

CHAPTER III

RESEARCH METHOD

This chapter provides a description of the research method used in this research, which is further divided into several points. They are the research design, population and sample, research instrument, treatment procedure, data collection, and data analysis techniques.

A. Research Design

This study employed a quasi-experimental design which involves the comparison of two groups: an experimental group and a control group. Both groups were administered a pre-test to assess their initial vocabulary proficiency. Subsequently, the experimental group was exposed to vocabulary learning using the Duolingo application, while the control group received vocabulary instruction through flashcards guided by the teacher. After the intervention period, both groups completed a post-test to evaluate vocabulary improvement.

This study employed a quasi-experimental because it allows the researcher to compare the effectiveness of two different instructional methods (Duolingo vs. flashcards) in a real classroom setting where random assignment is not feasible. According to Zakiyah, Renita, and Mirma (2024), quasi-experimental designs are suitable for examining instructional innovations such as mobile-assisted language learning tools like Duolingo, as they provide a systematic and controlled approach to measure treatment effects on language acquisition while accommodating the natural constraints of educational environments.

B. Population and Sample

The population of this study consists of all VII grade students at MTsN 2 Kediri in the ongoing academic year, totaling 363 students spread across classes VII-A to VII-K with approximately 33 students in each class. These students come from diverse academic backgrounds, which makes this population relevant for this study which aims to test the effectiveness of different methods in vocabulary learning. The diversity in the population provides a comprehensive context for understanding how different instructional approaches can affect vocabulary acquisition among junior high school students.

Table 3. 1 Sample of the Study

No.	Class	Number of Students
1.	VII – F	36
2.	VII – G	37
Total		73

From the population, the researcher selected two classes, namely VII-F which consists of 36 students and VII-G which consists of 37 students as the research samples. Total number of research participants is 73 students. These two classes were purposively selected because they had relatively similar academic abilities and class dynamics, making them suitable for comparison in experimental research. Class VII-F was designated as the experimental group, where students learned vocabulary using the Duolingo application, while Class VII-G served as the control group, receiving vocabulary instruction through flashcards guided by the teacher.

C. Instrument of the Research

In conducting this research, the researcher uses appropriate instruments to collect accurate and reliable data. The instrument plays a crucial role in measuring the variables and obtaining information related to the students' vocabulary mastery. The assessment process consists of three stages.

1. Pre-test

The pre-test was administered before the treatment phase to assess students' initial vocabulary knowledge. The test consists of 25 multiple choice questions covering vocabulary relevant to the material being taught. The purpose of the pretest is to establish an initial measurement of students' vocabulary acquisition, which will later be compared with the post-test results to determine the learning progress.

2. Treatment

During the treatment stage, both groups of students received instruction with different methods. The experimental group (Class VII-F) learned vocabulary through the Duolingo app, which provides interactive and engaging exercises to improve vocabulary acquisition. Meanwhile, the control group (Class VII-G) was taught using teacher-guided flashcards, a more traditional method that focuses on repetition and memorization of vocabulary. This treatment period aimed to expose students to the respective instructional strategies for a specific duration.

3. Post-test

After the treatment phase, a post-test was administered to both groups to evaluate their vocabulary improvement. Just like the pre-test, the post-test consisted of 25 multiple-choice questions that assessed the same vocabulary material taught

during the study. The results of the post-test were used to determine the effectiveness of each learning method by comparing the students' scores before and after the treatment.

The blueprint of the vocabulary test, which outlines the indicators and the number of items for each aspect, is attached below. Moreover, the complete pretest and post-test instruments can be seen in *Appendix 1*.

Table 3. 2 Blueprint of the Pre-test and Post-Test

Objective	Indicators	Test Item for Pre-test	Test Item for Post-test
To identify main ideas and specific information in descriptive text	Students are able to identify main ideas and specific details from a simple descriptive text about daily habits (food and drink)	1–5	1–5
To understand procedural text (sequence and content)	Students are able to identify the purpose, ingredients, and steps in a simple procedure text	6–10	6–10
To identify specific information and meaning in descriptive text	Students are able to recognize specific details such as favorite items, time expressions, and taste descriptions	11–15	11–15
To understand procedural text and vocabulary in context	Students are able to understand steps and vocabulary related to actions in procedure text (e.g., peel, cut, pour)	16–20	16–20
To understand descriptive text and word meaning in context	Students are able to identify details and interpret word meanings (e.g., delicious, vegetables) in context	21–25	21–25

The questions were designed based on the indicators in the blueprints to measure the students' ability to identify, use, and interpret vocabulary related to food and drinks context. The blueprint of the post-test is generally the same as the pre-test in terms of objectives and indicators. Both tests measure students' reading comprehension on descriptive and procedural texts related to food and drink,

including identifying main ideas, specific information, and word meanings in context. However, the post-test items are designed with slightly higher cognitive demand and more varied contexts to assess students' improvement after the treatment.

The following scoring rubric explains how the raw scores were determined and converted into percentages to facilitate the analysis of student progress after the intervention.

Table 3. 3 Scoring Rubric

Item Number	Description	Score
1-25	The student answers the question accurately.	1
	The student fails to answer or answers incorrectly.	0
	Total Score	Total × 4

4. Instrument Validity

The validity of the pre-test instrument was examined through a pilot study involving 30 students who were not included in the research sample. The purpose of the try-out was to determine whether each item in the test was valid and appropriate for measuring students' vocabulary mastery.

The validity analysis was conducted using the Pearson Product Moment correlation. The obtained r-count values were then compared with the r-table value at the significance level of 0.05 (r-table = 0.361, N = 30). The result can be seen in the Table 3.4 Below.

Table 3. 4 The Result of Validity Test (Pre-test)

No.	R Count	R Table	Status	No.	R Count	R Table	Status
Q01	0,394	0,361	Valid	Q14	0,424	0,361	Valid
Q02	0,424	0,361	Valid	Q15	0,635	0,361	Valid

Q03	0,566	0,361	Valid	Q16	0,541	0,361	Valid
Q04	0,821	0,361	Valid	Q17	0,692	0,361	Valid
Q05	0,516	0,361	Valid	Q18	0,717	0,361	Valid
Q06	0,424	0,361	Valid	Q19	0,454	0,361	Valid
Q07	0,692	0,361	Valid	Q20	0,543	0,361	Valid
Q08	0,556	0,361	Valid	Q21	0,423	0,361	Valid
Q09	0,568	0,361	Valid	Q22	0,468	0,361	Valid
Q10	0,454	0,361	Valid	Q23	0,39	0,361	Valid
Q11	0,692	0,361	Valid	Q24	0,712	0,361	Valid
Q12	0,391	0,361	Valid	Q25	0,605	0,361	Valid
Q13	0,605	0,361	Valid				

The results of the analysis showed that all 25 items in the pre-test instrument were valid. This is indicated by the r-count values of all items being higher than the r-table value. The r-count values ranged from 0.391 to 0.821, which means that all items met the criteria of validity. Therefore, it can be concluded that all items in the pre-test instrument were valid and suitable to be used in this study.

5. Instrument Reliability

The reliability of a research instrument is a fundamental requirement to ensure that the measurement tool provides consistent and stable results over time. In this study, the reliability of the 25-item pre-test instrument was evaluated using the internal consistency method. Internal consistency refers to the degree to which all items in a test measure the same construct or concept. For an objective test with dichotomous scoring (correct/incorrect), Cronbach's Alpha is a robust statistical measure used to determine if the instrument is dependable enough to be used as a primary data collection tool in a quasi-experimental setting. The result of the test can be seen in the Table 3.5 below.

Table 3. 5 Reliability Statistics of the Pre-test Instrument

Reliability Statistics	
Cronbach's Alpha	N of Items
.898	25

Based on the statistical analysis presented in Table 3.5 above, the Cronbach's Alpha coefficient for the 25 test items is 0.898. According to the reliability criteria established by Hair et al. (2019), a coefficient ranging from 0.80 to 1.00 is categorized as "Very High Reliability." This result indicates that the instrument possesses excellent internal consistency, as the score significantly exceeds the minimum acceptable threshold of 0.70. Consequently, the pre-test instrument is deemed highly reliable and scientifically valid for measuring the speaking proficiency of the tenth-grade students, ensuring that any observed changes in scores during the post-test can be attributed to the experimental treatment rather than measurement error.

D. Treatment Procedure

The treatment procedure of this study is divided into three main phases: pre-teaching, whilst teaching, and post-teaching. Each phase is applied to both the experimental and control classes with different treatments. The detailed procedure is presented in the table below.

Table 3. 6 Activities During Treatments

Phase	Experimental Class	Control Class
Pre-Teaching	a. The teacher opens the class, greets the students, and checks attendance. b. The teacher presents the learning objectives and the outline of the vocabulary material. c. The teacher asks a warm-up question related to food and drinks to stimulate students' interest.	a. The teacher opens the class, greets the students, and checks attendance. b. The teacher presents the learning objectives and the outline of the vocabulary material. c. The teacher asks a warm-up question related to food and drinks to stimulate students' interest.
Main Teaching	a. The teacher explains the material first to provide a clear understanding of the vocabulary to be learned. b. The teacher guides the students to use the Duolingo application to learn and practice vocabulary. c. The students complete interactive vocabulary exercises and games within the application. d. The teacher monitors their progress and provides help when needed.	a. The teacher explains the material first to provide a clear understanding of the vocabulary to be learned. b. The teacher introduces vocabulary using c. flashcards. The teacher shows and pronounces each flashcard d. while the students repeat. The teacher plays games or drills with the flashcards to help students memorize e. vocabulary. The teacher monitors their progress and provides help when needed.
Post-Teaching	a. Students take a post-test to assess their vocabulary development after learning with Duolingo. b. The teacher briefly reviews the material that was discussed during the lesson. c. The teacher provides feedback and closes the session.	a. Students take a post-test to assess their vocabulary development after learning with flashcards. b. The teacher briefly reviews the material that was discussed during the lesson. c. The teacher provides feedback and closes the session.

The treatment in this study was conducted in five meetings for both the experimental and control groups. The first meeting was used for the pre-test,

followed by three treatment meetings, and the last meeting was used for the posttest. Each meeting lasted approximately 90 minutes. The treatment was conducted in three meetings to provide students with sufficient exposure to the instructional media used in this study. The schedule of the treatment can be seen in Table 3.7 below.

Table 3. 7 Treatment Schedule

Meeting	Experimental Group	Control Group	Duration
Meeting 1	Pre-test	Pre-test	90 minutes
Meeting 2	Learning vocabulary using Duolingo	Learning vocabulary using flashcards	90 minutes
Meeting 3	Learning vocabulary using Duolingo	Learning vocabulary using flashcards	90 minutes
Meeting 4	Learning vocabulary using Duolingo	Learning vocabulary using flashcards	90 minutes
Meeting 5	Post-test	Post-test	90 minutes

E. Data Collection

In this study, the test method was used to collect data on the vocabulary skills of MTsN 2 Kediri class VII-F and VII-G students, consisting of control class and experimental class. Both tests were conducted (before and after) the treatment, by both groups. The pre-test aims to measure students' initial vocabulary mastery and becomes the basic data to analyze the comparison of test results. After the treatment was completed, a post-test was conducted to find out whether there was a difference between the vocabulary skills of students using Duolingo in the experimental class and the traditional method in the control class. Several statistical requirement tests were used to process the data from these two test stages.

F. Data Analysis

The analysis of the data obtained from the pre-test and post-test included normality testing and hypothesis testing. The normality test was conducted using the SPSS program to determine whether the data were normally distributed, with a significance criterion of more than 0.05 (Ghasemi & Zahediasl, 2012). This test was used to determine the appropriate statistical method for further analysis.

Based on the results of the normality test, it was found that some of the data were not normally distributed. Therefore, a non-parametric statistical test was considered more appropriate for analyzing the data. Non-parametric tests do not require the assumption of normal distribution and are suitable for data that violate normality assumptions.

The main statistical analysis employed in this study was the Mann–Whitney U test. This test was conducted using the SPSS program to determine whether there was a significant difference in students' vocabulary mastery between the experimental group, which received the Duolingo intervention, and the control group, which was taught using flashcards. The Mann–Whitney U test is appropriate for comparing two independent groups when the data are not normally distributed.

To test the hypotheses of this study, the Mann–Whitney U test was applied at a significance level of 0.05. The hypotheses were formulated as follows:

- Null Hypothesis (H_0): There is no significant difference in vocabulary mastery between the experimental group (using Duolingo) and the control group (using flashcards).

- Alternative Hypothesis (H_1): There is a significant difference in vocabulary mastery between the experimental group (using Duolingo) and the control group (using flashcards).

If the significance value (p-value) is less than 0.05, the null hypothesis is rejected, indicating that there is a significant difference between the two groups. Conversely, if the significance value is greater than 0.05, the null hypothesis is accepted, indicating that there is no significant difference between the groups.