

CHAPTER III

METHODOLOGY

A. Research Design

In this study the effect of Elsa Speak for increasing speaking ability in junior high school at SMPN 4 Bengkulu was investigated using quantitative research method. Quantitative method can be defined as a research method based on rationalism positivist philosophy, used to research on particular population or sample, data collection using research instruments, analysis of quantitative methods (Susanto & Jaya, 2023). The quantitative approach that was applied in this research was the quasi-experimental method, which aimed to determine the impact of using Elsa Speak in improving speaking ability. According to Creswell, a quantitative approach was a type of research that starts with a cause-and-effect hypothesis. Then, the author administered a test followed by the administration of the treatment to the experimental class, and the final test was conducted through a post test (Creswell, 2012). This research design was used to organize the implementation of this research in order to get accurate data. In conducting this research, the authors used a quasi-experimental design. This research design has been widely used by researchers. The main purpose of the research was to test the impact of a treatment on the results obtained, by controlling all the factors that affected these results.

According to Sugiyono the scheme of non-equivalent control design can be described as table 3.1 (Sugiyono, 2020).

Table 3. 1 Research design.

Group	Pretest	Treatment	Post test
A	0 ₁	X	0 ₂
B	0 ₃	-	0 ₄

(Adapted by
Sugiyono, 2020)

Definition :

Experiment = Without treatment (learning treatment with application)

Control = Treatment (learning treatment with application)

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

B. Place and Time of Research

This research was conducted at SMPN 4 Bengkulu City, for seventh grade students in the even semester of the 2025/2026 academic year on Jl. Gedang, Kec. Gading Cempaka Bengkulu City. Firstly, the author observed that

SMPN 4 Bengkulu City was a school full of achievements, even this school was one of the favorite schools in Bengkulu City, with the nickname child-friendly. Many students wanted to attend this school, because this school had created many outstanding students in the academic and non-academic fields. For class VIII there were 10 classes. Also, this school had complete facilities, and this research was carried out in October 6 to November 6, 2025.

The first step taken by the author was to submit a permit application to the Head of the Bengkulu City National Unity and Politics Agency (KESBANGPOL) and the Education and Culture Office (DIKNAS) on October 1, 2025. Then, a permit application was submitted to SMPN 4 Bengkulu City on October 3, 2025. This request was made to obtain permission to conduct research in the school environment. With this permission obtained, the author conducted the research according to the procedures that had been planned beforehand in the experimental and control classes.

This study was conducted from October 6 to November 6, 2025, and involved four meetings with the experimental group who were given treatment using the Elsa Speak application. The first meeting was held on October 14, 2025, to conduct a “Pre-Test” for both classes. The second meeting was held on October 21, 2025, to discuss “self-introduction with the application.” The third meeting was held on October 28, 2025, to discuss “self-introduction with materials and the application.” The

fourth meeting was held on November 3 to conduct a “post-test.”

C. Population and Sample

This research was conducted at SMPN 4 Bengkulu city, this research is located at Jl. Gedang Kec. Gading Cempaka Bengkulu City.

1. Population

According to Sugiyono population was a generalization area consisting of objects or subjects that had certain qualities and characteristics chosen by researchers to study and then draw conclusions (Sugiyono, 2020). And then, (Stadtländer, 2009) in Creswell also stated that population was the subject of research as a whole, and that samples can be taken from subjects that are too large for research. From some of the above opinions, it can be concluded that population referred to objects or subjects that are at the research site and met certain conditions which are certainly related to the problem or object of research. The author took the population in the VIII grade student class of SMPN 4 Bengkulu City, school year 2025/2026.

The population in this study were VIII grade students of SMPN 4 Bengkulu City, totaling 318 students. In sampling, the author choosed class VIII 4

as the control class totaling 34 people, and class VIII 1 as the experimental class totaling 34 students.

Table 3. 2 Population of the research.

No	Class	Female
1	VIII 1	34
2	VIII 2	35
3	VIII 3	34
4	VIII 4	34
5	VIII 5	34
6	VIII 6	34
7	VIII 7	34
8	VIII 8	34
9	VIII 9	34
10	VIII 10	34

(Adapted by SMPN 4 2025)

2. Sample

Part of the population was a sample, which was used as a source of data to represent the population as a whole. According to Cresswel a sample was the target group of the population that the author studied to generalize about the target population (Cresswel, 2018). In determining the sample, the author used purposive sampling technique. Of course, there were several criteria for selecting samples. The selected

groups had the same characteristics both from the same number of students, the same subject teacher, the average value, and the ability possessed. The samples that selected in this study were two classes from class VIII of SMPN 4 Bengkulu City.

In this research, the author used purposive sampling. Purposive sampling was a technique used in determining samples with certain considerations (Sugiyono, 2020). In purposive sampling, two class groups selected according to the criteria that the author had determined. The author took two classes divided them into 2 groups. The first class was the experimental class, namely (VIII 4) totaling 30 students and the second class was the control class, namely (VIII 5) totaling 30 people.

Table 3. 3 The sample of the Research.

No	CLASS	FEMALE
1.	VIII 1	34
2.	VIII 4	34
	Total	68

(Adapted by SMPN 4 2025)

D. Research Instrument

The main instrument in this research was a speaking test, intended to measure students' speaking abilities before and after using the Elsa Speak application. According to Skehan,

there were four main dimensions of speaking ability in English that can be used to assess someone's speaking skills: fluency, accuracy, pronunciation, and complexity. Fluency referred to the smoothness of speech characterized by minimal pauses and disruptions in language production. Accuracy referred to the correctness of grammatical structures and vocabulary use in speech.

However, complexity was related to the variety and sophistication of the language structures used by students. These dimensions were considered important in measuring speaking ability because they reflected the overall development of language skills. In further research, the aspect of pronunciation was also often used as an independent indicator to evaluate the extent to which learners could produce and understand English sounds correctly (comprehensibility).

Therefore, to provide a comprehensive overview of students' speaking abilities before and after using the Elsa Speak application, the assessment in this study included all four components. An analytical assessment rubric was used, with a scoring scale from 1 to 5 (Skehan, 1996). Before the main test, the instrument was tested by English language experts to ensure content and structural validity.

Additionally, to evaluate the clarity and difficulty level of the test, a trial was conducted with a number of students

outside the sample. With this design, the speaking test instrument was expected to provide accurate data regarding the impact of using the Elsa Speak application on students' speaking abilities. This was evaluated through inter-rater reliability methods and the calculation of Cronbach's Alpha.

Table 3. 4 Table Scoring Criteria.

Aspect	Score 5 (Excellent)	Score 4 (Good)	Score 3 (Satisfactory)	Score 2 (Poor)	Score 1 (Very Poor)
Fluency	Speaks smoothly with natural pace, few or no pauses; minimal hesitation	Mostly fluent with minor pauses or self-correction	Noticeable pauses; occasional hesitation affects flow	Frequent pauses and hesitation; difficult to maintain flow	Very limited fluency; constant hesitation; hard to understand
Accuracy	Consistently accurate grammar and vocabulary; rare errors that do not affect meaning	Minor grammar or word choice errors; meaning remains clear	Some noticeable errors that occasionally hinder meaning	Frequent grammar errors; meaning often unclear	Persistent errors; difficult to understand due to lack of accuracy
Pronunciation	Clear pronunciation; easily understood with native-like or intelligible accent	Generally clear; only minor mispronunciations	Some pronunciation issues, but generally understandable	Mispronunciation often causes misunderstanding	Very poor pronunciation; mostly unintelligible
Complexity	Uses varied and complex sentence structures and vocabulary appropriately	Some variety in sentence structures and word choice	Limited range of structures; mostly simple or repetitive vocabulary	Simple sentences with little variation; limited vocabulary	Extremely limited language use; mostly single words or memorized phrases

(Adapted by Skehan, 1996)

E. The Technique of Collecting Data

This research conducted pre-test and post-test for all samples. Pre-test was given to control class and experimental class. Data were obtained by giving tests to students. There were several steps that was taken for data collection, to obtain pre-test and post-test scores. With the aim of knowing the effect of using Elsa Speak Application on speaking ability and the scores obtained from pre-test and post-test.

1. Test

a. Pre-test

A pre-test was conducted before the intervention to measure the initial ability of the students. (Cresswel, 2018) stated that the pre-test was a very important initial step in the experiment to ensure that the groups were homogeneous before being given treatment. It allowed the researcher to discover any potential initial differences that may affect the outcome of the study. This test was an individual speaking test, more precisely a self-introduction, from name, age, hobbies, and family. During the test, the author will recorded it to provide further assessment.

b. Treatment

In this research there were two groups that were given material, namely the control and experimental groups. The experimental group was taught using Elsa Speak App, and the control group was used

conventional methods. Then, the treatment was carried out on the experimental group for 4 times, the author guided students in using the Elsa Speak application, by selecting the introductory English feature and choosing a self-introduction. The author will monitored the process of the treatment.

Table 3. 5 Treatment Plan.

Session	Activities & Focus	Objectives
1. Introduction & Basic Vocabulary	- Introduction to ELSA Speak - Login and placement test - Practice basic vocabulary: name, age, hobby, family	Provide a foundation in using the app and key vocabulary.
2. Sentence Building & Feedback	- Practice complete self-introduction sentences - Analyze ELSA's automated feedback	Improve sentence construction and correct common errors.
3. Fluency & Intonation Practice	- Repetition and shadowing exercises - Focus on speaking fluency and intonation	Enhance fluency and more natural pronunciation.

4. Speaking Test & Reflection	- Conduct post-test (recorded self-introduction) - Group reflection on learning experience	Evaluate progress and promote student self-awareness.
-------------------------------	---	---

c. Post-test

The post-test was conducted after the intervention to measure the extent to which the use of Elsa Speak application affected the improvement of students' speaking ability. Sugiyono said that the purpose of the post-test was to measure the results before and after treatment. As a result, it was used to determine the extent of the influence of Elsa speak after the student received the treatment (Sugiyono, 2020).

Then the author analyzed the students' post-test scores, to find out the significant differences in the achievement of students in the control class and the experimental class. This test was an individual speaking test, more precisely a self-introduction, from name, age, hobbies, and family. During the test, the author will recorded it to provide further assessment.

2. Documentation

Documentation was used to collect additional data, such as recordings of student activities during learning, test results, and attendance records. Documentation, according

to Sugiyono the process of collecting data was through written sources, such as records or reports related to the research subject (Sugiyono, 2020).

F. The Technique of Data Analysis

1. Validity of the Test

Before conducting the research, the author first tested the validity of the instruments. The instruments that were tested showed suitability for the aspects to be tested. According to Cresswel validity refers to the extent to which an instrument measured what it is supposed to measure (Cresswel, 2018). The content and construct validity of the study determined the validity of the speaking test. Content validity ensured that the test covered all aspects of speaking ability, such as pronunciation, fluency, grammar and vocabulary (Field, 2012). As explained by Brown construct validity was assessed by ensuring that the test conformed to the theoretical framework of speaking ability (H. D. Brown, 2007).

To ensure the validity of the study, an English language expert examined the test sections to assess their relevance to the research objectives (Bilodeau, M., & Brenner, 2004). In addition, a small sample of students was tested to determine whether if the instrument could measure their speaking ability (Sugiyono, 2020). Before the final test was

given, the results of the analysis were used to make changes.

Table 3. 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

(Adapted by Hussey et al., 2023)

Furthermore, the author also analyzed the test results using SPSS version 23. Below are the results of the validity test instrument data analysis:

Table 3. 7 Instrument Validity

Items	Sig. (2 Tailed)	Compare sig	r Table	r count	Result
D1	0.000	< 0,05	0.532	0.888	Valid
D2	0.040	< 0,05	0.532	0.518	Invalid
D3	0.183	< 0,05	0.532	0.351	Invalid
D4	0.027	< 0,05	0.532	0.551	Valid
D5	0.063	< 0,05	0.532	0.475	Invalid
D6	0.033	< 0,05	0.532	0.535	Valid
D7	0.004	< 0,05	0.532	0.681	Valid
D8	0.006	< 0,05	0.532	0.650	Valid
D9	0.026	< 0,05	0.532	0.553	Valid
D10	0.021	< 0,05	0.532	0.572	Valid
D11	0.017	< 0,05	0.532	0.588	Valid
D12	0.009	< 0,05	0.532	0.632	Valid
D13	0.006	< 0,05	0.532	0.656	Valid
D14	0.009	< 0,05	0.532	0.632	Valid

D15	0.000	< 0,05	0.532	0.805	Valid
D16	0.027	< 0,05	0.532	0.551	Valid
D17	0.008	< 0,05	0.532	0.637	Valid
D18	0.014	< 0,05	0.532	0.600	Valid
D19	0.017	< 0,05	0.532	0.588	Valid
D20	0.017	< 0,05	0.532	0.588	Valid
D21	0.063	< 0,05	0.532	0.475	Valid
D22	0.036	< 0,05	0.532	0.526	Invalid
D23	0.173	< 0,05	0.532	0.358	Invalid
D24	0.044	< 0,05	0.532	0.509	Invalid
D25	0.103	< 0,05	0.532	0.422	Invalid
D26	0.177	< 0,05	0.532	0.356	Invalid
D27	0.074	< 0,05	0.532	0.459	Invalid
D28	0.111	< 0,05	0.532	0.414	Invalid
D29	0.040	< 0,05	0.532	0.518	Invalid
D30	0.250	< 0,05	0.532	0.306	Invalid

In this validity test, the author used 30 questions with predetermined indicators, and the testing of this instrument has been conducted at SMPN 8 Bengkulu Selatan, which has the same accreditation level as the school that will be used as the research site. The instrument test was carried out from August 12 to 20, 2025, located on Tanjung Besar street, Kedurang, Bengkulu Selatan. After conducting the analysis, there are 18 out of 30 questions that have been tested and are declared valid.

With the following indicators:

Table 3. 8 Indikator of Speaking Validity Test

	No	Test Items	Test Indicators
A.		Basic Introduction	- Student can identify personal information in a short dialogue
	1.	ES : Hi, my name is Elsa A : Hi Elsa, nice to meet you	
B.		What's your dream job?	- Students can recognize and respond to questions about future professions. - Students can express personal aspirations and reasons for career choices.
	2.	ES : How about becoming a pilot ? A : No, being a pilot was only my childhood dream	
	3.	ES : Why entrepreneur A : Because I want to leave a legacy	
C.		My educatio background	- Students can identify educational qualifications (bachelor's, master's, PhD). - Students can recognize questions and answers about college background. - Students can understand future study plans and the concept of online/distance learning.
	4.	ES : Tell me about your education background ? A : I got bachelor's degree in computer science	
	5.	ES : Where did you go to college A : I graduated from standford university	
	6.	ES : That's wonderfull, my sister went there go A : Oh that's right, she's doing her phD in chemistry there right ? ES : Yeah, she's working on her thesis now	
	7.	A : I am also thinking to get an advanced degree next year ES : Go back to school ?	
	8.	A : Probably an online degree as I can do distance learning	

D.	The Collector	- Students can identify hobbies and interests (collecting books, coins, sports). - Students can distinguish between active hobbies and passive activities (playing vs. watching sports). - Students can recognize expressions of preference.
9.	ES : Wow, Justin! You are really into collecting old books A : Yes, I collect rare books	
10.	ES : It's my favorite things to do too. I collect coins A : Well, I am just an amateur. I started just recently	
11.	ES : Do you play any sport ? A : No, I am not particularly sporty. I like watching soccer	
12.	ES : Which team do you support ? A : I support Barcelona's team	
E.	Describe your hometown	- Students can understand descriptions of places (city, capital, economic hub). - Students can identify advantages and disadvantages of a hometown. - Students can recognize expressions related to transportation and pollution.
13.	ES : Tell me your hometown A : I grew up in Ho minh city, the largest city in Vietnam	
14.	ES : Is it the capital of the country A : No, but it's the major economic hub	
15.	ES : What do you like most about your hometown A : The hustling atmosphere and the excellent amenities	
16.	ES : But, I heard traffic congestion is pretty bad there, is that right ? A : Sadly yes, the air quality is terrible because of pollution	
17.	ES : Do you have to commute everyday ? A : Yes, but fortunately I live near a convenient bus route	

F.	Meeting at convention	- Students can recognize
18.	ES : That was a great presentation A : Thank you! Yours was great to	expressions of complimenting and responding politely.

2. Reliability of the Study

Reliability was the consistency and stability of test results over time (H. D. Brown, 2007). The reliable test ensured that the results reflected students' actual speaking ability, that was not random errors (Field, 2012). The reliability of the study was measured using the test-retest method and inter-rater reliability.

According to (Weir, 2007), the reliability of language assessment considered both aspects of test process reliability and score reliability. (Bachman, L., & Adrian, 2010) emphasized that reliability included consistency of results and reliability of test procedures and instruments. Retesting entailed the same group of students being given the same test twice. If there was a high correlation between the two sets of scores (above 0.70), then the test was considered reliable (Sugiyono, 2020). Inter-rater reliability was tested by asking two independent raters to score the students' speaking ability using a standardized rubric. If the scores of the two raters are comparable, then high reliability were confirmed (Pallant, 2005).

Cronbach's Alpha coefficient was calculated to further ensure reliability. Reliability with a coefficient above 0.70 was generally considered acceptable in educational research (Cresswel, 2018). High reliability increased the reliability of the research and the reproducibility of the results. The reliability value is measured using the Cronbach's Alpha value; if the Cronbach's Alpha value > r-table, then the variable can be declared valid.

Table 3. 9 Test Instrument Reliability Criteria

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

(Adapted by Iqbal &

Salsabila, 2023)

The author analyzes the reliability value using SPSS version 23. This test is conducted to determine whether the research instrument provides consistent results. The method used in the calculation is Cronbach's Alpha, which measures the extent to which the questions in the instrument are related to each other. After performing the calculation, the reliability coefficient value (Cronbach's Alpha) is as follows:

Table 3. 10 Instrument Reliability

Reliability Statistic	
Cronbach's Alpha	N of Items
0,919	30
Reliable	

In this reliability test, the author used 30 questions with predetermined indicators. Based on the reliability criteria instrument test table, if the Cronbach's Alpha value reaches 0.80-1.00, it is considered highly reliable. Therefore, the statistical calculation result in this study is 0.919, which means that the reliability level is Highly Reliable, so the tested data is declared reliable and can be used as an instrument for further research.

3. Item Difficulty Test

The analysis of item difficulty testing was used to examine the difficulty level of each test. This analysis determined whether the items on the test had an appropriate level of difficulty for students. To ensure fair assessment, the test must balance easy, moderate, and difficult aspects (H. D. Brown, 2007). Furthermore, the item difficulty test was an important aspect for the development of research instruments, including Elsa Speak, which aimed to improve speaking ability. The difficulty level of test items also refers to the difficulty of questions in an evaluation instrument. Several steps are taken to test the difficulty level: 1)

Arrange the answer sheets from the highest to the lowest scores, 2) Take the top 27% and bottom 27% of the score sheets, and set aside the middle 46%, 3) Create a distribution of answers for each group, and then calculate P based on the data in that table. This test was conducted to determine how difficult an item was for the target population being measured (Haladyna, 2004).

In the context of this research, this test aimed to ensure that the items or exercises presented had a level of difficulty that matched the students' abilities. The difficulty test process was conducted on junior high school students who had the same accreditation. After conducting the difficulty test, the author collected the results for further research. This test was usually calculated with the formula:

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

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

$$P = \frac{357+100}{4+4} \times 100\%$$

$$P = \frac{457}{8} \times 100\%$$

$$P = 57,1 \%$$

Table 3. 11 Interpretation of Difficult Question Category.

Question categories	Interpretation of result
Easy	71%-----100%
Medium	41%-----70%
Hard	21%-----40%

(Adapted by Haladyna, 2004)

In developing this research, it is important to balance the difficulty level of the problems or exercises provided. This ensures that the material presented is not too easy, which can reduce learning motivation, or too difficult, which can make students frustrated. Based on the difficulty level calculation, the tested items reached 57.1%. According to the criteria in the table, this calculated result can be stated as a moderate level, because the range 41%–70% is categorized as a moderate difficulty level. Therefore, the items are suitable for use in the research.

4. Defferential power test

The discriminating power index was calculated to measure how effective each item was in distinguishing between students who had good and less good abilities. Items with an index of discriminating power above 0.40

were considered effective in differentiating students' ability levels (Bilodeau, M., & Brenner, 2004). Items that had a low discriminating power index were changed or removed to improve the efficiency of the test. This process ensured that the test provided an accurate picture of students' speaking ability.

The differential power test was an important method in the development of evaluation instruments, including in this research designed to increasing students' speaking ability. This test aimed to determine how well an item could differentiate the skills possessed by students. According to Grounlund, differential power referred to an item's ability to compare the abilities of students. Items with high differential power could explain the differences between students who had understood the material well and those who had not mastered it (Gronlund, 1985).

This process created groups among students; the upper group consisted of a number of students with the highest scores on the overall tests in this research, while the lower group consists of students with the lowest scores on the overall tests in this research. The calculation of differential power was usually calculated by subtracting the proportion of lower group students who answered the item from the proportion of upper group students who answered correctly. The sample used for differential analysis was typically taken from the top 27% of the overall 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.

$$DP \frac{357-100}{8}$$

$$DP = 32,1$$

Table 3. 12 Interpretation of Defferential Power Test

Defferential power test	Interpretation
$DP \geq + 0,45$	Good
$+0,21 < DP < 0,40$	Satisfactory/Fairly acceptable
$0 < DP \leq 0,20$	Weak

(Adapted by Gronlund, 1985)

This test was conducted to ensure that the test given was able to effectively improve students' speaking skills. In this study, the test results reached 32.1%. Looking at the table, this result can be considered good, because the range is >

+ 0.45, therefore the Discrimination Test in this study meets the specified requirements.

5. Normality Test

The normality test is one of the statistical methods to determine whether the data is normally distributed. In parametric statistical analysis, normal distribution is a very important assumption. Ghasemi et al explain that the normality test is necessary to avoid bias in statistical estimation and to allow research results to be generalized (Ghasemi & Zahediasl, 2020). The normality test is often conducted to assess the normality of the data. The author uses SPSS to test and establish normality. There are criteria for testing in the normality test, namely:

- a. The significance value (P value) < 0.05 concludes that the data is not normally distributed.
- b. The significance value (P value) > 0.05 concludes that the data is normally distributed.

6. Homogeneity Test

Homogeneity of variances is a primary requirement for t-tests and analysis of variance (ANOVA). The homogeneity test is used to ensure that the variances among groups of data are the same. This is important for valid statistical analysis. Wang & Ma emphasize that the homogeneity test is necessary to ensure that the analysis results are not altered by differences in variances among

groups. There are several principles for determining homogeneity:

- a. Significance level (α) = 0.05
- b. If $\text{sig} > \alpha$, then the variances of each sample are the same (homogeneity).
- c. If $\text{sig} < \alpha$, then the variances of each sample are not the same (not homogeneous).

The author uses SPSS to determine this.

7. T-test

The T-test aims to compare the means of two groups and determine their statistical significance. Vankelecom et al state that, assuming normality of the data and homogeneity of variances, the T-test is very useful for evaluating differences between groups (Vankelecom et al., 2024). In this analysis, the author uses SPSS; this test aims to determine how significant the partial effect of independent factors on the dependent variable is when other independent variables are held constant. The principles used in this test are:

- a. The independent variable uniquely affects the dependent variable 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 alpha (0.05).

8. F-test

According to Robert Odek & Gordon Opuodho, the F Test serves as a basis for assessing the significance of group differences and the feasibility of statistical models (Robert Odek & Gordon Opuodho, 2023). This test aims to examine group similarities and significance in regression analysis and ANCOVA. It can reduce the influence of external variables that are not the primary focus of the research, making the comparative results between groups more accurate and objective. This technique is highly supportive in experimental research, especially when trying to determine the effects of a particular treatment or method.

9. The Effect Size

This test measures the magnitude of the influence or difference observed in the research. This is very important because it reveals the results regarding how large or strong the effect is on the research. Nordahl-Hansen et al emphasize that the interpretation of effect size is more important than the p-value because it shows the strength of the relationship between variables(Nordahl-Hansen et al., 2024). Effect size is commonly used to measure the average difference between two groups and Pearson's r , which measures how strong the relationship is between two variables. Here are the categories of effect size values that are often used to see how significant the influence is between two variables in research:

Table 3. 13 Value of The Effect Size

No	Value	Effect Size
1.	0.01-0.19	Very small
2.	0.20-0.49	Small
3.	0.50-0.79	Medium
4.	0.80-1.19	Large
5.	1.20-1.99	Very large
6.	≥ 2.00	Huge effect

Adapted by (Cohen, 2013)

10. Statistical Hypothesis

This test function in testing H_0 is accepted or rejected, and the alternative test H_a is accepted or rejected. This test can also be conducted using SPSS, with the following criteria:

- a. If the 2-tailed sig value < 0.05 , then the hypothesis is accepted.
- b. If the 2-tailed sig value > 0.05 , then the hypothesis is rejected.