CHAPTER IV

RESEARCH FINDING AND DISCUSSION

This chapter explains the result of the research. It describes the research finding and the discussion of the finding. The research finding consists of the result of pre-test, the result of post-test, inter rater reliability and the computation ANCOVA by using SPSS.

A. Research Finding

This study was conducted to answer the research questions through data analysis obtained from the administration of pre-tests and post-tests to two different classes. The pre-test was carried out on 23th January 2025, while the post-test was conducted on 20th February 2025, for both the experimental group and the control group. These tests were administered to eleventh-grade students at MA Nurul Islam Lampung.

The researcher presented the analysis results of the pre-test and post-test conducted on both groups. The information reflects the achievement levels of students who learned through the Praktika.ai application compared to those who learned using picture media. The data were analyzed using the ANCOVA technique with the assistance of the SPSS program to identify differences in learning outcomes between the two groups.

1. The Result of Pre-test

Before giving the treatment, the researcher conducted a pre-treatment test on both the experimental and control groups to assess the students' speaking skills. The speaking scores of eleventh grade students at MA Nurul

Islam Lampung were used to generate this data. In the speaking assessment rubric, the researcher evaluated the students' speaking ability based on five aspects: pronunciation, grammar, vocabulary, fluency, and comprehensibility. The results of the pre-test from both the control and experimental groups are presented in Table 4.1.

Table 4.1 The Result of Pre-Test

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Pre-Test Experiment	36	52	76	61.50	5.438
Pre-Test Control	35	52	76	63.49	5.409
Valid N (listwise)	35				

Table 4.1, which presents the descriptive statistics, summarizes the pre-test results of both the experimental and control groups. The sample consisted of 36 students in the experimental group and 35 students in the control group. Both groups had the same score range, from 52 to 76. The average score of the experimental group was 61.50 with a standard deviation of 5.438, while the control group recorded a slightly higher average score of 63.49 with a nearly identical standard deviation of 5.409. These findings indicate that the students in the control group performed slightly better on the pre-test compared to those in the experimental group. The "Valid N (listwise)" value of 35 indicates that 35 participants had complete data across all variables used in the analysis. These statistics provide a basic understanding of the condition of both groups before the treatment, which is

essential for comparing the post-test results and evaluating the effect of the intervention.

2. The Result of Post test

The post-test was administered after the student's received treatment from the researcher. This test aimed to evaluate the students' speaking skills following the treatment, with the experimental group using the Praktika application and the control group using picture media. The researcher analyzed the post-test results using SPSS, and the findings are presented in Table 4.2.

Table 4.2 The Result of Post-Test

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Post-Test Experiment	36	66	88	76.72	6.368
Post-Test Control	35	58	89	71.63	6.929
Valid N (listwise)	35				

Table 4.2 shows that the experimental group (N = 36) and the control group (N = 35) demonstrated different results after the treatment. The average post-test score of the experimental group was 76.72, which was higher than the control group's average of 71.63. The experimental group's scores were also more consistent, with a standard deviation of 6.368 compared to 6.929 in the control group. This data indicates that the experimental group experienced greater improvement in speaking skills; however, inferential analysis is needed to determine whether the difference is statistically significant.

3. Inter-Raters' Reliability

Two raters in this study assessed the students' speaking skills by assigning scores. The researcher identified the Pearson Product Moment correlation between the pre-test and post-test scores of both the experimental and control groups using SPSS. Table 4.3 presents the interpretation of the correlation coefficient, which refers to the guideline by Sugiyono (2013) as cited in Azmi (2024). Table 4.4 displays the result of the Pearson Product Moment calculation for the experimental group's pre-test.

Table 4.3 Interpretation of Coefficient Value

Interval correlation	Correlation level
0.00 – 0.199	Very Low
0.20 – 0.399	Low
0.40 – 0.599	Sufficient
0.60 – 0.799	Strong
0.80 – 1.000	Very Strong

Table 4.4 Inter-Rater Pre-test of Experimental Group

Correlations

		rater1	rater2
Rater1	Pearson Correlation	1	.580**
	Sig. (2-tailed)		.000
	N	36	36
Rater2	Pearson Correlation	.580**	1
	Sig. (2-tailed)	.000	
	N	36	36

**. Correlation is significant at the 0.01 level (2-tailed).

Based on the results above, the correlation coefficient between the pre-test scores of the experimental group from the two raters is highly significant at the 0.01 level, with a correlation coefficient of 0.580.

Therefore, the correlation coefficient between the two raters is considered sufficient.

The result of the reliability analysis for the inter-rater scoring of the pretest conducted in the control group is presented in Table 4.5, which provides statistical information regarding the consistency of scores given by different raters.

Table 4.5 Inter-Rater Pre-Test Control Group

Correlations

		rater1	rater2
Rater1	Pearson Correlation	1	.680**
	Sig. (2-tailed)		.000
	N	36	36
Rater2	Pearson Correlation	.680**	1
	Sig. (2-tailed)	.000	
	N	36	36

**. Correlation is significant at the 0.01 level (2-tailed).

Based on the findings above, the correlation coefficient between the pre-test scores of the control group from the two raters is considered strong, with a correlation value of 0.680 and high significance at the 0.01 level.

The result of the reliability analysis for the inter-rater scoring of the post-test conducted in the experimental group is presented in Table 4.6, which provides statistical information regarding the consistency of scores given by different raters.

Table 4.6 Inter-rater Post-Test Experimental Group

Correlations

		rater1	rater2
Rater1	Pearson Correlation	1	.764**
	Sig. (2-tailed)		.000
	N	36	36
Rater2	Pearson Correlation	.764**	1
	Sig. (2-tailed)	.000	
	N	36	36

**. Correlation is significant at the 0.01 level (2-tailed).

Based on the results above, the correlation coefficient between the post-test scores of the experimental group from the two raters is highly significant at the 0.01 level, with a correlation coefficient of 0.764. Therefore, the correlation coefficient between the two raters is considered strong.

The result of the reliability analysis for the inter-rater scoring of the post-test conducted in the control group is presented in Table 4.7, which provides statistical information regarding the consistency of scores given by different raters.

Table 4.7 Inter-Rater Post-Test Control Group

Correlations

		rater1	rater2
Rater1	Pearson Correlation	1	.839**
	Sig. (2-tailed)		.000
	N	41	36
Rater2	Pearson Correlation	.839**	1
	Sig. (2-tailed)	.000	
	N	36	36

**. Correlation is significant at the 0.01 level (2-tailed).

Based on the findings above, the correlation coefficient between the post-test scores of the control group from the two raters is highly significant at the 0.01 level, with a correlation coefficient of 0.839. Therefore, the correlation coefficient between the two raters is considered very strong.

4. The Fulfillment of The ANCOVA Assumption

In this section, the researcher presents the results of the pre-test and post-test calculations. Then, the researcher analyzed the results to conduct an Analysis of Covariance (ANCOVA). One of the purposes of this analysis

is to determine the significant effect of using the Praktika.ai application on students' speaking skill development. Since certain conditions must be met for data processing, several assumptions need to be fulfilled before applying ANCOVA to analyze the data. First, the data distribution must be normal, the variances between the experimental and control groups must be homogeneous, there must be no interaction between the pre-test and the groups, and the relationship between the pre-test and post-test must be linear. The following is the result of the assumption tests.

5. The Result of Normal Distribution

The normality test was used by the researcher to determine whether the data came from a population with a normal distribution. The purpose of this normality test was to assess how well the data from both groups—the experimental class and the control class—were normally distributed. The Kolmogorov-Smirnov test was chosen because this study involved more than 50 samples. Therefore, in this situation, the Kolmogorov-Smirnov test was considered appropriate. This is in line with the statement by Mishra et al. (2019), who stated that the Shapiro Wilk test is more recommended for small samples, namely those with fewer than 50 observations, whereas the Kolmogorov Smirnov test is more suitable for larger samples, specifically those with more than 50. The data analysis in this study was conducted using SPSS version 23.

Table 4.8 Test of Normality

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Class	Statistic	df	Sig.	Statistic	df	Sig.
Result	Pre-Test A (experiment)	.122	36	.196	.963	36	.266
	Pre-test B (control)	.126	35	.176	.936	35	.141
	Post-Test A (experiment)	.110	36	.200*	.961	36	.238
	Post-Test B (control)	.122	35	.200*	.957	35	.185

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Based on the results shown in Table 4.8, it was found that the significance value (p-value) for the experimental class pre-test was 0.196, while the post-test was 0.200. Meanwhile, for the control class, the significance value for the pre-test was 0.176 and for the post-test was 0.200. All of these significance values are greater than 0.05, which indicates that all data in this study are normally distributed.

6. The Result of Homogeneity Test

After the initial assumption that the variances between groups are homogeneous, the Levene's test was used to determine the homogeneity of variances among the groups. A Levene's test result greater than 0.05 indicates that the variances between groups are homogeneous. The results of the variance homogeneity test can be seen in Table 4.9.

Table 4.9

Levene's Test of Equality of Error Variances[a]

Dependent Variable: Nilai Posttest			
F	df1	df2	Sig.

.651	1	69	.422

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.
a. Design: Intercept + PRETEST + CLASS + CLASS * PRETEST

The result of Levene's test using SPSS shows a significance value of 0.422, which is greater than 0.05. According to Priyatno (2018), if the significance value is higher than 0.05, it indicates that the data variance is homogeneous. Therefore, based on the result of the test, it can be concluded that the variance of the post-test scores between the groups is homogeneous.

7. The Result of Homogeneity Regression

Before conducting the ANCOVA test, a test of regression homogeneity, also known as the homogeneity of regression slopes, must be performed. This test evaluates the relationship between the covariate and the independent variable in predicting the dependent variable. A significant interaction between the covariate and the independent variable must be present, and the result should not exceed 0.05 in order to proceed with the ANCOVA analysis. This indicates that differences in the dependent variable between groups change as a result of the covariate's influence. If there is a significant interaction, the ANCOVA result becomes invalid and the test should not be conducted. Table 4.10 presents the result of the regression homogeneity test, also known as the slope test.

Table 4.10

Tests of Between-Subjects Effects

Dependent Variable: Nilai Posttest

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1285.520[a]	3	428.507	12.896	.000
Intercept	627.158	1	627.158	18.874	.000
PRETEST	805.438	1	805.438	24.239	.000
CLASS	2.482	1	2.482	.075	.785
CLASS * PRETEST	**14.965**	**1**	**14.965**	**.450**	**.504**
Error	2226.311	67	33.229		
Total	394531.000	71			
Corrected Total	3511.831	70			

a. R Squared = .366 (Adjusted R Squared = .338)

According to the table above, it can be seen that the significance value of the interaction between the covariate (pretest) and the independent variable (group), labeled as class pretest, is 0.504. Since the significance value is greater than 0.05, it means that there is no significant interaction between the covariate and the independent variable. This indicates that the assumption of homogeneity of regression slopes has been met. Therefore, the researcher can proceed to perform the ANCOVA analysis.

8. The Result of Test of Linear Relationship Covariate and Dependent Variable

The analysis uses a covariate to control the independent variable. For a specific group, the covariate test is used to control the dependent variable in order to evaluate the relationship between the covariate and the dependent variable. To proceed with the ANCOVA analysis, the p-value must be less

than 0.05. Table 4.11 shows the results of the linear relationship test between the covariate and the dependent variable.

Table 4.11

Tests of Between-Subjects Effects

Dependent Variable: Nilai Posttest

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1270.555[a]	2	635.277	19.274	.000
Intercept	621.070	1	621.070	18.843	.000
PRETEST	**810.117**	**1**	**810.117**	**24.579**	**.000**
GROUP	691.332	1	691.332	20.975	.000
Error	2241.276	68	32.960		
Total	394531.000	71			
Corrected Total	3511.831	70			

a. R Squared = .362 (Adjusted R Squared = .343)

According to the table above, it can be seen that the Sig. value of the group is 0.000, which is lower than 0.05 (0.000 < 0.05). This indicates that there is a significant relationship between the covariate and the dependent variable. Therefore, it can be concluded that the covariate is linearly related to the dependent variable, and the analysis can proceed using ANCOVA.

9. The Result of Hypothesis Testing

Based on the table 4.11, the significance value for the group is 0.000, which is lower than the threshold of 0.05. This result provides sufficient evidence to reject the null hypothesis and accept the alternative hypothesis. Therefore, it can be concluded that students who were taught using the Praktika.ai application as a strategy for teaching speaking performed significantly better than those who were taught without it. To strengthen this

finding, a Tests of Between-Subjects Effects analysis was also conducted, as shown in Table 4.12. This test aims to examine the influence of each independent variable on the post-test scores after controlling for the pre-test scores.

Table 4.12

Tests of Between-Subjects Effects

Dependent Variable: Nilai Posttest

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1270.555[a]	2	635.277	19.274	.000
Intercept	621.070	1	621.070	18.843	.000
PRETEST	810.117	1	810.117	24.579	.000
CLASS	**691.332**	**1**	**691.332**	**20.975**	**.000**
Error	2241.276	68	32.960		
Total	394531.000	71			
Corrected Total	3511.831	70			

a. R Squared = .362 (Adjusted R Squared = .343)

Based on Table 4.12, the significance value for the variable Class is 0.000, which is less than the significance threshold of 0.05. This indicates that there is a statistically significant difference in post-test scores between the groups, even after controlling for the pre-test scores. The F-value of 20.975 further supports this strong effect. This finding strengthens the conclusion that the use of the Praktika.ai application has a significant impact on students' speaking performance. In addition, the Pretest variable also shows a significant effect (Sig. = 0.000), meaning that students' prior ability contributed to their post-test performance. The R Squared value of 0.362 suggests that approximately 36.2% of the variance in post-test scores can be

explained by the model, including both the pre-test and the class treatment effects.

B. Discussion

The results of this study indicate that the Praktika application is effective in improving students' speaking skills. The experimental group that used the application obtained significantly higher post-test scores compared to the control group, with statistical analysis (ANCOVA) showing a significant difference (Sig. = 0.000 < 0.05). Students in the experimental class also demonstrated greater enthusiasm, confidence, and active participation during speaking activities. They responded positively to features such as AI-based avatars, real-time feedback, and interactive dialogues, which created an enjoyable. These factors likely contributed to their improvement, especially in aspects such as pronunciation, fluency, and comprehension.

The experimental class proved to be more effective in the learning process compared to the control class due to several reasons. The first is that the Praktika application allowed students to access learning materials anytime and from any location, which increased their opportunities to practice outside of class. This flexibility supported continuous learning and gave students the autonomy to study at their own pace. Donoghue (2015), stated that mobile learning enables learners to access resources anytime and anywhere, thus supporting more independent and self-directed learning. The second is that Praktika offers real-time feedback that helps learners immediately recognize and correct their speaking errors. This instant correction builds learners' awareness of their own performance and boosts their confidence. According to Godwin

(2019), timely feedback in mobile-assisted language learning significantly enhances learner autonomy and accuracy in speaking. The third is the application support for authentic and interactive communication, which exposes learners to realistic language use in meaningful contexts. This is consistent with the study by Sun et al. (2017), who found that students' speaking fluency improved more through mobile-based authentic communication than through traditional classroom instruction. These three features flexibility, immediate feedback, and authentic interaction collectively contributed to the better speaking outcomes observed in the experimental group.

The effectiveness of the Praktika application in enhancing students' speaking skills is supported by previous studies in the field of technology-assisted language learning. The finding of this study is in line with a study by Hanifah and lisan. (2024), showed that the use of applications such as Duolingo and ELSA Speak significantly improved students' speaking skills, including pronunciation and fluency, through structured practice and real-time feedback. ELSA Speak, in particular, provided immediate feedback that was effective in enhancing students' pronunciation, focusing on accurate articulation and intonation. In addition, research by Anggraini (2022) found that the use of the ELSA Speak application helped non-native English-speaking students improve their pronunciation and fluency. The application uses advanced voice recognition technology to assess users' pronunciation and provide feedback on accent, intonation, and other speech characteristics. Compared to those applications, Praktika offers a more immersive experience through real-life conversation simulations and automated correction. This provides students with

more speaking practice opportunities and encourages self-correction. Therefore, the findings of this study are consistent with and reinforce the results of previous research.