# CHAPTER III
# RESEARCH METHOD

## H. Research design

This research applied a quantitative approach as its methodological framework. According to Creswell (2017, p. 4), quantitative research is designed to test objective theories by analyzing the relationships between variables. This approach emphasizes objective measurement and relies on statistical, mathematical, or numerical analysis of data collected through instruments such as questionnaires, surveys, or existing datasets that have been systematically processed.

A quasi-experimental design was used in this study. Time series design and nonequivalent control group design are the two categories of quasi-experimental designs, according to Sugiyono (2010, p. 75). A nonequivalent control group design was utilized, with pretests administered to both groups before the intervention and posttests conducted afterward to evaluate changes.

The scheme of the nonequivalent control group design can be depicted as follows, per Sugiyono (2011, p. 116):

**Table 1: Scheme Research design**

| Class | Pretest | Treatment | Posttest |
|---|---|---|---|
| Experimental | 01 | X | 02 |
| Control | 03 | - | 04 |

**Source: Campbell, D. T., & Stanley, J. C. (1963)**

**Note :**

Experimental : Group of students who get teach with local culture material

Control : Group of students eho get conventional teaching or no given treatment

01 : Experimental group pretest result before given treatment

02 : Experimental group pretest result after given treatment

03 : Experimental group post test result before given treatment

04 : Experimental group post test result after given treatment

## I. Population and sample

### 1. Population

Creswell (2012, p. 381) defines a population as a collection of individuals who possess unique traits or shared characteristics that distinguish them from other groups. Similarly, Sugiyono (as cited in Supardi, 2013, p. 187) describes a population as a generalization region consisting of subjects or objects that have specific quantities and characteristics determined by the researcher to be analyzed and from which conclusions are drawn. Consequently, the population in this study comprises all students within the specified academic level who share the same educational

environment and curriculum relevant to vocabulary learning. Based on these definitions, population can be concluded as objects or subjects within a specific area meeting certain conditions related to the research problem or object. In this study, the population consisted of eighth-grade students from SMP Negeri 18 Bengkulu City. The population of this study consisted of 11 eighth-grade classes at SMP Negeri 18 Bengkulu City, comprising a total of 363 students, including 184 male and 179 female students.

**Table 2: Population of the research**

| No | Class | Male | Female | Total |
|---|---|---|---|---|
| 1 | 8.1 | 14 | 19 | 33 |
| 2 | 8.2 | 24 | 9 | 33 |
| 3 | 8.3 | 13 | 20 | 33 |
| 4 | 8.4 | 23 | 9 | 32 |
| 5 | 8.5 | 16 | 17 | 33 |
| 6 | 8.6 | 16 | 17 | 33 |
| 7 | 8.7 | 8 | 25 | 33 |
| 8 | 8.8 | 11 | 22 | 33 |
| 9 | 8.9 | 21 | 13 | 34 |

| | | | | |
|---|---|---|---|---|
| 10 | 8.10 | 21 | 12 | 33 |
| 11 | 8.11 | 17 | 11 | 28 |
| | Total | 184 | 179 | 358 |

**Source: SMPN 18 Kota Bengkulu**

2. **Sample**

A sample is a subset of the population chosen to provide data for the study and may be representative of the whole populace. Creswell (2012:142) defines a sample as a representative subset of the target population that is systematically selected and examined to enable researchers to draw valid generalizations about the entire population. The researcher employed the purposive sampling strategy to choose the sample. Sampling is a methodological process used to select representative elements from a population for data collection and analysis. The chosen group's members share the same attributes, including the same number of pupils, aptitude, age, grade point average, and English instructor. Two classes from SMP 18 Kota Bengkulu's 8.5 and 8.11 classes served as the researcher's samples for this investigation.

Purposive sampling was utilized in this study. Purposive sampling is a method for choosing data source samples depending on specific factors, according to Sugiyono (2013:218). In purposive sampling, subjects are

intentionally chosen according to specific attributes that align with the objectives and characteristics of the population under investigation. The researcher is already aware of these traits. As a result, all they have to do is link the sample units according to certain standards.

The researcher selected two specific classes from SMPN 18 Kota Bengkulu: Class VIII.5 as the experimental group (33 students) and Class VIII.11 as the control group (28 students), totaling 61 students. The primary criterion for selecting these two classes was their baseline equivalence or homogeneity in academic performance. To provide empirical evidence of this homogeneity, the researcher analyzed the students' latest English daily test scores. The statistical analysis showed that The mean score of Class VIII.5 was 78.18, while the mean score of Class VIII.11 was 78.21. The negligible difference of only 0.03 between the two means demonstrates that both classes possessed an almost identical level of vocabulary mastery before the treatment. This equivalent starting point is vital in a quasi-experimental study to ensure Internal Validity. By minimizing pre-existing disparities, the researcher can confidently assert that any significant improvement in the experimental group is the result of the Spelling Bee Game intervention rather than inherent differences in students' initial abilities.

**Table 3 sample of the research**

| No | Class | Male | Female | Total |
|---|---|---|---|---|
| 1 | 8.5 Experiment Class | 16 | 17 | 33 |
| 2 | 8.11 Control Class | 17 | 11 | 28 |
| | Total | 33 | 33 | 61 |

Source: Campbell, D. T., & Stanley, J. C. (1963).

## J. Place and time of the research

This study was carried out with eighth-grade students of SMP Negeri 18 Bengkulu City, situated on Jalan KS. Tubun, Bengkulu City. While the subjects of this study were two eighth grade classes of SMP 18 Bengkulu City.

## K. Operational definition of variables

The operational definition of research variables, according to Sugiyono (2010:38), is the components or values derived from actions or objects with specific changes that would subsequently be determined by the researcher to be examined and conclusions formed from.

1. Spelling bee game

   Spelling Bee is a competition that tests spelling skills and enhances linguistic abilities. It helps participants improve their understanding of language structure and vocabulary, contributing to their overall linguistic development (Cowie, 2024, p. 56).

2. Vocabulary

"The body of words that a person or group knows and uses is referred to as their vocabulary. This includes both productive (words utilized in speech or writing) and receptive (words understood when heard or read) vocabulary. It is emphasized that vocabulary development is crucial for academic success, as a well-developed vocabulary enhances an individual's ability to communicate effectively and comprehend complex texts (Snow, 2024, p. 112).

**L. Data collection technique**

1. Test

Each of the samples in this study had a pre-test and a post-test. Both the experimental class and the control class took the pretest. Students were given tests in order to gather data. Gathering information to ascertain the study's outcomes is one of the most crucial phases in the research process. There were multiple steps involved in the data collection process. The pre-test and post-test results provided the data for this investigation. to determine whether the spelling bee activity at SMP Negeri 18 Bengkulu City can be used to teach vocabulary. Additionally, a comparison was made between the pre-test and post-test results. Researchers examined students' reading skills. Prior to the pupils taking the test..

a. Pre-Test

Creswell (2012, p. 297) explains that a pretest is used to measure participants' initial attributes or abilities prior to the administration of an experimental treatment. In this study, the pretest was administered once to both the experimental and control groups to assess students' reading comprehension before the treatment was implemented. The test consisted of 30 multiple-choice questions.

b. Post-Test

As defined by Creswell (2012, p. 297), a post-test is an assessment used to measure specific attributes or outcomes of participants after the experimental treatment has been administered. The purpose of the post-test is to ascertain how much the pupils have improved their reading comprehension following treatment. The pre-test and post-test questions have the same points. A post-test is used to gauge how well students have improved their academic reading comprehension following treatment. By analyzing the post-test results, the researcher can determine whether a statistically significant difference exists in students' achievement between the experimental and control groups. The researcher assessed the students'

pre-test and post-test results and put them in a rank order after administering the tests to the two groups. Each of the test's thirty items had a score of 1 for the correct response and 0 for the incorrect one. If every question was answered correctly, the students would receive a score of 100. The following formula was employed by the researcher to determine the students' score: $\frac{Correct\ Answer\ X\ 100}{Maximal\ Score}$

**Table 4: scoring Criteria**

| The range of score | Scores category | |
|---|---|---|
| 85-100 | Excellent | A |
| 75-84 | Good | B |
| 56-74 | Fair | C |
| <55 | Poor | D |

c. Treatment

The experiment's action is referred to as treatment, which is setting up circumstances that will be evaluated for their impact. When conducting experimental research, it is important to set up the experimental group and control group so that both variables have similar or identical properties. Here, two groups were given the material, namely the

experimental group and the control group. The experimental group received instruction through the Spelling Bee Game, whereas the control group was taught using conventional teaching methods. In this study, there were four meetings and each meeting consisted of 30 minutes.

2. Vocabulary test

In this study, multiple-choice tests are used to assess students' vocabulary mastery. This type of test is chosen because it provides objective scoring and allows for the evaluation of a wide range of vocabulary within a relatively short period. Additionally, the answer choices are designed to minimize the chances of students guessing without understanding the meaning of the words. According to Nation (2001, p. 347), multiple-choice tests are effective for measuring vocabulary knowledge as they allow for objective assessment, cover a broad range of vocabulary items in a single test, and reduce random guessing by providing well-selected distractors. Given these advantages, this method is considered appropriate for assessing students' vocabulary proficiency in this study.

At the beginning of the study, a pre-test comprising 26 multiple-choice questions was administered to determine students' vocabulary mastery before the treatment was

applied. After the pre-test, students participated in learning activities using the Spelling Bee Game for four meetings. The Spelling Bee Game is a vocabulary-based game that requires students to correctly spell and pronounce words within a limited time, which helps improve their focus, spelling accuracy, and word recognition. After the students took part in learning with the Spelling Bee Game, they were given a post-test. At the outset of the study, a pre-test consisting of 26 multiple-choice items was administered to measure students' vocabulary mastery prior to the implementation of the treatment.

Each question had only one correct answer. The scoring was done in a simple way, correct answers were given a score of 1, and incorrect answers were given a score of 0. After all students completed the pre-test and post-test questions, the researcher collected and calculated the pretest and post-test scores of each student. Furthermore, Improvements were identified by comparing pre-test and post-test scores, and the effectiveness of the Spelling Bee Game was evaluated through statistical analysis.

**Table 5: Score Interval for English Teaching Vocabulary**

| Range of Score | Category | Description |
|---|---|---|
| 20 – 18 | Excellent | Sophisticated range, effective word/idiom choice and usage. |
| 17 – 14 | Good to Average | Adequate range, occasional errors of word/idiom form, choice, or usage, but meaning not obscured. |
| 13 – 10 | Fair to Poor | Limited range, frequent errors of word/idiom form, choice, or usage. |
| 9 – 7 | Very Poor | Essentially translation level, little knowledge of English vocabulary. |

**Source: Adapted from J. B. Heaton (1991, p. 146)**

3. Documentation

Documentation was a data collection technique that gathers information from various types of documents or written sources. These can be reports, notes, texts, archives, journals, or other recordings relevant to the research objectives. In this study, documentation was needed to collect data relevant to the title (Amelia et al., 2023, p. 45). The documentation includes a list of names

of students who are the subject of the study, as well as grades needed for the study. The documentation also includes photographs of learning activities and pre- and post-test results on the given material.

**M. Research instruments**

"Because they are used to assess the variables under investigation, instruments are essential to quantitative research (Sugiyono, 2019, p. 156). A vocabulary test with a pre-test and a post-test was used in this investigation. The test was created to collect information about students' language mastery-related knowledge, abilities, attitudes, beliefs, and preferences (Olajide, 2018, p. 214). Its main goal was to assess how well students' vocabulary mastery was improved by the Spelling Bee game. Both the experimental and control groups took the test, which consisted of twenty items. The instrument met the necessary eligibility requirements for use in both the pre-test and post-test after undergoing validity and reliability testing prior to implementation to guarantee its correctness and consistency (Kimberlin & Winterstein, 2008).

**N. Data analysis technique**

he prerequisite data analysis in this study comprised tests of normality and homogeneity, followed by a t-test, N-Gain analysis, and hypothesis testing.

1. **Validity of the test**

Before the researcher conducts the research, the researcher conducts the validity of the instrument first. The instrument to be tested must show suitability in the aspects to be tested. Validity is fundamental to research integrity. According to Kimberlin and Winterstein (2008, p. 2278), validity is the extent to which a study accurately measures what it is intended to measure. This is further supported by Kouam Arthur William (2024, p. 12), who emphasizes that validity ensures the instruments used are relevant to the research objectives and capable of producing credible data.

**Table 6. Test Instrument Validity Criteria**

| No | Cronbach's Alpha | Validity Level |
|---|---|---|
| 1 | 0.81 – 1.00 | Highly |
| 2 | 0.61 – 0.80 | High |
| 3 | 0.41 – 0.60 | Fairly |
| 4 | 0.21 – 0.40 | Low |
| 5 | 0.00 – 0.20 | Very Low |

**Source: Adapted from Widodo et al. (2023, p. 92)**

In addition, researcher also analyzed the test results used computer statistics such as SPSS version 25. And the test results consisted of 65 questions containing three vocabulary indicators. following were the results of the data analysis of the instrument validity test:

**Table 7: Instrumen Validity**

| Item | r_hitung | r_tabel | Sig. (p) | Keterangan |
|---|---|---|---|---|
| Soal 1 | 0.822 | 0.468 | 0.0 | Valid |
| Soal 2 | 0.766 | 0.468 | 0.0 | Valid |
| Soal 3 | 0.175 | 0.468 | 0.46 | Tidak Valid |
| Soal 4 | 0.239 | 0.468 | 0.31 | Tidak Valid |
| Soal 5 | 0.716 | 0.468 | 0.0 | Valid |
| Soal 6 | 0.011 | 0.468 | 0.962 | Tidak Valid |
| Soal 7 | 0.638 | 0.468 | 0.002 | Valid |
| Soal 8 | 0.315 | 0.468 | 0.176 | Tidak Valid |
| Soal 9 | 0.51 | 0.468 | 0.022 | Valid |
| Soal 10 | 0.814 | 0.468 | 0.0 | Valid |
| Soal 11 | -0.242 | 0.468 | 0.305 | Tidak Valid |
| Soal 12 | 0.423 | 0.468 | 0.063 | Tidak Valid |
| Soal 13 | 0.665 | 0.468 | 0.001 | Valid |
| Soal 14 | -0.191 | 0.468 | 0.42 | Tidak Valid |
| Soal 15 | 0.387 | 0.468 | 0.092 | Tidak Valid |

| | | | | |
|---|---|---|---|---|
| Soal 16 | 0.67 | 0.468 | 0.001 | Valid |
| Soal 17 | 0.366 | 0.468 | 0.113 | Tidak Valid |
| Soal 18 | 0.405 | 0.468 | 0.076 | Tidak Valid |
| Soal 19 | -0.057 | 0.468 | 0.811 | Tidak Valid |
| Soal 20 | 0.273 | 0.468 | 0.244 | Tidak Valid |
| Soal 21 | 0.366 | 0.468 | 0.113 | Tidak Valid |
| Soal 22 | -0.094 | 0.468 | 0.694 | Tidak Valid |
| Soal 23 | 0.827 | 0.468 | 0.0 | Valid |
| Soal 24 | 0.678 | 0.468 | 0.001 | Valid |
| Soal 25 | 0.011 | 0.468 | 0.962 | Tidak Valid |
| Soal 26 | 0.011 | 0.468 | 0.962 | Tidak Valid |
| Soal 27 | 0.822 | 0.468 | 0.0 | Valid |
| Soal 28 | 0.805 | 0.468 | 0.0 | Valid |
| Soal 29 | 0.127 | 0.468 | 0.594 | Tidak Valid |
| Soal 30 | 0.452 | 0.468 | 0.045 | Valid |
| Soal 31 | 0.899 | 0.468 | 0.0 | Valid |
| Soal 32 | -0.065 | 0.468 | 0.787 | Tidak Valid |

| | | | | |
|---|---|---|---|---|
| Soal 33 | 0.602 | 0.468 | 0.005 | Valid |
| Soal 34 | 0.366 | 0.468 | 0.113 | Tidak Valid |
| Soal 35 | 0.826 | 0.468 | 0.0 | Valid |
| Soal 36 | 0.072 | 0.468 | 0.764 | Tidak Valid |
| Soal 37 | 0.638 | 0.468 | 0.002 | Valid |
| Soal 38 | 0.138 | 0.468 | 0.562 | Tidak Valid |
| Soal 39 | 0.035 | 0.468 | 0.884 | Tidak Valid |
| Soal 40 | 0.091 | 0.468 | 0.704 | Tidak Valid |
| Soal 41 | 0.09 | 0.468 | 0.706 | Tidak Valid |
| Soal 42 | 0.5 | 0.468 | 0.025 | Valid |
| Soal 43 | 0.138 | 0.468 | 0.562 | Tidak Valid |
| Soal 44 | 0.052 | 0.468 | 0.829 | Tidak Valid |
| Soal 45 | -0.388 | 0.468 | 0.091 | Tidak Valid |
| Soal 46 | 0.711 | 0.468 | 0.0 | Valid |
| Soal 47 | -0.153 | 0.468 | 0.519 | Tidak Valid |
| Soal 48 | 0.659 | 0.468 | 0.002 | Valid |
| Soal 49 | 0.587 | 0.468 | 0.007 | Valid |

| | | | | |
|---|---|---|---|---|
| Soal 50 | -0.143 | 0.468 | 0.547 | Tidak Valid |
| Soal 51 | -0.299 | 0.468 | 0.201 | Tidak Valid |
| Soal 52 | 0.48 | 0.468 | 0.032 | Valid |
| Soal 53 | 0.368 | 0.468 | 0.11 | Tidak Valid |
| Soal 54 | -0.039 | 0.468 | 0.872 | Tidak Valid |
| Soal 55 | 0.493 | 0.468 | 0.027 | Valid |
| Soal 56 | 0.661 | 0.468 | 0.002 | Valid |
| Soal 57 | 0.245 | 0.468 | 0.297 | Tidak Valid |
| Soal 58 | 0.545 | 0.468 | 0.013 | Valid |
| Soal 59 | 0.177 | 0.468 | 0.456 | Tidak Valid |
| Soal 60 | 0.671 | 0.468 | 0.001 | Valid |

A total of 60 multiple-choice questions were created. After the try-out, 26 questions were declared valid. A question is considered valid if the calculated r-value is greater than the r-table value, and it is also valid if the significance value is less than 0.05.

In this validity test, the researcher employed 60 multiple-choice items developed based on predetermined indicators. The instrument try-out was conducted at SMP Negeri 5 Kota

Bengkulu, which has an accreditation level equivalent to the school selected as the main research site. The trial involved 20 students who served as participants for the instrument testing. Upon completion of the trial, the researcher analyzed the data by calculating the validity of each item. The analysis results indicated that 26 out of the 60 items were considered valid, with the following indicators:

**Table 8 : Test Indicator**

| No | Test item | Test indicator |
|---|---|---|
| 1 | What is the meaning of ***apple?***<br>a) A type of drink c) A kind of animal<br>b) A type of fruit d) A piece of furniture | Students can define nouns (fruit) in English. |
| 2 | What do you ***use to write?***<br>a) Eraser c) Pencil<br>b) Notebook d) Chair | Students can identify objects related to writing. |
| 3 | What do you ***drink in the morning?***<br>a) Coffee c) Chair<br>b) Pencil d) Shirt | Students can identify beverages in daily life. |

| | | |
|---|---|---|
| 4 | What do you do ***when you are tired?***<br>a) Sleep c) Swim<br>b) Cook d) Talk | Students can identify daily activities. |
| 5 | What do ***birds do?***<br>a) Swim c) Run<br>b) Fly d) Talk | Students can define verbs (animals' actions). |
| 6 | What do you do ***when you are thirsty?***<br>a) Eat c) Walk<br>b) Drink d) Laugh | Students can identify actions related to daily needs. |
| 7 | The opposite of ***cold*** is…<br>a) Slow c) Heavy<br>b) Hot d) Quiet | Students can identify adjectives (antonyms). |
| 8 | ***Chocolate*** is…<br>a) Bitter c) Salty<br>b) Sweet d) Spicy | Students can identify adjectives (taste). |
| 9 | What do you do with a ***camera***?<br>a) Cook c) Sleep | Students can identify the use of objects. |

| | | |
|---|---|---|
| | b) Take photos d) Wash clothes | |
| 10 | What ***color is the sky on a clear day?*** a) Green c) Yellow b) Blue d) Red | Students can identify colors in nature. |
| 11 | Which one is a ***day*** of the week? a) January c) Summer b) Monday d) December | Students can identify days of the week. |
| 12 | What do you use to tell ***time?*** a) Clock c) Pencil b) Lamp d) Chair | Students can identify objects related to time. |
| 13 | What do you wear to protect your ***eyes from the sun?*** a) Scarf c) Hat b) Sunglasses d) Socks | Students can identify clothing items for specific purposes. |
| 14 | What is the ***synonym of happy?*** a) Sad c) Angry b) Glad d) Tired | Students can identify synonyms (emotions). |

| | | |
|---|---|---|
| 15 | Which one is a ***fruit?***<br>a) Carrot c) Potato<br>b) Apple d) Cabbage | Students can identify nouns (fruits). |
| 16 | Which word is the ***correct spelling?***<br>a) Environmant c) Environment<br>b) Enviroment d) Envaironment | Students can identify correct spelling. |
| 17 | What is the ***opposite of expensive?***<br>a) Price c) Cost<br>b) Cheap d) Money | Students can identify antonyms (adjectives). |
| 18 | What is the ***synonym of angry?***<br>a) Happy c) Glad<br>b) Mad d) Silent | Students can identify synonyms (emotions). |
| 19 | Which word is the ***synonym of beautiful?***<br>a) Ugly c) Bad<br>b) Pretty d) Plain | Students can identify synonyms (adjectives). |
| 20 | Which word means ***rumah sakit in English?*** | Students can translate |

| | | |
|---|---|---|
| | a) Library c) School<br>b) Hospital d) Market | Indonesian to English nouns. |
| 21 | What is the ***synonym of child?***<br>a) Adult c) Kid<br>b) Baby d) Parent | Students can identify synonyms (nouns). |
| 22 | Which is a ***color?***<br>a) Blue c) Book<br>b) Bread d) Bag | Students can identify colors. |
| 23 | What is the ***opposite of easy?***<br>a) Hard c) Smooth<br>b) Soft d) Nice | Students can identify antonyms (adjectives). |
| 24 | ***Which one is a day?***<br>a) April c) Year<br>b) Friday d) Summer | Students can identify days of the week. |
| 25 | What is the ***opposite of light?***<br>a) Heavy c) Easy<br>b) Bright d) Dark | Students can identify antonyms (adjectives). |
| 26 | Which word is the ***correct spelling?***<br>a) Hospital c) Hospptal<br>b) Hospittal d) Hosspital | Students can identify correct spelling. |

## 2. Reliability

Reliability is a fundamental aspect of research instrumentation. According to Kouam Arthur William (2024, p. 15), reliability refers to the consistency and stability of the results produced by a measurement tool across different instances of testing. After conducting the validity test, several items met the validity criteria. Therefore, the next step was the reliability test. The reliability value was measured using Cronbach's Alpha. If the Cronbach's Alpha value is greater than the r-table, then the variable can be considered reliable.

**Table 9: Test Instrument Reliability Criteria**

| No | Cronbach's Alpha | Reliability Level |
|---|---|---|
| 1 | 0.00 – 0.20 | Not Reliable |
| 2 | 0.20 – 0.40 | Less Reliable |
| 3 | 0.40 – 0.60 | Fairly Reliable |
| 4 | 0.60 – 0.80 | Reliable |
| 5 | 0.80 – 1.00 | Highly Reliable |

Source: Adapted from Nuzulia (2019, p. 74)

The researchers analyzed the reliability value using SPSS version 25. This reliability test was conducted to determine whether the research instrument produced consistent results. The Cronbach's Alpha technique, which measures how closely connected the instrument's items are to one another, was used in the computation. The dependability

coefficient value (Cronbach's Alpha) was displayed in the computation results as follows:

**Table 10 : Realibility**

| Reliability Statistics | |
|---|---|
| Cronbach's Alpha | N of Items |
| .902 | 60 |

In this reliability test, the researcher used 65 multiple-choice questions with predetermined indicators. The instrument was tested at SMP Negeri 5 Kota Bengkulu, a school with an accreditation level equivalent to the main research location, namely SMP 2 Kota Bengkulu. The trial involved 20 students as test subjects. After the trial was completed, the researcher analyzed the test results by calculating the reliability. The calculation showed a reliability coefficient of 0.902. Based on the established criteria, an instrument is considered highly reliable if it reaches a value of 0.600 or higher (Setyaedhi, 2024, p. 112). Thus, it can be concluded that a test meeting this threshold possesses a high degree of dependability (Herlawan et al., 2021, p. 88), ensuring that the collected data is consistent and trustworthy for further analysis.

.

**3. The level difficulty test**

One important component in developing research instruments is the item difficulty test. Item difficulty refers to analyzing test questions in terms of how easy or difficult they are for students (Magdalena et al., 2021, p. 154). This analysis was done to make sure that every test item was suitable for the students' skills and to assess the impact of the Wordscapes application on vocabulary mastering. Students who finished the vocabulary test provided the data, which was then processed to classify the items as easy, moderate, or tough. The number of students who successfully answered the test is divided by the total number of test takers to determine the difficulty index. The general formula is:

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

Explanation :

P = level of difficulty in percentage

Nh = number of test participants in the high-score group

Rh = number of correct answers in the high-score group

Ni = number of test participants in the low-score group

Ri = number of correct answers in the low-score group

$$P = \frac{465+144}{6+6} \text{ x } 100\%$$

$$P = \frac{609}{12} \text{ x } 100\%$$

P = 50,7 %

**Table 11 : Interpretation of difficult question categories**

| Question category | Interpretation of results |
|---|---|
| Classified as easy | 0,71% ---------- 100% |
| Classified as medium | 0,31% ---------- 0,70% |
| Classified as difficult | 0% ---------- 0,30% |

**Source: Adapted from Gde Yadnyawati (2019, p. 68)**

In testing research instruments, it is important to ensure that the level of difficulty of the items is balanced so that the questions are neither too easy nor too difficult. Based on the calculation of the difficulty index, the test items achieved a value of 50,7%. According to the established criteria, this percentage falls into the medium category (31%–70%). Therefore, the test items can be considered appropriate for use in the research.

**4. Item discrimination test**

The degree to which an item may differentiate between students with high ability (upper group) and those with low ability (lower group) in comprehending the subject matter being evaluated is known as differentiating power (Pradita et al., 2023, p. 115). This exam's objective is to ascertain whether the vocabulary test items can distinguish between students who have a high and low level of English vocabulary proficiency. Items with high discriminating power indicate a clear

distinction between students who have mastered the vocabulary and those who have not. In this study, the discriminating power test was carried out by dividing students into two groups based on their vocabulary test scores: the upper group consisted of students with the highest scores, while the lower group consisted of students with the lowest scores. The general formula is:

$$DP\frac{u-L}{N}$$

Definition:

Dp = Item defferential power test

U = Number of correct answers from the high-score group

L = Number of correct answers from the low-score group

N = Number of students in each group.

$$DP\ \frac{465-144}{6}$$

DP = 53,5

**Table 12. Interpretation of the Power Difference Test**

| T-test | Interpretation of results |
|---|---|
| 0,71 ≤ DP ≤1,00 | Very good |
| 0,41 ≤ DP ≤ 0,70 | Good |
| If +0,21 < DP < 0,40 | Sufficiently qualified |
| If 0 < DP ≤ 0,20 | Weak |

**Source: Adapted from Desman et al. (2025, p. 104)**

The distinguishing power test was conducted to ensure that the instrument could effectively differentiate the level of vocabulary mastery among students. In this research, the calculation of distinguishing power reached 53,5%. Based on the applicable criteria, which classify values within the range of 20% ≤ DP < 40% as "Good," it can be concluded that the distinguishing power of the test items in this study was sufficiently qualified.

**5. Normality test**

Normality testing is a statistical method to determine whether data follows a normal distribution, which is a fundamental requirement in parametric statistical analysis (Sihotang, 2023, p. 118). A test was often used to determined if the data was normal. To make sure the instrument was normal and whether the research was generally excellent,

normality tests were carried out. SPSS is used by researchers to assess and determine normalcy. The Normality Test then has the following testing criteria: a. It was determined that the data was not normally distributed when the Sig Value (P Value) was less than 0.05. b. The data has a normal distribution if the Sig Value (P Value) is greater than 0.05.

**6. Homogenity test**

The homogeneity test is conducted to ascertain whether data from two or more groups possess the same variation (variance). This process is essential for ensuring a valid parametric statistical analysis (Sihotang, 2023, p. 121). The homogeneity test determines whether or not population data was homogeneous. Use the following principles to determine homogeneity:

a. Significance test (α) = 0.05

b. If sig>α, then the variance of each sample is the same (homogeneity).

c. If sig<α, then the variance of each sample is the not same (not homogeneity) to determine it the researcher uses SPSS.

**7. T-test**

The T-test is an effective statistical tool; however, its effectiveness may be compromised under certain conditions, such as when the data is not normally distributed or when it is applied to mismatched data sets (Novak, 2022, p. 189). The

analysis will be carried out using the SPSS T-test, the test was carried out to determine how significant the partial effect of the independent factor was on the dependent variable when the other independent variables are assumed to be constant. and the formula to be used is as follows.

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).

**8. F-test**

George W. Snedecor named this statistical tool the F-test in honor of its inventor, Sir Ronald Fisher, a renowned statistician. The F-test is a crucial tool in statistics used to compare two or more variances, measuring the extent to which data is spread across different groups (Odek & Opuodho, 2023, p. 42). One of the most commonly used applications of the F test is in Analysis of Covariance (ANCOVA). ANCOVA is used to compare the means of two or more groups by taking into account one or more control variables (called covariates) that can affect the results of the analysis. With covariates, ANCOVA can reduce the influence of outside variables that are not the main focus of the research, so that the results of comparisons between groups become more accurate and objective. This technique is very useful in experimental research, especially when

trying to determine the effectiveness of a particular treatment or method.

**9. The effect size**

The effect size is a measure used to describe how much influence or relationship between two variables in research. This measure was important because it provides information about how large or strong the effect was not just whether or not it exists. Two commonly used effect sizes are Cohen's d, which measures the mean difference between two groups, and Pearson's r, which measures the strength of the relationship between two variables (Funder & Ozer, 2019, p. 157). The following was the category of effect size values used to measure how much influence or relationship between two variables in research:

**Table 13. Value of the effect size**

| No | Velue | Effect size |
|---|---|---|
| 1. | 0-00-0,1,95 | very weak effect |
| 2. | 0,20-0,395 | weak effect |
| 3. | 0,40-0,595 | modest effect |
| 4. | 0,60-0,749 | strong effect |
| 5. | 0,80-1,00 | very strong effect |

**Source: Adapted from Cohen (1988) in Funder & Ozer (2019)**

**10. Statistical**

Hypothesis Hypothesis testing functions in testing Ho was accepted or rejected and Testing the alternative Ha was accepted or rejected. Hypothesis testing can also be done using SPSS version 25 with basic guidelines for collecting the following test results:

a. If the value of hig> or Sig. (2-tailed) < 0.05 it is said that the hypothesis is accepted.

b. If the hig value < or Sig.(2-tailed) > 0.05 it is said that the hypothesis is rejected.