity factor is an evaluation of how accurate a measure of data or results is. It is defined as how accurately a methodology measures the variable it is intended to measure, that a result is valid, and an unreliable methodology collecting unreliable data may not be valid(Ahmed & Ishtiaq, 2021). If the significance value < 0.05, the item is considered valid. The score of each question will be correlated with the total score. An item is valid if the r-count > the r-table; if the r-count ≤ the r-table, the item is considered invalid. Overall, the data was computed with the aid of computer facilities, SPSS version 26.

Table 3.4 Instrumen speaking Validity

Indikator	Sig.	r Tabel	r Hitung	Keterngan
Item 1	0.000	0.497	0.872	Valid
Item 2	0.005	0.497	0.670	Valid
Item 3	0.205	0.497	0.335	Unvalid
Item 4	0.010	0.497	0.622	Valid
Item 5	0.000	0.497	0.787	Valid
Item 6	0.410	0.497	0.221	Unvalid
Item 7	0.001	0.497	0.745	Valid
Item 8	0.014	0.497	0.599	Valid
Item 9	0.337	0.497	0.257	Unvalid
Item 10	0.023	0.497	0.564	Valid

Item 11	0.000	0.497	0.801	Valid
Item 12	0.000	0.497	0.827	Valid
Item 13	0.793	0.497	0.071	Unvalid
Item 14	0.001	0.497	0.731	Valid
Item 15	0.000	0.497	0.828	Valid
Item 16	0.013	0.497	0.604	Valid
Item 17	0.204	0.497	0.335	Unvalid
Item 18	0.006	0.497	0.649	Valid
Item 19	0.101	0.497	0.425	Unvalid
Item 20	0.001	0.497	0.731	Valid
Item 21	0.000	0.497	0.828	Valid
Item 22	0.004	0.497	0.675	Valid
Item 23	0.014	0.497	0.599	Valid
Item 24	0.000	0.497	0.770	Valid
Item 25	0.013	0.497	0.604	Valid
Item 26	0.001	0.497	0.753	Valid
Item 27	0.000	0.497	0.842	Valid
Item 28	0.000	0.497	0.813	Valid
Item 29	0.008	0.497	0.634	Valid
Item 30	0.007	0.497	0.642	Valid

Table 3.5. Validity Test 2 Tailed

Indicators	Sig. (2-tailed)	Compare sig	notes
Item 1	0.000	<005	Valid
Item 2	0.005	<005	Valid
Item 3	0.205	<005	Unvalid
Item 4	0.010	<005	Valid
Item 5	0.000	<005	Valid
Item 6	0.410	<005	Unvalid
Item 7	0.001	<005	Valid
Item 8	0.014	<005	Valid
Item 9	0.337	<005	Unvalid
Item 10	0.023	<005	Valid
Item 11	0.000	<005	Valid
Item 12	0.000	<005	Valid
Item 13	0.793	<005	Unvalid
Item 14	0.001	<005	Valid
Item 15	0.000	<005	Valid
Item 16	0.013	<005	Valid
Item 17	0.204	<005	Unvalid

Item 18	0.006	<005	Valid
Item 19	0.101	<005	Unvalid
Item 20	0.001	<005	Valid
Item 21	0.000	<005	Valid
Item 22	0.004	<005	Valid
Item 23	0.014	<005	Valid
Item 24	0.000	<005	Valid
Item 25	0.013	<005	Valid
Item 26	0.001	<005	Valid
Item 27	0.000	<005	Valid
Item 28	0.000	<005	Valid
Item 29	0.008	<005	Valid
Item 30	0.007	<005	Valid

In this validity test, the researchers conducted the validity test using 30 dialogues with missing questions. The researchers then conducted a trial at SMP Negeri 02 Kaur, a school where the instrument was tested, which was different from the school where the research would be conducted. The trial was conducted with a total of 16 students as test subjects. After conducting the pilot test, the researchers processed the data from the interviews or pilot test questions by calculating the validity of each question. Therefore, the researchers found that out of the 30 questions, a total of 24 questions were valid

Table 3.6. The Instrument Validity Criteria

No	Cronbach's Alpha	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

b. Reliability test

Using the concept of reliability, we can define reliable quantitative research as research that uses methods, instruments, or measures that produce consistent results and the extent to which those results can be replicated if repeated under the same conditions. Assessing reliability allows us to assess how much of the variability in the measured results is due to errors in measurement and how much is due to the actual results (Armour & Williams, 2022).

Table 3.7. Case Processing Summary

Case Processing Summary			
		N	%
Cases	Valid	16	100.0
	Excluded[a]	0	.0
	Total	16	100.0
a. Listwise deletion based on all variables in the procedure.			

Table 3.8. reliability statistics

Reliability Statistics	
Cronbach's Alpha	N of Items
.950	30

Table 3.9. Test Instrument Reliability Criteria

No	Cronbach's Alpha	Reliability Level
1.	0,00-0,20	Lack of Reliable
2.	0,61-0,80	Somewhat Reliable
3.	0,41-0,60	Fairly Reliable
4.	0,21-0,40	Reliable
5.	0,00-0,20	Highly Reliable

Based on the table above, it can be concluded that of the 30 questions given to 16 students, 24 were declared valid and suitable for use in pre- and post-tests to be conducted by researchers in the future. This is because Cronbach's Alpha result was 0.950.

c. Difficulty level test

Difficulty classification is an important theory and technique in machine learning that aims to group objects or data into specific classes or categories based on their properties or features. In this research, difficulty classification divides exam questions into three main categories: easy, medium, and difficult. The difficulty classification theory aims to assist instructors and evaluators in understanding the difficulty level of each question, which helps them create fair exams that are appropriate for students' abilities (Siregar et al., 2023). According to Haladyna (2004), the purpose of testing item difficulty is to find out how many students give the correct answer. This can be achieved by knowing how many students work on or choose the correct answer

Difficulty index:

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

Description:

P: Difficulty level in percent

Nh: Number of test takers in the highest score group

Rh: Number of correct answers in the highest score group

Ni: Number of test takers in the low score group

Ri: Number of correct answers in the low score group

Table 3.10. Interpretation of Item Difficulty Test

Score p	Interpretasi
0.00-0.30	Difficult
0.31-0.70	Medium
0.71-1.00	Easy

$$p = \frac{203}{8} + \frac{70}{8} \times 100\%$$

$$p = 34{,}5\%$$

Based on the above calculations, it can be concluded that the difficulty level of each question tested on 16 students with a total of 30 questions can be considered medium, based on the interpretation of the test difficulty level.

Table 3.11. Interpretation of Item Difficulty Test

Question Category	Interpretation
0%-15%	Very Difficult
16%-30%	Difficult
31%-70%	Medium
71%-85%	Easy
86%-100 and	Very Easy

d. Discriminating Power Test

The number of students in the upper group (Ru) and lower group (Rl) who answered the question correctly can be calculated to estimate the item. Questions with differentiating power show how well they distinguish students who master the material from students who do not. You can calculate the discriminating power index analysis by subtracting the number of the lower group who solved the question correctly from the number of the upper group who solved the question correctly, and then dividing by the number of students in each group (Karim et al., 2021).

The following is the formula for computing discriminating power as suggested by Gronlund (1993).

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

Description:

U = is the number of correct responses from high scorers.

L = is the number of correct responses from low scorers.

N = is the number of students in each group (27%)

Table 3.12. Discriminating power

Discriminating power	Interpretation of results
Bila DP ≥+ 0,40	Good
Bila +0,21 <DP <0,40	Enough
Bila 0<DP ≤0,20	Poor

$$DP = \frac{\mathbf{203}}{\mathbf{8}} - \frac{\mathbf{70}}{\mathbf{8}} : \mathbf{30}$$

DP: 0,56

Table 3.13. Interpretation of Item Differentiability Test

Differentiability Test	*Interpretation*
0,00 ---------- 0,20	Poor
0,21 ---------- 0,40	Enough
0,41 ---------- 0,70	Good
0,71 ---------- 1,00	Very Good

e. Normality test

The data normality test is an important component of using the normality test to determine whether the dependent and independent variables have a normal distribution. The test results indicate that the data is normally distributed, or the regression model looks normal, and researchers use the T-test to determine whether the independent variable partially or individually affects the dependent variable. In other words, the T-test aims to test whether the estimated parameters (regression

coefficients and constants) can predict whether the multiple linear regression equation/model is the right parameter. This test can explain the independent variables that affect the dependent variable. The estimated parameters include the intercept (constant) and slope (coefficient) (Alita et al., 2021). Normality tests ensure that statistical assumptions are met by ensuring that the data collected follows a normal distribution. For this test, SPSS was used to check the normality of the pre-test and post-test data. If the data is normally distributed, parametric statistical tests such as the T-test can be used. This test was conducted with SPSS 26.

Table 3.14. test of normality

Tests of Normality						
	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Pre-Test Experiment	.139	32	.119	.957	32	**.228**
Post-Test Experiment	.145	32	.086	.956	32	**.219**
Pre-Test Control	.142	32	.100	.962	32	**.319**
Post-Test Control	.141	32	.107	.962	32	**.307**
a. Lilliefors Significance Correction						

1. The significance level for the pre-test of the experimental class data was 0.228, which was higher than the significance level of 0.05. Based on the decision criteria, it was inferred that the pre-test of the experimental class data was normally distributed because the significance value was greater than 0.05. This indicated that the pre-test results in the experimental class were normally distributed and that there were no deviations from normality.
2. The significance level for the post-test class was 0.219, which was also higher than 0.05. Thus, it was inferred that the post-test class experimental data had a normal distribution. In other words, the students' test results at the end of the experiment showed a normal distribution, which met one of the fundamental assumptions of parametric statistical analysis.
3. In contrast, the significance level for the pre-test control class data was 0.319, which was greater than 0.05. Based on the aforementioned results, it was inferred that the pre-test control class data were normally distributed.
4. With a significance value of 0.307, which was greater than 0.05, the post-test data for the control class were also normally distributed. This indicated that the final test results for the control group were normally

distributed and did not deviate from the assumption of normality.

f. Homogeneity test

The homogeneity test tests whether the sample populations have equal variances, which is essential for making accurate comparisons between the experimental and control groups. Before conducting a t-test, you must ensure that the data is normally distributed (normality test) and the variance between groups is homogeneous (independent t-test).

Table 3.15. Interpretasi Levene's test

Sig. (p-value)	Conclucion
≥ 0,05	Homogeneous (same variance)
≤0,05	Is not homogeneous (different variance).

This test was conducted using SPSS 26

g. T-test Independent

One statistical analysis method is the independent t-test, which is used to compare the means of two different and unrelated groups. In this study, the independent t-test was used to determine whether there was a significant difference between the experimental group and the control group after being given various treatments. In other words, the independent t-test helps researchers determine whether an approach, medium, or learning

strategy actually affects student learning outcomes or whether changes in scores are merely coincidental. The following are the results of independent calculations for the experimental class and the control class.

h. Effect Size

To describe how large the effect of a variable or treatment is in a study, the effect size is used. In quantitative research, this measure can be calculated using statistics such as Cohen's d or eta squared. The effect is measured to determine whether there is a significant difference and how large the effect is in practice. The following effect size values are used to measure the influence or relationship between two research variables:

Table 3.15. Value of the effect size

No	Value	Effect size
1	0,00-0,1,95	Very weak effect
2	0,20-0,3,95	Weak effect
3	0,40-0,5,95	Modest effect
4	0,60-0,7,95	Strong effect
5	0,80-1,00	Very strong effect

Adapted by (Cohen, 2013)

Cohen's d	Effect Size Category
0.01 – 0.19	Very small
0.20 – 0.49	Small
0.50 – 0.79	Medium
0.80 – 1.19	Large
1.20 – 1.99	Very Large
≥ 2.00	Huge effect

H. Statistical analysis

Researchers want to find out whether the task-based language teaching method has an influence on students' speaking skills. The authors also calculated data from the experimental and control classes using the following T-test formula.

a. The null hypothesis (Ho) and the alternative hypothesis (Ha) are accepted, indicating significant effects of the speaking skills method on students.

b. The null hypothesis (Ho) is rejected, while the alternative hypothesis (Ha) is accepted, indicating that the influence of task-based language teaching on students' speaking skills is significant.