

CHAPTER III

RESEARCH METHODS

A. Research Design

This study utilizes a quantitative research method with quasi-experimental design. Quantitative methods are research approaches that rely on numerical data obtained through measurement and observation processes to explain and describe phenomena objectively. Taherdoost, (2022) explains that quantitative research relies on numerical data derived from observation to represent the phenomenon under investigation, and it uses empirical assessment to gauge how far particular standards, norms, policies, or programs have been fulfilled. In addition, quantitative methods aim to identify patterns and trends, calculate average values, predict results, test cause-and-effect relationships, and generalize research findings to broader groups or populations (Pandey et al., 2023). Therefore, this method is considered appropriate for use in research that focuses on measuring the influence or impact of treatment on certain variables in a systematic and measurable manner.

This study adopted a quasi-experimental design since the researcher was unable to randomly allocate participants, meaning the classes used were already pre-formed groups. a quasi-experiment as a research approach that is similar to a true experiment but does not fulfill every criterion of one, particularly regarding the random

assignment of subjects (Singh, 2021). This design is used when the researcher does not have full control over the assignment of subjects to the treatment group or control group. This study employed a quasi-experimental approach, specifically a none equivalent control group design, in which participants were divided into two group, experiment and control without random assignment. Prior to any treatment, both group completed a pre-test aimed at measuring students' initial listening ability. Following this, the experiment group received the intervention, namely the use of Listening English Podcast (LEP) in listening instruction, whereas the control group continued to receive conventional teaching without such intervention. Once the treatment phase concluded, a post-test was administered to both groups in order to examine to extent to which learning outcomes different as a result of the treatment. The pattern of the nonequivalent control group design applied in this research is illustrated below:

Table 3.1 Pre-Test Post-Test Nonequivalent control group design

Experiment	O ₁	X	O ₂
Control	O ₃	-	O ₄

Description:

- E : Experimental Group
- K : Control Group
- X : Application Media
- O₁ : Pretest In experimental class
- O₂ : Posttest In experimental class
- O₃ : Pretest In control class
- O₄ : Posttest In control class

B. Location and Time of Research

This study will take place at SMPN 11 in Bengkulu Tengah, a site selected because it corresponds with the research aims specifically, examining English learning at the junior secondary level and because the school is supportive of the research being conducted there. In addition, the selection of the research location was based on the consideration that seventh grade students at SMPN 11 in Bengkulu Tengah, still face several obstacles in listening skills, so the use of alternative learning media such as English Podcast (LEP) is considered relevant to study. This research will be conducted in the even semester of the 2025/2026 academic year, specifically on April-May 2026. The research period covers several stages, namely preparation of research instruments, implementation of pre-tests, administration of treatment, and implementation of

post-tests. The entire research series will be carried out in accordance with the English learning schedule applicable at the school.

C. Population and sample

1. Population

The term "population of interest" refers to the collection of individuals or entities that serve as the central focus of a study, which may include people, pairs, groups, organizations, or comparable units (Hossan et al., 2023). This will be the main focus of the study, and the findings or research will be applied to this group. In this study, the researcher selected a population from the seventh grade at SMPN 11 Bengkulu Tengah.

Table 3.2. Population Of the Research

No	Classes	Population
1.	7A	20
2.	7B	20
Total		40

Source: SMPN 11 Bengkulu Tengah , 2025/2026

2. Sample

A "sampling frame" denotes the set of units from which a sample is drawn; it functions as an operational depiction of the target population namely, the specific group of individuals invited to take part in the research (Hossan et al., 2023). In this study, the researchers took

samples from two classes, consisting of 7A and 7B, at SMPN 11 Bengkulu Tengah.

Table 3.3 Total sample of the Research

Class	subject of student	Male	Female	Number of students
1.	7A	9	11	20
2.	7B	9	11	20
Total				40

Source: SMPN 11 Bengkulu Tengah, 2025/2026

Based on the explanation regarding the population and sample above, it is known that in the 2025/2026 academic year, there were only two Year 7 classes at SMPN 11 Bengkulu Tengah, namely Year 7A and Year 7B. Therefore, this study employed a purposive sampling technique. According to (Esezi Isaac, 2023), purposive sampling is a non-probability sampling technique in which participants are deliberately selected based on specific characteristics relevant to the research objectives. This technique is considered appropriate because the researcher does not randomly assign pupils to groups. Instead, the existing classes were deliberately selected to serve as experimental and control groups in accordance with a quasi-experimental design with unequal control groups. Class VII A was designated as the experimental class and received the intervention using the Learning English Podcast (LEP),

whilst Class VII B served as the control class and received conventional listening instruction. The selection of these two classes was based on accessibility, similar academic levels, and suitability for the research objectives.

D. Operational Definition of variables

This research involves two variables, namely the independent variable is English Podcast (LEP) and the dependent variable in this study is English listening skills.

1. Independent variable (X): English Podcast LEP

In this research, the independent variable is the English Podcast (LEP), which serves as a digital audio-based instructional medium used as the primary material for teaching listening. In the context of this research, LEP refers to English audio podcasts containing conversations or discussions that are implemented during the treatment sessions. The podcasts are delivered to students through digital devices under the teacher's guidance during listening activities. The use of English Podcast (LEP) is applied in several learning sessions as part of podcast-based listening instruction to support students' listening development.

2. Dependent variable (Y): Listening skills

The dependent variable examined here is students' English listening competence, which refers to their capacity to comprehend spoken English covering

the identification of main ideas, specific details, and contextual meaning conveyed through audio content. In this research, students' listening skills are measured using listening comprehension tests administered before and after the treatment. The level of students' listening skills is indicated by the difference between the pre-test and post-test scores, which reflects the improvement in students' listening ability after the implementation of English Podcast (LEP).

1. Test

The main data collection technique in this research is a test, which consists of a pre-test, treatment and post- test.

a. Pre-test

Pre-test is a test conducted to measure students' initial understanding. According to (Yulia Efendi & Astutik, 2024). Pretests are tests conducted before the sample is given treatment, with the aim of assessing students' listening skills and knowledge before using podcast media to improve listening skills. The test will be conducted by providing audio for students to listen to and asking them to answer the questions provided

In this study, a pre-test is administered to the seventh-grade students of Junior High School 11 in Bengkulu Tengah before the implementation of the treatment to measure their initial listening ability. The

pre-test aims to identify the baseline level of students' listening skills before they are taught using English Podcast (LEP). The test consists of 30 items in the form of multiple-choice questions. The multiple-choice items measure students' ability to identify the main idea, specific information, and the meaning of vocabulary from the audio. The results of the pre-test function as a basis for comparison to determine students' improvement after the treatment.

b. Treatment

Treatment is the process whereby research samples are given treatment to evaluate the effect of independent variables on dependent variables in research. According to (Purwanto et al., 2020)), research instruments are basically tools used to collect data in research. In this study the treatment was given to students in the experimental class using English Podcasts (LEP) as the main learning medium to improve students' English listening skills. Through the use of English podcasts, students were exposed to authentic spoken English in the form of conversations and discussions. During listening activities, students were directed to focus on understanding the main ideas, specific information, and contextual meaning presented in the audio material. The listening process

was divided into three stages: pre-listening, during listening and post-listening to support student comprehension. Meanwhile, the control class did not receive the same treatment and continued to learn listening skills using conventional teaching methods. In the control class, listening instruction relied on teacher explanations, textbook-based exercises, and limited audio materials without the integration of podcast-based digital media. As a result, students in the control class had fewer opportunities to interact with authentic listening materials.

Table 3.4 Treatment for Experiment and control Classes

Aspects	Experimental Class (Using LEP)	Control Class (Conventional Method)
Learning Objective	To improve students' listening comprehension skills (identifying main idea, specific information, vocabulary, and meaning) through Learning English Podcast (LEP).	To improve students' listening comprehension skills through textbook-based and teacher-led instruction without podcast media.
Teaching Approach	Audio-based learning using Learning English Podcast (LEP) – <i>8 Minute English podcast</i> .	Conventional listening instruction using textbook and teacher explanation
Participants	Seventh-grade students of SMPN 11 Bengkulu Tengah (Class VII A) academic year 2025-	Seventh-grade students of SMPN 11 Bengkulu Tengah (Class VII B) academic year 2025-2026.

	2026.	
Teaching Media	English Podcast (audio), speaker, worksheet (LKPD), printed transcript, whiteboard.	English textbook, printed dialogue text, whiteboard, worksheet.
Number of Meetings	Four meetings (2 x 40 minutes each meeting)	Four meetings (2 x 40 minutes each meeting)

The treatment will be conducted in four sessions, with each session lasting 40 minutes, resulting in a total treatment time of 160 minutes. The use of four treatment sessions will be comparable with previous research on listening practice. Forexample, Yusuf et al., (2026) implemented structured listening practice across four sessions between the pre-test and post-test. In the present study, four sessions will be considered sufficient to provide students with repeated exposure to the Learning English Podcast (LEP) materials while remaining consistent with the available instructional schedule at the school. The same amount of instructional time will also be maintained between the experimental and control classes to ensure comparable treatment conditions.

During the treatment, the researcher will guide students in accessing and listening to the English podcast materials, assist them in understanding

unfamiliar vocabulary, and provide explanations related to the audio content. Students will also be encouraged to actively participate in listening activities, such as answering comprehension questions and discussing the podcast content. The treatment will be designed to introduce an engaging, interactive, and technology-based approach to listening instruction through Learning English Podcast (LEP). Through these activities, the treatment will aim to examine whether the use of English podcasts can improve students' listening comprehension, particularly their ability to understand main ideas, identify specific information, recognize vocabulary, and make inferences from spoken English, compared with the conventional listening instruction used at SMPN 11 Bengkulu Tengah.

Table 3.5 Meeting 1: Introduce Yourself

Stages	Experimental Class	Control Class	Time
Opening	Before the main act teacher greets the stu attendance, then expla focus is listening to a v self-introduction.	The lesson begins with a greeting and attendance check, followed by the teacher outlining the learning objective related to self-introduction using textbook materials.	10 minutes

Main activities	Pre-Listening The teacher introduced several key vocabulary items and invited students to predict the video's content based on its title. While-Listening Students watch the English podcast video without subtitles, after which their answers are confirmed under the teacher's guidance. Post-Listening Students are asked to identify important expressions commonly used in self-introductions.	Pre-activity: Teacher explains expressions from textbook. Whilst-activity: The teacher reads the dialogue aloud while students repeat it and underline unfamiliar words. Post-activity: Students complete a set of comprehension questions taken from the textbook.	60 minutes
	Closing To close the session, the teacher offers feedback and acknowledges students' participation before ending the class with a prayer.	The teacher goes over the material once more, has students summarize what they have learned, and then dismisses the class.	10 minutes

Table 3.6 Meeting 2: How start Learning English

Stages	Experimental Class	Control Class	Time
Opening	The class open with greeting and a short prayer, following by attendance check and an introduction to the day's topic on strategies for learning English.	Following the greeting, attendance check, the teacher introduction the same topic through explanations drawn from the textbook.	10 minutes
Main activities	Pre-Listening Before listening began, the teacher asked students to share their personal experiences in learning English and introduced key vocabulary such as <i>lazy</i>, <i>habit</i>, <i>strategy</i>, and <i>practice</i>. While-Listening Students watch a podcast video containing tips for learning English, noted down important information, and rewatched the segment for a more detailed understanding. Post-listening Students' identification the main strategies presented in the video and composed a summary consisting of five to seven sentences.	Pre-activity: Teacher explains expressions from textbook. While-activity: The teacher read the dialogue aloud; students then repeated the lines and underlined vocabulary they found difficult. Post-activity: Students answered comprehension questions provided in the textbook.	60 minutes

Closing		Teacher reviews material then Students summarize the lesson and Teacher closes the class.	10 minutes
----------------	--	---	------------

Table 3.6 Meeting 3: Daily Life Reflection

Stages	Experimental Class	Control Class	Time
Opening	After the customary prayer and attendance check, the teacher introduces the topic of daily activities to set the context for the listening task.	The same routine of prayer and attendance checking is followed, after which the teacher presents the topic through textbook-based explanation.	10 minutes
Main activities	Pre-Listening Students are asked about their daily routines while the teacher introduces relevant vocabulary, including <i>habit</i>, <i>daily life</i>, <i>improve</i>, and <i>organize</i>. While-Listening While watching a podcast video on daily life reflection, students write down key points they hear and replay the video to confirm specific details. Post-Listening	Pre-activity: Teacher explains expressions from textbook. While-activity: Students listen to the teacher read the dialogue aloud, then repeat it and mark unfamiliar words. Post-activity: Students respond to comprehension questions from the textbook.	60 minutes

	Students discuss the benefits of documenting daily activities and produce a concise written summary.		
Closing	The lesson concludes with a joint summary, teacher feedback, words of encouragement, and a closing prayer.	The teacher reviews the material first, followed by a student-led summary, before the class ends.	10 minutes

Table 3.7 Meeting 4: Self Study Routine

Stages	Experimental Class	Control Class	Time
Opening	The teacher greets the class, leads a short prayer, checks attendance, and introduces the topic by building a self-study routine.	A similar opening procedure is followed, with the topic instead delivered through explanations based on the textbook.	10 minutes
Main activities	Pre-Listening Students share their existing study habits	Pre-activity: Teacher explains expressions from textbook.	60 minutes

	while the teacher introduces vocabulary related to self-study routines. While-Listening Students watch a podcast video explaining how to build a self-study routine, identify the main ideas and steps involved, and rewatch the video to verify details. Post-listening Students outline simple steps for creating their own study routine and present their ideas orally in front of the class.	While-activity: The teacher reads the dialogue aloud; students repeat it and underline words they are unfamiliar with. Post-activity: Students complete comprehension questions provided in the textbook.	
Closing	The teacher provides overall feedback on students' performance, expresses appreciation for their participation, and closes the session with a prayer.	The teacher reviews the material, summarize the lesson, and the class is then dismissed.	10 minutes

c. Post-test

Following the intervention, a post-test was given to assess students' listening achievement after using the English Podcast (LEP). The post-treatment test was administered to the experimental group and the control group using the same test format and level of difficulty as the pre-treatment test to ensure consistency and fairness. The purpose of the post-

test was to identify changes in students' listening performance after receiving podcast-based listening instruction and to determine the effectiveness of English Podcast (LEP) compared to conventional listening teaching methods.

2. Listening test

Listening tests will be used as the primary data collection technique in this study to measure students' listening comprehension before and after the implementation of the Learning English Podcast (LEP). The tests will be administered twice: as a pre-test before the intervention and as a post-test after the intervention. The pre-test will be conducted to determine the students' initial level of listening comprehension, while the post-test will be administered after the intervention to measure changes in their listening comprehension. The scores obtained from both tests will then be analyzed to determine the effect of the Learning English Podcast (LEP) on students' listening comprehension. The listening tests will be administered to both the experimental and control classes using the same testing procedures to ensure consistency and comparability of the data.

The same test items will be used in both the pre-test and post-test to maintain consistency in measuring

students' listening comprehension. The use of identical test items has also been applied in previous listening research. For example, Merzifonluoglu, (2023) used 20 identical multiple-choice questions in the pre-test and post-test and explained that the use of the same questions allowed a direct comparison of students' progress before and after the treatment. Therefore, the same test items will be used in the present study to maintain consistent measurement of students' listening comprehension before and after the implementation of Learning English Podcast (LEP).

3. Documentation

According to Sugiyono (2013), documentation is a data collection technique that uses records of past events in the form of written documents, photographs, or other relevant materials. In this study, documentation was used to collect supporting data related to the implementation of the research, including students' pre-test and post-test scores, lesson plans, attendance lists, and photographs of classroom activities during the treatment. These documents were used to provide evidence of the research implementation and to complement the data obtained from the listening tests.

E. Research Instrument

The primary utilized in this research was a listening comprehension test in multiple choice form, develop by the researcher. Its purpose was to asses students listening comprehension both before and after the LEP intervention was implemented. In developing the instrument, the researcher referred to Brown's (2004) principles of language assessment as general guidelines for constructing a listening comprehension test. Based on these principles and the objectives of the study, the researcher initially developed 45 multiple-choice items covering four listening comprehension indicators: listening for gist, listening for specific information, vocabulary recognition, and listening for inference.

Prior to its implementation in the actual study, the instrument was first tried out in order to assess the quality of the test items. The try-out was conducted with seventh-grade students at SMPN 15 Bengkulu Tengah because the research site, SMPN 11 Bengkulu Tengah, only had two seventh-grade classes, both of which were used as the experimental and control groups. Therefore, conducting the try-out at the same school was not feasible. SMPN 15 Bengkulu Tengah was selected because it has students at the same grade level and with comparable characteristics to those of the research participants.

The try-out data were analyzed through item validity, reliability, difficulty index, and discrimination index analyses to determine the quality of the instrument. Based on the results of these analyses, 30 items met the required criteria and were selected as the final instrument, while the remaining items were excluded. These 30 valid items were subsequently used as the pre-test and post-test for both the experimental and control classes. The blueprint of the final listening comprehension test is presented in Table 3.8

Table 3.8 Listening Test Blueprint

No	Indicator	Item Number	Total
1	Listening for Gist	1, 2, 3, 4, 5, 6, 7, 8	8
2	Identifying specific information	9, 10, 11, 12, 13, 14, 15, 16	8
3	Vocabulary Recognition	17,18,19,20,21,22,23	7
4	Making inference	24, 25, 26, 27, 28, 29, 30	7
TOTAL			30

As shown in Table 3.8, the final listening comprehension test consists of 30 multiple-choice items distributed across four listening comprehension indicators. Each indicator is proportionally represented to ensure comprehensive measurement of students' listening

comprehension. The test items were developed based on the Learning English Podcast (LEP) materials and were adapted to the English proficiency level of seventh-grade students by referring to the CEFR listening descriptors (Council of Europe, 2001). Each item was presented in the form of a short conversation or monologue, followed by four answer options (A, B, C, and D). Students selected the most appropriate answer after listening to the audio recording. Each correct answer was awarded one point, while each incorrect answer received zero points. The students' final scores were calculated using the following formula:

$$S = \frac{R}{N} \times 100$$

Where:

S = Students' score

R = Number of correct answers

N = Total number of test items

F. Instrument Testing

1. Validity of the test

The validity of test pertains to how accurately a research instrument is able to measure to the construct it is designed to asses. In this study, the validity of the listening test items was examined using the Pearson Product Moment correlation. Each test item was correlated with the total score to determine its validity.

According to Sugiyono (2013), an instrument is considered valid if the correlation coefficient (r-count) of each item is higher than the (r-table) value at a 5% significance level. This validity analysis was conducted to ensure that the listening test accurately measured students' listening comprehension.

No	Pearson Coefficient (r)	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

Table 3.9. Test Instrument Validity Criteria

The validity tes in this study, the researchers used 45 multiple-choice items developed on the basis of four indicators of listening skills. A pilot study was conducted at SMPN 15 Bengkulu Tengah in 2026; this school was selected as the pilot site for the instrument due to its academic standards and its location in Bengkulu Tengah, which is similar to the target location of the research. This pilot study was carried out to assess the quality of each test item before it was used in the

main study. The results of the data processing from the instrument validity test are presented as follow:

Table 3.10 Instrument Validity

Variable	Item	r table	r count	Result
SOAL	s1	0,443	0,547	VALID
SOAL	s2	0,443	0,633	VALID
SOAL	s3	0,443	0,431	INVALID
SOAL	s4	0,443	0,229	INVALID
SOAL	s5	0,443	0,603	VALID
SOAL	s6	0,443	0,593	VALID
SOAL	s7	0,443	-0,052	INVALID
SOAL	s8	0,443	0,477	VALID
SOAL	s9	0,443	0,431	INVALID
SOAL	s10	0,443	0,552	VALID
SOAL	s11	0,443	0,616	VALID
SOAL	s12	0,443	0,459	VALID
SOAL	s13	0,443	0,527	VALID
SOAL	s14	0,443	0,619	VALID
SOAL	s15	0,443	0,270	INVALID
SOAL	s16	0,443	0,522	VALID
SOAL	s17	0,443	0,356	INVALID
SOAL	s18	0,443	0,498	VALID
SOAL	s19	0,443	0,626	VALID
SOAL	s20	0,443	0,341	INVALID
SOAL	s21	0,443	0,616	VALID
SOAL	s22	0,443	0,504	VALID
SOAL	s23	0,443	0,458	VALID
SOAL	s24	0,443	0,642	VALID
SOAL	s25	0,443	0,193	INVALID
SOAL	s26	0,443	0,452	VALID
SOAL	s27	0,443	0,437	INVALID
SOAL	s28	0,443	0,642	VALID
SOAL	s29	0,443	0,547	VALID
SOAL	s30	0,443	0,642	VALID
SOAL	s31	0,443	0,310	INVALID
SOAL	s32	0,443	0,478	VALID
SOAL	s33	0,443	0,437	INVALID
SOAL	s34	0,443	0,084	INVALID
SOAL	s35	0,443	0,452	VALID
SOAL	s36	0,443	0,578	VALID
SOAL	s37	0,443	0,315	INVALID

SOAL	s38	0,443	0,462	VALID
SOAL	s39	0,443	0,564	VALID
SOAL	s40	0,443	0,328	INVALID
SOAL	s41	0,443	0,472	VALID
SOAL	s42	0,443	0,633	VALID
SOAL	s43	0,443	0,401	INVALID
SOAL	s44	0,443	0,452	VALID
SOAL	s45	0,443	0,485	VALID

The validity results shown in table 3.10 above, each item was analyzed by comparing the r-count value with the r-table value (0.443). The results showed that of the 45 test items, some were considered valid, whilst others were considered invalid. Valid items are those with an r-count value higher than the r-table value ($r\text{-count} > 0.443$), whilst items with an r-count value lower than the r-table value are considered invalid. The analysis revealed that 30 items were deemed valid and 15 items were deemed invalid. Consequently, only the valid items were included in the final research instrument.

2. Reliability of the test

Test reliability pertains to how consistently an instrument measures students' performance over repeated use. In this study, the reliability test was examined by means of the Cronbach's Alpha coefficient. Based on Sugiyono (2013), an instrument is considered reliable if the alpha coefficient is equal to or greater than 0.70. The reliability analysis was conducted to ensure that the listening test

produced stable and consistent results and was therefore appropriate for use in this research.

Table 3.11 Test Instrument Reliability Criteria

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

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

Table 3.12 Instrument Reliability

Reliability Statistic	
Cronbach's Alpa	N of items
0.918	45
Reliable	

An analysis of the reliability test through spss version 27 revealed a cronbach's coefficient of 0.918, based on 45 items included in the test. This value is used

to measure the internal consistency of the instrument, which indicates how closely related the items are in measuring the same construct.

Referring to the reliability criteria, a Cronbach's Alpha value above 0.70 indicates that the instrument has a good level of reliability. Since the obtained value is 0.918, which is higher than 0.70, it can be concluded that the instrument has a very high level of reliability. This confirms that the test items are consistent in measuring students' listening ability and are suitable for use in the main study.

3. The level difficulty

The level of difficulty was used to determine whether the test items were easy, medium, or difficult. The item difficulty test in this study was conducted to determine the level of difficulty of each test item. The analysis was carried out using Microsoft Excel by calculating the proportion of students who answered each item correctly. Each correct answer was scored as "1" and each incorrect answer was scored as "0". The total number of correct answers for each item was then divided by the total number of students to obtain the difficulty index. The difficulty index was calculated using the following formula:

$$P = \frac{B}{JS}$$

P = Difficulty Index

B = Number of students who answered correctly

JS = Total number of students

The interpretation of the difficulty index was based on Arikunto (2018), as presented in the following table:

Table 3.13 Interpretation of Difficulty Level

Categories

Question Category	Interpretation Of Result
Easy	0,71% -100%
Medium	0,31% - 0,70%
Difficult	0% -0,30%

The results of the difficulty index analysis are presented in the following table:

Table 3.14 Difficulty Index of Test Items

c	Difficult index	Category
1	0.70	Medium
2	0.65	Medium
3	0.40	Medium
4	0.75	Easy
5	0.70	Medium
6	0.65	Medium
7	0.60	Medium
8	0.60	Medium
9	0.50	Medium
10	0.65	Medium
11	0.85	Easy
12	0.60	Medium
13	0.80	Easy
14	0.60	Medium
15	0.55	Medium
16	0.65	Medium

17	0.60	Medium
18	0.55	Medium
19	0.60	Medium
20	0.70	Medium
21	0.60	Medium
22	0.55	Medium
23	0.55	Medium
24	0.50	Medium
25	0.50	Medium
26	0.60	Medium
27	0.60	Medium
28	0.65	Medium
29	0.55	Medium
30	0.60	Medium

Item difficulty was found that out of 30 test items, many of the items were categorized as medium difficulty. Specifically, 27 items (No. 1, 2, 3, 5, 6, 7, 8, 9, 10, 12, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, and 30) were classified as medium with difficulty index values ranging from 0.40 to 0.70. Meanwhile, 3 items (No. 4, 11, and 13) were categorized as easy because they had higher difficulty index values above 0.70, indicating that most students were able to answer these items correctly.

Furthermore, no items were categorized as difficult (0.00–0.30), which shows that none of the test items were too difficult for the students. Therefore, it can be concluded that the test items used in this study are appropriate for measuring students' listening ability

because they are balanced and suitable for classroom assessment.

4. Item discrimination test

The item discrimination test was conducted to determine how well each test item could differentiate between high achieving and low achieving students. A good test item should be able to distinguish students who have mastered the material from those who have not. Therefore, the discrimination index was calculated to evaluate the effectiveness of each test item. The discrimination index was calculated using the following formula:

$$DP = \frac{BA}{JA} - \frac{BB}{JB}$$

Where:

DP = Discrimination Index

PA = Proportion of correct answers in the upper group

PB = Proportion of correct answers in the lower group

In this study, the calculation was carried out using Microsoft Excel. Students' answers were coded as "1" for correct answers and "0" for incorrect answers. The students were then divided into upper and lower groups based on their total test scores. The average score of each group was calculated, and the discrimination index for each item was obtained by subtracting the lower group mean from the upper group mean. The

interpretation of the discrimination index was based on Arikunto (2018), as presented in Table:

Table 3.15 Interpretation of the discrimination index

T-test	Interpretation of Result
0,71 – 1,00	Very good
0,41 - 0,70	Good
0,21 - 0,40	Sufficiently Qualified
0,00 - 0,20	Weak

Table 3.16 Item Discrimination Index of Test Items

No	Upper Mean	Lower Mean	Discrimination Index	Category
1	0.90	0.50	0.40	Sufficient
2	0.90	0.40	0.50	Good
3	0.80	0.00	0.80	Very good
4	0.90	0.60	0.30	Sufficient
5	0.80	0.60	0.20	Weak
6	0.80	0.50	0.30	Sufficient
7	0.80	0.40	0.40	Sufficient
8	0.50	0.70	0.20	Weak
9	0.90	0.10	0.80	Very good
10	0.90	0.40	0.50	Good
11	1.00	0.70	0.30	Sufficient
12	0.80	0.40	0.40	Sufficient
13	0.90	0.70	0.20	Weak
14	0.80	0.40	0.40	Sufficient
15	0.70	0.40	0.30	Sufficient
16	0.80	0.50	0.30	Sufficient
17	1.00	0.20	0.80	Very good
18	0.80	0.30	0.50	Good
19	1.00	0.20	0.80	Very good
20	0.90	0.50	0.40	Sufficient
21	1.00	0.20	0.80	Very good
22	0.70	0.40	0.30	Sufficient
23	0.80	0.30	0.50	Good

24	0.70	0.30	0.40	Sufficient
25	0.80	0.20	0.60	Good
26	0.80	0.40	0.40	Sufficient
27	0.80	0.40	0.40	Sufficient
28	1.00	0.30	0.70	Good
29	0.80	0.30	0.50	Good
30	0.70	0.50	0.20	Weak

Based on the results of the analysis, the discrimination index of the test items varied from weak to very good categories. The classification of the items is presented as follows:

- Very Good (0.71 – 1.00): 6 items
- Good (0.41 – 0.70): 8 items
- Sufficient (0.21 – 0.40): 13 items
- Weak (0.00 – 0.20): 3 items

The results show that most of the test items fall into the sufficient and good categories, indicating that many of the items are able to differentiate students' abilities at an acceptable level. Several items are categorized as very good, meaning they are highly effective in distinguishing high- and low-performing students. However, there are also a small number of weak items which indicate that those items are less effective and may need revision or replacement.

G. Data analysis Technique

Normality tests, homogeneity tests, T-tests, F-tests, and hypothesis testing were among the essential data analysis methods employed in this investigation.

1. Normality test

The normality test serves to check whether students' listening scores follow a normal distribution, with a significance value that was greater than 0.05. SPSS is used by researchers to assess and determine normalcy.

The testing criteria, which are as follows:

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

2. Homogeneity test

This test was used to determine whether the variances of the experimental and control groups were equal. This test is a prerequisite for conducting parametric statistical analysis. The data were considered homogeneous if the significance value was greater than 0.05.

The homogeneity test establishes the homogeneity of the population data. To ascertain homogeneity, apply the following guidelines:

- a. Significance test (α) = 0.05
- b. If $s_1^2 = s_2^2$, then the variance of each sample is the same (homogeneity).
- c. If $s_1^2 \neq s_2^2$, then the variance of each sample is not the same (not homogeneity) to determine if the researcher uses SPSS.

3. T-test

A T-test will be conducted to examine whether a significant difference exists in listening achievement between the experiment and control groups. A t-test will be conducted to examine whether a significant difference exists in listening achievement between the experimental and control groups. Analysis will be performed using SPSS. In general, this test aims to determine the extent to which the partial effect of an independent variable influences the dependent variables remain constant. The formula applied in this analysis is presented below.

- 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 α (0.05).

4. F-test

An F-test refers to a statistical technique employed to determine the homogeneity of variance among two or more groups. In quantitative research, especially in quasi-experimental design, the F-test is applied as a prerequisite test before conducting parametric analysis such as the independent sample t-test. Variance homogeneity is important because parametric tests assume that the populations from which the samples are drawn have equal variances.

According to Sugiyono (2013), the F-test is used to examine the similarity of variances between two groups. If the variances are homogeneous, further parametric statistical analysis can be conducted. In this study, the F-test was used to determine whether the variance of the listening test scores from the experimental group and the control group was homogeneous.

H. Statistical Hypothesis

Hypothesis testing in this study served to determine whether H_0 would be accepted or rejected, and correspondingly, whether H_a would be accepted or rejected. This process was carried out using SPSS version 25, guided by the following decision criteria:

- a. If the t-value exceeds the t-table value, or if the significance (2-tailed) is below 0.05, the hypothesis is accepted.
- b. If the t-value falls below the t-table value, or if the significance (2-tailed) exceeds 0.05, the hypothesis is rejected.

