

CHAPTER 3

METHODOLOGY

This part of the paper explains the methodology utilized to conduct the research, which is as follows: research design, sites and participants, data collection method, and data analysis.

3.1. Research Design

To evaluate students' enhancement in performing speaking, this study employed a quantitative approach and employed a quasi-experimental design. This design, ideal for examining the effectiveness of one variable over another, allows manipulation of variables to establish cause-and-effect (Kothari, 2004). Creswell and Creswell (2018) also emphasize its use for assessing attitudes before and after treatment. Hatch and Faraday (1982) describe it as a realistic balance between human linguistic nature and experimental rigor. Porte (2012) supports its use in educational research, where established courses are often involved.

This design was selected for a number of important reasons, as follows: First, the population and participants are selected randomly in suit of authentic experimental design (Cresswell, 1994); second, in well-designed experimental characteristics, the quasi-experimental design must consist of both the control and experimental groups to allow comparison between the two (Emilia, 2000); and the third, pretest part of the quasi-experimental design attempts to compare the features between the control and experimental groups before any treatment is administered. The posttest results were investigated to determine the influence of the treatments on the participants.

In addition, there are five particular key criteria that characterize quasi-experimental research, namely: 1) the data collection consists of two groups, the experimental group and the control group; 2) both groups are measured and compared in terms of the dependent variable; 3) both groups are measured two times, in a pretest and a posttest; 4) the dependent variable is measured by the same test and at the same time; and 5) a particular treatment is employed in the experimental group (Creswell & Miller, 2000).

Ryan Adriansyah Ramdani, 2024

UTILIZING VIDEO-BASED DISCUSSION PLATFORM IN ASSISTING THE IMPROVEMENT OF EFL HIGH SCHOOL STUDENTS' SPEAKING PERFORMANCE

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

The study utilized the One Group Pretest Posttest Design as a pre-experimental group, where the dependent variable (pretest) was measured, followed by the application of a stimulus in the classroom (treatment), and a subsequent measurement of the dependent variable (posttest) without any comparison groups. This design involved two tests: a pretest (O1) conducted before the treatment and a post-test (O2) conducted after the experiment. This methodology, as per Shadish et al. (2002), entails a single pretest observation on a group of respondents (O1), the implementation of treatment (X), and a single posttest observation on the same measure (O2), illustrated as in Table 3.1.

Table 3.1 Pretest and Posttest Group Scheme

Experiment Design Used in This Study			
Experimental group	Pretest (O1)	X (treatment with Flipgrid)	Posttest (O2)
Control group	Pretest (O1)	Treatment without Flipgrid	Posttest (O2)

During the study, the researcher provided teaching material, assessed students' speaking progress, and delivered instruction in two groups of classes. The treatment was administered during the teaching and learning process, preceded by a pretest assessing students' speaking performance. Flipgrid was incorporated into the teaching method. Following the treatment, a posttest was conducted to measure the impact, allowing the observation of pretest and posttest results.

3.2. Hypothesis

To examine the research question, quantitative studies necessitate the utilization of quantitative hypotheses. As per Creswell and Creswell (2018), these hypotheses constitute the researcher's forecasts concerning the anticipated outcomes regarding relationships among variables. This concept aligns with the viewpoint of Fraenkel et al. (2012), which defines a hypothesis as a projection of potential study outcomes. In quantitative research, the researcher establishes both a research hypothesis and a null hypothesis. To evaluate the research hypothesis,

Ryan Adriansyah Ramdani, 2024

UTILIZING VIDEO-BASED DISCUSSION PLATFORM IN ASSISTING THE IMPROVEMENT OF EFL HIGH SCHOOL STUDENTS' SPEAKING PERFORMANCE

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

it is imperative to articulate a null hypothesis (Fraenkel et al., 2012). Consequently, in order to refute the null hypothesis (H₀), an alternative hypothesis (H_a) is also formulated, representing the hypothesis deemed acceptable for the research at hand. The hypotheses for this study are outlined as follows:

(H₀) = There is no significant difference in students' speaking performance before and after instruction employing Flipgrid.

(H_a) = There is a significant difference in students' speaking performance before and after instruction employing Flipgrid.

The null hypothesis (H₀) posited in this investigation suggests that there is no discernible variance in students' speaking performance pre- and post-usage of Flipgrid for instructional purposes. The study anticipates rejecting the null hypothesis (H₀) and confirming the alternative hypothesis (H_a), thus underscoring the effectiveness of employing Flipgrid in instructional activities aimed at enhancing students' speaking performance.

3.3. Data Collection

3.3.1. Population and Sample

Research participants were selected from the 10th grade students of one of the public schools located in West Java. This particular grade level was chosen because speaking skills have yet to be taught to them; thus, their speaking performance has never been measured in this degree as well. Additionally, since the students were approximately new to the high school phase, it was considered an excellent start for them to get some exposure to the current technology to support their language learning. The experimental class selected for this study is composed of students from the 10th grade. This group comprised 144 students, with an equal distribution of 72 students in the experimental group and 72 others in the control group. The school designated this particular class due to its perceived cooperative nature and anticipated willingness to actively participate in the research procedures.

3.3.2. Research Instruments

An instrument, in the context of this study, refers to a tool utilized for data collection purposes. Instrument refers to a device or tool used to collect data. In educational research, particularly quantitative approaches, instruments are often tests, questionnaires, or rating scales. Specifically, the instruments employed in this research were speaking tests and questionnaires.

3.3.2.1. Pretest and Posttest

Initially, a pretest was conducted to assess the students' speaking ability before the intervention commenced. Subsequently, a posttest was administered to determine whether there was an enhancement in the students' speaking ability following the utilization of Flipgrid. Both assessments required students to perform orally, tailored to their grade level. The pretest prompt was centered around "My First Crush," while the posttest prompt focused on "The Best Day in My Life." Introducing distinct topics for each test aims to prevent bias and ensure consistency in student responses across assessments. Evaluation criteria encompassed various aspects, including fluency, cohesion, and content, as outlined in the scoring rubric.

3.3.2.2. Questionnaire

To obtain comprehensive research data, questionnaires were administered after the post-test to investigate how Flipgrid's features help high school students' speaking performance. Questionnaires efficiently gather information from large populations (Ponto, 2015), making them ideal for studies with numerous participants (Bhattacharjee, 2012; Munasiroh, 2023). They effectively capture data on attitudes and perspectives that can't be directly observed (Bhattacharjee, 2012). Only the experimental group received the questionnaire, which was translated into Bahasa Indonesia to avoid misunderstandings. It featured 20 close-ended, Likert-scale questions covering students' experience with Flipgrid, its utility, features, and impact on speaking performance (Fraenkel, Wallen, & Hyun, 2012).

Ryan Adriansyah Ramdani, 2024

UTILIZING VIDEO-BASED DISCUSSION PLATFORM IN ASSISTING THE IMPROVEMENT OF EFL HIGH SCHOOL STUDENTS' SPEAKING PERFORMANCE

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

Table 3.2 Questionnaires on the Experience of Using Flipgrid

No.	Items	Strongly Disagree	Disagree	Agree	Strongly Agree
1.	I like learning using Flipgrid.				
2.	Learning using Flipgrid increases my enthusiasm for learning English in the classroom.				
3.	Learning using Flipgrid creates a more enjoyable atmosphere.				
4.	Learning using Flipgrid increases my engagement in the learning process.				
5.	Learning using Flipgrid makes me more confident to speak English.				
6.	Flipgrid is easy to use.				
7.	Flipgrid is flexible since I can use it anywhere.				
8.	Flipgrid's features are interesting and engaging.				
9.	Flipgrid is device- and internet-friendly.				
10.	Learning using Flipgrid helps me to prepare more before performing speaking.				
11.	"Grids" feature helped me to manage the online classroom better.				
12.	"Topics" feature allowed me to understand the instructions better.				
13.	"Video Responses" feature allowed me to prepare until I perform my best in speaking.				
14.	"Comments" feature in Flipgrid engaged me in the interactive online classroom.				
15.	Enhanced features in Flipgrid, such as "Whiteboard, Music, Video Editing", helped me to develop my creativity in				

Ryan Adriansyah Ramdani, 2024

UTILIZING VIDEO-BASED DISCUSSION PLATFORM IN ASSISTING THE IMPROVEMENT OF EFL HIGH SCHOOL STUDENTS' SPEAKING PERFORMANCE

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

No.	Items	Strongly Disagree	Disagree	Agree	Strongly Agree
	speaking activities.				
16.	“Mic Only” feature helped me to speak more fluently since I don’t have to present myself in real time.				
17.	“Lecturer’s Comment” feature allowed me to review the linguistic mistakes in my speech.				
18.	“Record/Pause” feature assisted me to fix my accuracy and intonation.				
19.	“Board” feature helped me contextually explain better to my target audience.				
20.	“Sticky Note” feature assisted me maintain the cohesion and coherence in my speaking.				

3.4. Research Procedure

It is essential for research to follow certain procedures in order to gain the required data. The procedures conducted in this study refer to the steps that have employed by Sugiyono (2017) and Creswell (2012), which include:

- 1) Establishing the research questions and hypothesis;
- 2) Selection of the participants;
- 3) Grouping to control and experimental group;
- 4) Establishing the research instruments;
- 5) Conduction of pretest;
- 6) Treatment;
- 7) Conduction of posttest;
- 8) Spread of the questionnaires;
- 9) Data Analysis;

The procedure is adapted and modified with an addition of the spread of questionnaires before the data analysis. This research was conducted before the end of service of Flipgrid as an independent platform and its integration to Microsoft (October 2024). To be specific, the idea of this study was already

Ryan Adriansyah Ramdani, 2024

UTILIZING VIDEO-BASED DISCUSSION PLATFORM IN ASSISTING THE IMPROVEMENT OF EFL HIGH SCHOOL STUDENTS’ SPEAKING PERFORMANCE

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

established in 2023 and thus the research was conducted in April 2024. Particularly in this study, to measure the impact of using Flipgrid in the classroom, the procedures are explained as follows:

3.4.1. Administering the Pilot Test

The pilot test was conducted in order to measure both the validity and reliability of the research instrument. Its objective was to confirm the appropriateness of the instruments to be used for the experimental group. In this particular study, the pilot test was applied to a group consisting of 20 students. This group was established from 4 classrooms, which later will be classified as 2 control groups and 2 experimental groups. The form of the test was a recorded speaking test with the topic of "Interesting Day This Week." Whereas the scoring rubric includes four criteria, namely fluency, accuracy, coherence, and content.

3.4.2. Choosing participants

The study took place at a senior high school located in Bandung. The population consisted of 10th grade students attending junior/senior high school. For the purpose of this research, one class from the 10th grade was selected as the sample. The selection was made by the teacher responsible for the 10th grade students of one of the public schools located in West Java, who chose a class capable of actively participating in the research and willing to cooperate in following the prescribed treatment procedures.

3.4.3. Designing research instruments

Designing the research instruments represented a crucial stage, as it directly influences the research outcomes and addresses the research question. During this phase, the researcher developed speaking tests for both the pretest and posttest. These tests were intended to gauge any improvement in students' speaking scores before and after the treatment. The assessment of students' speaking was conducted using a rubric that encompassed key aspects of speaking: fluency, accuracy, coherence, and content.

3.4.4. Administering pretest

In utilizing Flipgrid into students' speaking activities, a pretest was administered to assess their speaking performance's proficiency. This activity took place during the initial meeting with the students. The pretest took the form of speaking activity of one person. The theme is "My First Crush," and they are asked to record their own experience in an oral manner with no certain requirement. They were allotted 30 minutes to have a training section before performing the speaking, with a duration of 3-5 minutes. To facilitate the smooth conduct of the pretest, students were supervised to ensure they proceeded in an organized manner.

3.4.5. Conducting treatments

Following the administration of the pretest, treatments were administered to the students. These treatments involved the implementation of teacher feedback to aid students in performing the speaking. Each instructional session lasted for one hour, with an hour comprising forty minutes. The breakdown of the meetings is detailed below:

Table 3.3 Research and Treatment Timeline

No.	Date	Material/Classroom Activity		Time Allocation (Minutes)
		Experimental Group	Control Group	
1.	29-30 April 2024	Introduction	Introduction	2x45
2.	6-7 May 2024	Pretest	Pretest	4x45
3.	13-14 May 2024	Discussion: Content and features of speaking	Discussion: Content and features of speaking	2x45
4.	20-21 May 2024	Introduction to Flipgrid and its features, practice session using Flipgrid	Conventional speaking practice	2x45
5.	27-28 May 2024	Speaking practice using Flipgrid, commenting on a friend's video	Conventional speaking practice	2x45

Ryan Adriansyah Ramdani, 2024

UTILIZING VIDEO-BASED DISCUSSION PLATFORM IN ASSISTING THE IMPROVEMENT OF EFL HIGH SCHOOL STUDENTS' SPEAKING PERFORMANCE

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

6.	3-4 June 2024	Speaking practice using Flipgrid	Conventional speaking practice	2x45
6.	10-11 June 2024	Posttest	Posttest	4x45
7.	17-18 June 2024	Spread of Questionnaire and Closing	Closing	-

Most of the procedures were conducted in equal and similar terms except for the treatment. The speaking performance assessed was a monologue and a planned talk.

Table 3.4 Research and Treatment Description

Timeline	Description	
	Experimental Group	Control Group
Pretest	Students were prompted to make a monologue narration with a planned talk setting with the topic “My First Crush”. The duration of the video was 3 minutes maximum. The type of the text employed was Recount Text since it was based on personal experience. Students were given 30 minutes to practice before performing and recorded the video before upload it to Google Drive.	Students were prompted to make a monologue narration with a planned talk setting with the topic “My First Crush”. The duration of the video was 3 minutes maximum. The type of the text employed was Recount Text since it was based on personal experience. Students were given 30 minutes to practice before performing and recorded the video before upload it to Google Drive.
Treatment	For one week, the students were introduced to Flipgrid and its features. Then, the students were treated using Flipgrid to practice their speaking in similar manner and setting. The topic was “Interesting Thing This Week”. The form of the speaking was monologue and planned talk, with maximum 3 minutes. In two weeks,	For one week, the students were introduced to traditional speaking and its components. Then, the students were treated using the conventional method to practice their speaking in similar manner and setting. The topic was “Interesting Thing This Week”. The form of the speaking was monologue and planned talk, with

Ryan Adriansyah Ramdani, 2024

UTILIZING VIDEO-BASED DISCUSSION PLATFORM IN ASSISTING THE IMPROVEMENT OF EFL HIGH SCHOOL STUDENTS' SPEAKING PERFORMANCE

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

	students were required to record two videos with identical topic and upload them to google drive.	maximum 3 minutes. In two weeks, students were required to record two videos with identical topic and upload them to google drive.
Posttest	Students were prompted to make a monologue narration with a planned talk setting with the topic “The Best Day of My Life”. The duration of the video was 3 minutes maximum. The type of the text employed was Recount Text since it was based on personal experience. Students were given 30 minutes to practice before performing and recorded the video before upload it to Google Drive.	Students were prompted to make a monologue narration with a planned talk setting with the topic “The Best Day of My Life”. The duration of the video was 3 minutes maximum. The type of the text employed was Recount Text since it was based on personal experience. Students were given 30 minutes to practice before performing and recorded the video before upload it to Google Drive.

Whereas for the data that would be picked as an example, it was not possible to transcribe every single one of the students’ speaking performance and present them in this report, we decided to pick 12 students in total as excerpt, with equally distributed dividing between the control and the experimental group. We picked 3 students from the control group and other 3 from the experimental group for the pretest session, and the similar selection was applied for the posttest session. The students were selected based on their respective speaking performance level.

3.4.6. Administering posttest

After the treatment, a posttest was given to compare students' speaking performance before and after the intervention. Like the pretest, students delivered a speech on a different topic, allowing a fair comparison of their progress.

3.5. Data Analysis

3.5.1. Pretest and Posttest

3.5.1.1. Scoring Technique

A scoring rubric was used to assess students' speaking proficiency in both the pretest and posttest phases. Each criterion was assigned specific scores to ensure accurate evaluation. The researcher analyzed the students' speaking from both phases using an adapted version of the American Council on the Teaching of Foreign Languages (2012) rubric. Table 3.4 shows the simplified version.

Table 3.5 Rubric of Speaking

Aspects	Criteria	Scores	Weight
Fluency (F) 25%	- Limited, hesitant speech. Frequent pauses. Little to no spontaneity.	1	2.5X
	- Somewhat hesitant, with noticeable pauses. Emergent spontaneity.	2	
	- Averagely fluent, with occasional hesitations. Developing spontaneity.	3	
	- Mostly fluent with minimum pauses.	4	
	- Fluent and spontaneous. Near-native or native-like flow of speech.	5	
Accuracy (A) 25%	- Limited range of vocabulary and basic grammatical structures. Frequent errors.	1	2.5X
	- Developing range of vocabulary and grammatical structures. Some errors.	2	
	- Decent numbers of vocabulary and grammatical structures. Occasional errors.	3	
	- Varied vocabulary and grammatical structures. Few errors.	4	
	- Rich and varied vocabulary. High level of grammatical accuracy. There are very few or even no errors.	5	
Cohesion (Ch) 25%	- Limited ability to link ideas. Speech may lack coherence.	1	2.5X
	- Developing the ability to link ideas. Speech is somewhat cohesive.	2	
	- Average cohesiveness, with some	3	

Ryan Adriansyah Ramdani, 2024

UTILIZING VIDEO-BASED DISCUSSION PLATFORM IN ASSISTING THE IMPROVEMENT OF EFL HIGH SCHOOL STUDENTS' SPEAKING PERFORMANCE

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

Aspects	Criteria	Scores	Weight
	difficulty in maintaining coherence.	4	
	- Decently cohesive, with a few difficulties in cohesing ideas.	5	
	- Cohesive, with smooth transitions between ideas. Demonstrates excellent coherence.		
Content (Cn) 25%	- Limited ability to convey information or express ideas. May rely heavily on memorized phrases.	1	2.5X
	- Can convey basic information and express simple ideas.	2	
	- Generally can express some complex ideas and some detailed information.	3	
	- Can express complex ideas and convey detailed information.	4	
	- Can handle abstract topics and convey nuanced meaning.	5	

Adapted from TOEFL iBT® Tests. (n.d.). Speaking rubric. In TOEFL iBT® Tests. <https://www.ets.org/toefl.html> with some adjustments and simplifications to make the rubric more comprehensible both for the participants and the reviewer.

$$\text{Score} = 4 \times (\text{F} + \text{A} + \text{Ch} + \text{Co})$$

Table 3.5. demonstrates the computation of the students' speaking score.

Table 3.6 Computation of the Rubric Score

FINAL SCORE		
Criteria	Raw Score	Converted Score (x4)
Fluency	4	20
Accuracy	4	20
Cohesion	5	25
Content	5	25
Total	18	90

After measuring the student's performance, the average scoring was classified as in Table 3.6.

Table 3.7 Level Categorization of the Speaking Score

Score	Category
81-100	Excellent
61-80	Average to Good
41-60	Poor to Average
0-40	Poor

Adapted from Brown (2019), *Language Assessment: Principles and Classroom Practices*, and Council of Europe (2001).

3.5.1.2. Data Analysis on the Pilot Test

A pilot test was conducted to assess the validity and reliability of the study instruments. Validity, as defined by Riyanto and Hatmawan (2020), refers to how well a research instrument performs its intended function. Twenty students from a separate class participated in the pilot test. Their speaking performances were analyzed using the Pearson Product Moment method in SPSS.

$$r = \frac{N \cdot \Sigma xy - (\Sigma x) (\Sigma y)}{\sqrt{\{N \cdot \Sigma x^2 - (\Sigma x)^2\} \{N \cdot \Sigma y^2 - (\Sigma y)^2\}}}$$

Figure 3.1 Pearson Product Moment Validity Formula

(Riyanto & Hatmawan, 2020, p. 63)

Furthermore, to assess the degree of validity of the instruments, the final results utilized a scale of coefficient correlation criteria for comparison. The table outlining these criteria is as follows:

Table 3.8 The Criteria of Coefficient Correlation

Coefficient Interval	Interpretation
0.00 – 0.199	Very Low
0.20 – 0.399	Low
0.40 – 0.599	Fair
0.60 – 0.799	High
0.80 – 1.000	Very High

(Sugiyono, 2011)

Ryan Adriansyah Ramdani, 2024

UTILIZING VIDEO-BASED DISCUSSION PLATFORM IN ASSISTING THE IMPROVEMENT OF EFL HIGH SCHOOL STUDENTS' SPEAKING PERFORMANCE

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

The validity test was done for all of the scoring criteria, which are fluency (F), accuracy (A), coherence (Ch), and content (Co). The results of a validity test conducted on Try-Out Class show that the instrument is valid. It can be seen in Table 3.8 below:

Table 3.9 The Results of Pretest-Posttest Validity Test

Item	Score	Interpretation
F	0.512	Fair
A	0.568	Fair
Ch	0.706	High
Co	0.665	High

According to Table 3.6, the result of the validity test of the scoring instrument can be concluded as valid because the scoring aspects were on the Fair-Very High level. The score of each scoring aspect on the validity result is as follows: (1) Fluency is 0.512, (2) Accuracy is 0.568, (3) Cohesion is 0.506, and (4) Content is 0.665. It can be concluded that the instrument used for this research is a valid research instrument.

After calculating the validity of the test, the reliability was also calculated to measure the reliability of the instrument. Charney (1984) stated that a reliable measurement is capable of replication under equivalent conditions; it also concluded that a reliable method of assessing speaking performance would yield a consistent judgment of a student's abilities being equal. The reliability of a measuring instrument is the consistency with which the instrument measures what it measures (Riyanto & Hatmawan, 2020).

Following the assessment of validity, reliability was also evaluated to gauge the consistency and replicability of the instrument. Charney (1984) proposed that a reliable measurement should be capable of replication under similar conditions, ensuring consistent judgments of a student's abilities. Reliability, as defined by Riyanto and Hatmawan (2020), refers to the consistency with which an instrument measures what it is intended to measure.

To measure the reliability of the test, the speaking performance of the tryout students was analyzed using the Cronbach's Alpha formula in SPSS. The formula for Cronbach's Alpha is depicted in Figure below.

$$r = \left(\frac{k}{(k-1)} \right) \left(1 - \frac{\sum \sigma b^2}{\sigma t^2} \right)$$

Figure 3.2 Reliability Test Formula

(Riyanto & Hatmawan, 2020, p. 75)

The criteria of coefficient correlation were used to determine the scale of reliability on the result; the criteria conducted on Table 3.9 below:

Table 3.10 The Criteria of Coefficient Correlation

Coefficient Interval	Interpretation
$r \leq 0.20$	Very Low
$0.20 < r \leq 0.40$	Low
$0.40 < r \leq 0.70$	Fair
$0.70 < r \leq 0.90$	High
$0.90 < r \leq 1.00$	Very High

(Sugiyono, 2011)

The reliability test of the Try-Out class was done by using SPSS; the results are described in Table 3.10 below:

Table 3.11 The Result of Pretest Posttest Reliability Test

Cronbach's Alpha	N of Items
0.730	4

Based on the reliability results presented in Table 3.10, it can be inferred that the instrument demonstrates reliability, with a Cronbach's Alpha value of 0.730. This indicates that the reliability test value falls within the range of $0.70 < r \leq 0.90$, categorizing it as a high level of reliability. Therefore, it can be concluded

that the instrument utilized in this research is reliable for measuring the intended variables.

3.5.1.3. Data Analysis on Pretest and Posttest

3.5.1.3.1. Normality Distribution Test

Normality testing was conducted to determine whether the data followed a normal distribution, ensuring appropriate statistical analysis (Riyanto & Hatmawan, 2020). The Shapiro-Wilk test in SPSS was employed for statistical calculations. The procedure for testing the normality of the data involved three steps:

1. Setting the level of significance (p) at 0.05 to establish the hypotheses:
 - H_0 : The test scores are normally distributed.
 - H_a : The test scores are not normally distributed.
2. Analyzing the normality distribution using the Shapiro-Wilk test on SPSS.
3. Interpreting the results of the test based on the comparison between the significance value (Sig.) and the level of significance (p). The interpretation criteria are as follows:
 - If the significance value (Sig.) $>$ level of significance (0.05), then the data are considered normally distributed.
 - If the significance value (Sig.) $<$ level of significance (0.05), then the data are not considered normally distributed.

3.5.1.3.2. Homogeneity Distribution Test

Homogeneity testing was conducted to assess whether the variance of the data was homogeneous. The Levene test in SPSS was utilized for statistical calculations. The procedure for testing the homogeneity of the data involved three steps:

1. Setting the level of significance (p) at 0.05 to establish the hypotheses:
 - H_0 : The variance of the test scores is homogeneous.
 - H_a : The variance of the test scores is not homogeneous.
2. Analyzing the homogeneity of variance using the Levene test on SPSS.

Ryan Adriansyah Ramdani, 2024

UTILIZING VIDEO-BASED DISCUSSION PLATFORM IN ASSISTING THE IMPROVEMENT OF EFL HIGH SCHOOL STUDENTS' SPEAKING PERFORMANCE

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

3. Interpreting the results of the test based on the comparison between the significance value (Sig.) and the level of significance (p). The interpretation criteria are as follows:
 - If the significance value (Sig.) $>$ level of significance (0.05), then the data are considered homogeneous.
 - If the significance value (Sig.) $<$ level of significance (0.05), then the data are not considered homogeneous.

3.5.1.3.3. Paired Wilcoxon Signed-Rank Test

Since the data is not normally distributed, the paired sample Wilcoxon Signed-Rank Test was employed to determine whether the mean scores of students' pretest and posttest had any fluctuation. The paired sample Wilcoxon Signed-Ranks was conducted using SPSS, following these steps:

1. Setting the level of significance for the asymp. Sig on 0.05. The rule is necessary to establish the following hypothesis:
 - H_0 : The mean scores of students' pretest and posttest are not different.
 - H_a : The mean scores of students' pretest and posttest are different.
2. Analyzing the paired sample Wilcoxon Signed-Rank test using SPSS.
3. Interpreting the results of the test based on table:
 - Negative ranks mean that the score of the posttest $<$ the pretest.
 - Positive ranks mean that the score of the posttest $>$ the pretest.
 - Ties means that the score of the posttest = the pretest.

Furthermore, the determination of the conclusion is decided by the following rules:

- If the significance value (2-tailed) is <0.05 , then there is a difference between the pretest and posttest scores. Thus, the alternative hypothesis (H_a) is accepted.
- If the significance value (2-tailed) is <0.05 , then there is a difference between the pretest and posttest score. Thus, the alternative hypothesis (H_0) is rejected.

3.5.2. Questionnaire

Three steps are taken to analyze the data from the questionnaire. First, the statements in the questionnaire were classified based on themes (Cresswell, 2008). Second, the data was measured with computations and percentages. Third, the results were tabulated. Then, it was analyzed and interpreted. The Likert scale was employed to measure the extent of scores on which a person agreed or disagreed. The score ranges from 1 to 5. This is in line with Creswell (2008), which illustrates the scale as follows: 1 as "strongly disagree," 2 as "disagree," 3 as "agree," and 4 as "strongly agree." The scoring results of the questions were analyzed by transforming them into percentages using the frequency base with the following formula:

$$\frac{\text{Number of students choosing certain option}}{\text{Total number of the students}} \times 100\%$$

3.5.2.1. Questionnaire Validity Test

Similar to the pretest and posttest instrument, a validity test is essential to determine the credibility of the instrument. The content was reviewed by an English lecturer with additional analysis of SPSS 27. The Pearson product moment correlation coefficient was deployed to assess the validity with the interpretation based on a significance level of 0.05. In specific, if the correlation coefficient exceeds 0.207, the item is deemed sufficient and valid. The validity test results for each item are presented in the table below.

Table 3.12 The Results of Questionnaire Validity Test

Items	Correlation value (r-count)	Description
Q1	0.589	Valid
Q2	0.457	Valid
Q3	0.612	Valid
Q4	0.662	Valid
Q5	0.592	Valid
Q6	0.527	Valid

Ryan Adriansyah Ramdani, 2024

UTILIZING VIDEO-BASED DISCUSSION PLATFORM IN ASSISTING THE IMPROVEMENT OF EFL HIGH SCHOOL STUDENTS' SPEAKING PERFORMANCE

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

Q7	0.684	Valid
Q8	0.482	Valid
Q9	0.524	Valid
Q10	0.512	Valid
Q11	0.532	Valid
Q12	0.452	Valid
Q13	0.483	Valid
Q14	0.524	Valid
Q15	0.562	Valid
Q16	0.625	Valid
Q17	0.582	Valid
Q18	0.612	Valid
Q19	0.558	Valid
Q20	0.642	Valid

From table 3.11, it can be indicated that all the questionnaire items have a bigger Correlation r-count Value than the value of r Table (0.207), which can be interpreted that the items on the questionnaire are valid.

3.5.2.2. Questionnaire Reliability Test

The reliability test is utilized to determine the reliability and consistency of the questionnaire if the similar measurement is carried out repeatedly on different sample or population. The Cronbach's Alpha was employed. The criteria for reliability test are if the alpha value is >0.60 , then the statement is reliable. If the alpha value is <0.60 , it means that the statement is unreliable. The reliability test results are demonstrated in detail in the table below.

Table 3.13 The Results of Questionnaire Reliability Test

Item	Cronbach's Alpha	Description
Q1	0.822	Reliable
Q2	0.813	Reliable
Q3	0.782	Reliable

Ryan Adriansyah Ramdani, 2024

UTILIZING VIDEO-BASED DISCUSSION PLATFORM IN ASSISTING THE IMPROVEMENT OF EFL HIGH SCHOOL STUDENTS' SPEAKING PERFORMANCE

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

Q4	0.769	Reliable
Q5	0.812	Reliable
Q6	0.852	Reliable
Q7	0.824	Reliable
Q8	0.833	Reliable
Q9	0.842	Reliable
Q10	0.797	Reliable
Q11	0.781	Reliable
Q12	0.813	Reliable
Q13	0.816	Reliable
Q14	0.828	Reliable
Q15	0.819	Reliable
Q16	0.836	Reliable
Q17	0.812	Reliable
Q18	0.814	Reliable
Q19	0.842	Reliable
Q20	0.796	Reliable

According to Table 3.11, since all of the Cronbach's Alpha value is higher than 0.60, it can be indicated that all the questionnaire items used are reliable.

3.6. Concluding Remarks

This chapter has presented the methodology utilized in this study in order to answer the research questions mentioned in chapter one. The methodology includes research design, sites and participants, data collection, as well as data analysis.