

CHAPTER III

RESEARCH METHOD

A. Research Setting

1. Place Of The Research

This research was conducted at MTs Roudlotur Rosmani located in Padang Serai Village, Kampung Melayu District, Bengkulu City. The reason the researcher conducted the research at MTs Roudlotur Rosmani was because the researcher during the Internship Program of approximately three months at the place found obstacles in grade 8 students of MTs Roudlotur Rosmani in the form of a lack of understanding of the Reading Skills mastered so that it affected the English learning process in the classroom.

therefore an effort was needed to improve reading skills in the English subject. In addition, the researcher at the same time will be practiced a game-based learning method, namely a team game tournament (TGT) which was considered more effective than the traditional learning method (lecture).

2. Research Of Time

This research was conducted in the second semester of 2026. During this period, researchers collected data,

information, and documentation related to the title of this research.

B. Type of Research

This quantitative study uses an experimental design. Quasi-experimental research, according to Rindrayani (2025), is a way to examine the causal relationship between variables without randomizing the subjects as a whole.(M. Pembelajaran et al., 2022). According to Abraham (2022), in English, experiment means an experiment on an object. An experiment can be interpreted as the treatment of an activity through a certain method to produce or confirm something. Meanwhile, experimental research is research that seeks to find the cause and effect relationship obtained from independent variables with dependent variables.(M. Pembelajaran et al., 2022)

Researchers conclude that independent variables are systematically designed to be controlled, manipulated, or selected to influence other variables. Meanwhile, dependent variables are variables that are deliberately measured, observed, and ultimately changed in response to independent variables. According to Abraham (2022), experimental research is divided into three major groups, namely pre-experimental, true experimental, quasi-experimental, and action research.

Each of the four major types of research above has its own advantages and disadvantages. In addition, the four have different levels of confidence, adjusting to different beliefs in accordance with the experimental research paradigm commonly referred to as positivism.(M. Pembelajaran et al., 2022).

According to Amaliya (2024) in (Salim & Haidir, 2019), quasi-experimental research aims to determine potential causes and effects. By using experimental methods, this research can determine how experiments or treatments affect the characteristics of the subjects being studied.(Amaliya & Anas, 2024). Based on the above theory, the researcher concludes that quasi-experimental research (pretest-posttest control group) is research involving two groups: a control group and an experimental group.

The control group receives a specific treatment (method, media, application, etc.), while the experimental group does not receive any treatment. After that, a pretest and posttest are conducted to measure whether there is a difference in learning outcomes before and after the treatment. For details, see the table below.

Tabel 3.1 Pre-Test Post-Test Control Group Design

Class	Pretest	Treatment	Posttest
Experiment	O1	TGT + monopoly games	O2
Control	O3	-	O4

Source: (Elpita, 2026)

The researchers chose a quasi-experimental study with a control group to determine whether the TGT cooperative learning method had an impact on students' reading skills. They also wanted to find out whether there was a difference between the skills of students in the experimental class using the experimental method and those in the control class using the conventional method.

In this design, the researchers used pretest and posttest instruments; the pretest was conducted before the special treatment was given and the posttest was conducted after the treatment was given. This was done to determine the impact.

C. Population

According to Isra (in Hutami, 2024), population has a meaning that includes humans or animals that are a separate group that is considered to have different characteristics from other groups. This serves as a guideline in understanding the population that has a broad scope in the form of the total

number of individuals and more specifically, it is a collection of group units that are analyzed through specific characteristics, such as the nature of the individual itself or the conditions that are currently occurring in the individual.(Makassar et al., 2026). The population in this study is all eighth grade students at MTs Roudlotur Rosmani in Bengkulu City.

Table 3.2 The Population of the Research

NO	CLASS	TOTAL NUMBER
1.	Grade 7 A	5
2.	Grade 7 B	6
3.	Grade 8 A	9
4.	Grade 8 B	9
5.	Grade 9 A	7
6.	Grade 9 B	6
Total		42

Source: MTs Roudlotur Rosmani Kota Bengkulu, 2026

Based on the table above, the population of this study is trith grade students at MTs Roudlotur Rosmani Kota Bengkulu.

D. Samples

A sample is a member or part of a larger unit (population). According to Hazara (in Musthofa, 2024), a sample is defined as a subset of a population or the largest unit. Population sampling is based on a systematic approach that encompasses specific characteristics within the sample or

the phenomena occurring within the sample so that they can be used for scientific research purposes.(Nazara, 2026). The sample in this study consisted of the eighth-grade students at MTs Roudlotur Rosmani in Bengkulu City.

Table 3.3 The Samples of the Research

No	CLASS	TOTAL NUMBER
1	Grade 8 A	9
2	Grade 8 B	9

Source: MTs Roudlotur Rosmani Kota Bengkulu, 2026

Based on the table above the samples of this study is grade eighth A and eighth B. All eighth-grade students were used as the sample because there was only one eighth-grade class at the school, so the entire sample was taken from the entire population.

E. Variable and indicator of the research

According to Variable is a characteristic, property, or feature that is variable and becomes the focus of measurement in research. Variables can be independent variables (which are thought to influence) and dependent variables (which are influenced or measured). Variables are objects of research observation, often referred to as factors that play a role in research or phenomena to be studied. There are many types of research variables, but in this study, the following variables were used:

1. Independent Variables (x)

Independent variables (x) are variables that influence or cause changes or the emergence of dependent variables. Therefore, the independent variable (x) in this study is TGT cooperative learning using media, namely educational games that help teachers relate teaching materials to real-world situations. The indicators used by the researcher in learning are : Students participate in using game media, student dare to express their opinions together with the group, student read the text through challenges in game media, and students can learn actively by using game media.

2. Dependent Variable (y)

The dependent variable (y) is a variable that is influenced or affected by the independent variable. Therefore, the dependent variable (y) in this study is student learning activities after conducting experiments on 8th grade students at MTs Roudlotur Rosmani in Bengkulu City. The indicators of students' English reading ability in this study are: students work together with their peers to solve challenges, students are able to read prepared texts, students dare to ask/answer questions, students understand the meaning of reading text and how to read them, students acquire new vocabulary, students can summarize the main ideas from the reading texts, and Students dare to

summarize their learning outcomes.

F. Data Collection technique

According to the Thesis Writing Guide – UMN (2021). Data collection techniques are scientific methods used by researchers to obtain the information needed to answer research questions. These can include observation, questionnaires/surveys, interviews, documentation, tests, etc. In this study, data was collected through. In this study, the researcher used pre-tests and post-tests in the experimental class and control class because both tests were important for determining the results of the study.

In this controlled class experimental study, two eighth-grade classes were used: the experimental class (Class 8A) and the control class (Class 8B). The data collection process was carried out in several stages, namely pre-tests, treatment, and post-tests. The purpose of this process was to determine whether TGT using Monopoly game media could improve students' English reading skills at MTs Roudlotur Rosmani in Bengkulu City. The pre-test and post-test scores were then compared to see the difference in learning outcomes and to determine the students' speaking skills before the treatment was given. The following are the three data collection techniques:

1. Test Technique

According to Rozikin (2026), the instrument is the most crucial stage in the data collection process because it is the tool. Instruments are an integral part of the mandatory process prepared by researchers before data collection. A valid instrument goes through various stages, including creation, testing by a validator, and evaluation, so that the instrument can be deemed valid and suitable for use in research. (Kualitatif, 2026)

Magdalena (2021) states that test instruments consist of objective tests and subjective tests. Objective tests are tests that assess individuals using scores, such as multiple choice, true or false, etc. Subjective tests are descriptive tests that assess individuals using a scale. An example of an objective test is an essay. In this study, the researcher used multiple choice by pre-tests and post-tests. Then, the treatment, namely TGT (Time Game Tournament), was applied at the stage before the post-test in the control class and was not applied to the experimental class. (Penggunaan & Dan, 2021)

In the final stage, namely the post-test, where the number of correct student answers will be calculated to see whether there is an increase, each awarding 1 point for correct answers and 0 points for incorrect answers.

a. Pre-test

According to Ayuningtyas (2026), the pre-test activity is the initial series of testing tasks. This pre-test serves to provide a more realistic perspective as an initial test that can be used to measure students' initial abilities.(J. Multidisiplin et al., 2026). The test material must be relevant to the material to be taught. In this pre-test phase, the researcher administered 30 multiple-choice questions to eighth-grade students in the experimental and control groups.

b. Post-Test

Final tests, also known as post-tests, are conducted at the end of the learning process for a particular subject. Magdalena (2021) states that these tests are conducted to assess the extent to which students understand the material and important topics they have learned. The test material is combined with the material that students have learned previously. (Penggunaan & Dan, 2021)

If students understand the material better after the learning process, the learning program is considered successful. The goal is for teachers to be able to determine which students have better test results. The teaching program is considered successful if students understand the material better after the learning process.

In this post-test phase, the researcher also administered 30 multiple-choice questions to eighth-grade students in the control and experimental groups.

c. Treatment

Treatment is a systematic procedure or intervention deliberately administered to research subjects (experimental group) to determine its causal effect or impact on the dependent variable. The treatment used in this study was the application of TGT using a Monopoly game that had been designed to improve and stimulate students to read English texts. In this treatment phase, the researcher administered a treatment in the form of the TGT method, supplemented with the Monopoly game as a teaching aid; this treatment was conducted only in the control class, namely B class.

2. Documentation.

According to the official Oxford Learner's Dictionaries website, Hajar Hasan's (2022) research states that there are two definitions of documentation. First, it refers to the presentation of information or official evidence that is useful for records. Second, it refers to the process of recording and categorizing information in the form of text, photos, videos, and so on. Documentation is

a systematic activity or process involving the search, use, investigation, collection, and provision of documents to obtain knowledge, information, and evidence, as well as to disseminate them to interested parties.(Hasan et al., 2022)

Documentation is the final stage of this study. It is a crucial stage because it serves as evidence that the researcher has conducted the study and also as evidence of the study's results. The researcher documented the entire sequence of stages, from the pre-test and post-test to the intervention. The documentation includes the students' behavior in the classroom, their performance on multiple-choice questions during the pre-test and post-test, and the implementation of the intervention.

G. Instrument test

Louis Guttman, an American sociologist and statistician, first introduced the Guttman Scale, which is characterized by two answer choices. Faizah, Wahyudin, and Juniarti, 2024, state that in this scale, answers are given in a dichotomous form, such as "yes-no," "ever-never," or similar categories (Fatimah & Puspaningtyas, 2020). Two score categories are used to assess answers: a score of one indicates an answer that is considered correct or shows a positive response, and a careless score indicates an answer that is

considered incorrect or shows a negative response.

According to Lumuko et al. (2018), the measuring instrument is made based on the material available with TGT and adjusted to the students' ability level to support the achievement of learning objectives. The main idea, stated details, unstated details (inference), reading in context, and reference words (he, she, it, they, this, that, etc.) are the research tools used in this study. An English text reading test was used. The purpose of this test was to indirectly measure students' reading abilities, particularly their ability to understand and fluently read texts used in everyday conversation. The test consisted of two stages: a pretest and a posttest. Based on the Guttman scale, a dichotomous system was used to assess students' answers.

H. Try Out Test Results

According to Apriza (2025), a tryout is a series of evaluations within an assessment that serves to provide a concrete picture of the question format and exam implementation. The results of the tryout provide guidance on the quality and composition of the questions we've created to determine whether they're adequate. (Apriza et al., 2025)

In this Try Out test, researchers used 30 multiple-choice questions from several sentences with predetermined indicators. The questionnaire trial was conducted at MTs Nur

Rahma in Bengkulu City, which has the same accreditation level as the school that would be the main research location, namely MTs Roudlotur Rosmani in Bengkulu City. This trial involved 18 students as test subjects. After the trial test was completed, the researchers analyzed the data by calculating the validity of each answer.

From the analysis results, it was found that 25 of the 30 multiple choice questions. However, before conducting the practice test, the researcher first tested the instrument to ensure it was suitable for use with the target population in the upcoming study at MTs Roudlotur Rosmani. The validation test was administered to three expert validators, after which the validity was calculated by excel and revisions were made to the instrument.

I. Validity

1. Validity from Validator Judgment

The first test is the validity test. According to Jannah (2026), a validity test is a test conducted by researchers to determine whether a measurement tool, or instrument, can be deemed valid or invalid. The measurement tool in question consists of a series of questions posed to respondents or multiple-choice questions. Validity is used to ensure that measuring instruments (such as tests or questionnaires) actually measure what is intended to be

studied, not something else.(Jannah et al., 2026)

Based on the definition above, researchers conclude that a validity test is a test conducted by researchers to determine whether a measurement tool or instrument is valid or not. There are two types of validity tests. The first is a test to correlate each test item commonly referred to as an item with the total number of items. The second is to correlate each score on an item indicator with the total score. In this study, because multiple-choice instruments were used, the second type of validity test was employed, which involves correlating each score with the total score Using Aiken's V Formula.

Before the research instrument was administered in the try-out, the researcher first conducted an expert judgment to ensure the content validity of the test items. The validation process involved three lecturers from the English Education Study Program at UIN Fatmawati Sukarno Bengkulu who acted as expert validators. They evaluated the instrument based on several aspects, including the relevance of the test items to the research objectives, the suitability of the material with the learning objectives, the clarity of instructions, the appropriateness of language use, and the construction of the multiple-choice questions.

Based on the suggestions and feedback provided by

the validators, the researcher revised several test items to improve their quality and ensure that they appropriately measured students' reading skills. After the revision process had been completed, the validated instrument was administered in the try-out to examine its validity, reliability, item difficulty level, and discrimination index before being used as the research instrument in the pre-test and post-test.

Table.3.4 Validator Data Score Result

Validator			Interpretation
I	II	III	
S1	S2	S3	
114,1	102,5	317	HIGH
Σs	0,880556		

(See Appendixs 6 ,P.CXXVI)

To determine the significance of the correlation, a test is conducted by comparing the calculated r with the table r for the product-moment correlation at a 5% significance level. If the calculated r is greater than the table r for the product-moment correlation at the 5% significance level, then the instrument is considered valid.

The instrument validity test was conducted by having three experts in the field of English evaluate the instrument; the validators were as follows:

2. Validity from Try Out

After the content validity test, the researcher also conducted a field validity test by administering a pilot test to 8th-grade students at MTs Nur Rahma in Bengkulu City.

Table 3.5 Frequency Distribution of Try Out Test

No	Interval	Try Out Test		
		Frequency	Percentage	Interpretation
1.	3-7	5	27,8%	Verry Low
2.	8-12	4	22,2%	Low
3.	13-17	2	11,1%	Medium
4.	18-22	3	16,7%	High
5.	23-27	4	22,2%	Very high
Total		18	100%	

(See Appendixs 7,P.CXXXV)

The table above shows that the 18 students from MTs Nur Rahma, Bengkulu City, are distributed across five categories based on the score intervals obtained during the trial exam. Five students fell into the 3–7 interval, placing them in the "very low" category, while four students achieved scores in the highest interval (23–27). Overall, the data is considered valid and normally distributed, as nine students—a larger group than those

with low scores—achieved results ranging from the medium to the highest levels.

Figure 2 Try out test



Source : Try out test At MTs Nur Rahma, 2026

After completing the practice test, the data collected will be analyzed using validity tests, reliability tests, difficulty level tests, discriminative power tests, and item

$$r_{xy} = \frac{N \sum XY - (\sum X)(\sum Y)}{\sqrt{\{N \sum X^2 - (\sum X)^2\} \{N \sum Y^2 - (\sum Y)^2\}}}$$

Note:

r_{xy} : Correlation Coefficient Between Variable X and Variable Y

N : Number Of test takers

X : Score of each item

Y : total score

$\sum XY$: sum of XY multiplecations

$(\sum X)^2$: Sum of the Squared item scores

$(\sum Y)^2$: total test score, square

Source: (Elpita sari, 2025)

To determine the significance of the correlation, a test is conducted by comparing the calculated r with the table r for the product-moment correlation at a 5% significance level. If the calculated r is greater than the table r for the product-moment correlation at the 5% significance level, then the instrument is considered valid.

This pilot test was analyzed using SPSS 16.0. Validity test criteria:

- If the calculated r value > than the table r value, it is deemed valid
- If the calculated r value < than the table r value, it is deemed invalid

Tabel 3.6 Result Validity Test

No item	R count	R Tabel	Interpretation
1	0,527	0,468	Valid
2	0,716	0,468	Valid
3	0,691	0,468	Valid
4	0,189	0,468	Invalid
5	0,530	0,468	Valid
6	0,473	0,468	Valid

7	0,608	0,468	Valid
8	0,570	0,468	Valid
9	0,352	0,468	Invalid
10	0,530	0,468	Valid
11	0,541	0,468	Valid
12	0,691	0,468	Valid
13	0,544	0,468	Valid
14	0,585	0,468	Valid
15	0,285	0,468	Invalid
16	0,284	0,468	Invalid
17	0,569	0,468	Valid
18	0,540	0,468	Valid
19	0,530	0,468	Valid
20	0,127	0,468	Invalid
21	0,662	0,468	Valid
22	0,626	0,468	Valid
23	0,624	0,468	Valid
24	0,610	0,468	Valid
25	0,626	0,468	Valid
26	0,709	0,468	Valid
27	0,864	0,468	Valid
28	0,805	0,468	Valid
29	0,681	0,468	Valid
30	0,756	0,468	Valid

(See Appendixs 8 ,P.CXXXVI)

From the output table above, it can be seen that 25 test items have a calculated r greater than the table r , meaning those 25 items are considered valid. Meanwhile, 5 test items have a calculated r smaller than the table r , meaning those 5 items are invalid. Since the 5 questions are invalid, they are unsuitable and cannot be used in the study. Based on this, only 25 test items are suitable for use in measuring students' English reading ability in this study.

3. Reliability

According to Nurhalimah (2025), a reliability test is a design carried out after a validity test. A reliability test focuses on how to measure a measuring instrument consistently. Consistency in reliability is assessed when the test instrument has the same results repeatedly when tested on different places or objects. (Kualitatif, 2026).

Reliability is a tool for assessing the extent to which a questionnaire is able to produce consistent answers over time. An instrument is said to be reliable if the responses given by a person to the statement items remain stable when measurements are taken repeatedly under the same conditions. In other words, reliability indicates the level of stability, consistency, predictive ability, and accuracy of a measuring instrument. The

higher the reliability value, the greater the confidence that the instrument can produce accurate and reliable data in describing the construct being measured.

Therefore, measurements with high reliability are an important requirement in research so that the data obtained truly reflects the actual situation. According to Nur Halimah (2026), a reliability test is a measure of the consistency of results when repeated measurements are taken of a particular characteristic. A measurement is considered reliable if the results of an evaluation tool are the same or relatively the same. Reliability testing for a test using multiple-choice questions employs Cronbach's Alpha, with the formula for calculating item reliability as follows:

$$r_i = \frac{k}{(k-1)} \left\{ 1 - \frac{\sum S_i^2}{S_t^2} \right\}$$

Note:

r_i : Test reliability coefficient

k : Number of items

S_i^2 : Variance of item scores

S_t^2 : Variance of total scores

Source : (Elpita Sari, 2025)

If the calculated r is ≥ 0.7 , the reliability is high; if the calculated r is < 0.7 , the instrument used does not yet have high reliability. The interpretation of the correlation coefficient values is as shown in the following table:

Tabel 3.7 Interpretation coefficient correlation

Koefisein Korelasi ®	Interpretasi
$0.80 < r \leq 1.00$	Very High
$0.60 < r \leq 0.80$	High
$0.40 < r \leq 0.60$	Moderete
$0.20 < r \leq 0.40$	Low
$r \leq 0.20$	Very low

A reliability test is a measurement tool where an instrument is considered reliable if it yields consistent results when measured repeatedly under the same conditions.

Reliability test criteria:

- If the Cronbach's Alpha value > 0.60 , the instrument is reliable
- If the Cronbach's Alpha value < 0.60 , the instrument is not reliable.

The reliability test in this study was conducted using SPSS 16.0, resulting in the output shown in the following table:

Table.3.8 Test Reliability Criteria

No	Cronbach's Alpa	Reliability Level
1.	0,00 – 0,20	Lack of Reliabel
2.	0,20 – 0,40	Some what Reliable
3.	0,40 – 0,60	Fairly Reliabel
4.	0,60 – 0,80	Reliable
5.	0,80 – 1,00	Highly Reliable

Source: (Sugiyono, 2017) (Dan, Angket, & Emosi, 2023)

Tabel.3.9 Result reliability test

Reliability Statistics	
Cronbach's	
Alpha	N of Items
.936	25

(See Appendixs 9 ,PCXXXVII.)

From the table above, the Cronbach's Alpha value is $0.936 > 0.60$, meaning the instrument to be used is reliable with a high level of reliability.

4. The level difficulty test

Analyzing the difficulty level of each exercise or question is an important step in developing technology-based research instruments. Difficulty level analysis is the process of assessing test items based on their level of difficulty so that it can be determined which questions are easy, medium, or difficult.(Penggunaan & Dan, 2021). In

this study, the Cake application was used as a learning medium to improve students' speaking skills.

Difficulty level analysis was conducted to ensure that the speaking exercises provided were in line with the students' abilities, so that the material was neither too easy nor too difficult. Data was collected through speaking exercises given using the Cake application, and then the students' results were analyzed. The number of students who were able to complete the exercises correctly compared to the total number of participants was used to determine the difficulty level of an item. The most commonly used formula is

$$P = \frac{R_h + R_i}{N_h + N_i} \times 100\%$$

Description:

P = Difficulty level in percent

N_h = Number of test takers in high score group

R_h = Number of correct answers in high score group

N_i = Number of test takers in low score group

R_i = Number of correct answers in low score group

$$P = \frac{145 + 65}{5 + 5} \times 100\%$$

$$P = \frac{210}{5} \times 100\%$$

$$P = 42\%$$

Table.3.10 Level Of Difficulty scale

Level of Difficulty scale
0.00–0.30 = difficult
0.31–0.70 = moderate
0.71–1. 00 = easy

At this difficulty level, there are many questions classified as moderate, and only two classified as easy. The data is summarized in the table below.

Tabel 3.11 level difficulty Result

No item	Valid	Missing	Mean	Interpretation
1	18	0	0,56	Moderate
2	18	0	0,56	Moderate
3	18	0	0,50	Moderate
4	18	0	0,50	Moderate
5	18	0	0,50	Moderate
6	18	0	0,56	Moderate
7	18	0	0,44	Moderate
8	18	0	0,50	Moderate
9	18	0	0,72	Easy
10	18	0	0,50	Moderate
11	18	0	0,67	Moderate
12	18	0	0,50	Moderate

13	18	0	0,61	Moderate
14	18	0	0,61	Moderate
15	18	0	0,67	Moderate
16	18	0	0,68	Moderate
17	18	0	0,67	Moderate
18	18	0	0,56	Moderate
19	18	0	0,50	Moderate
20	18	0	0,72	Easy
21	18	0	0,56	Moderate
22	18	0	0,61	Moderate
23	18	0	0,50	Moderate
24	18	0	0,50	Moderate
25	18	0	0,61	Moderate
26	18	0	0,61	Moderate
27	18	0	0,44	Moderate
28	18	0	0,39	Moderate
29	18	0	0,39	Moderate
30	18	0	0,39	Moderate

(See Appendixs10 ,P.CXXXVIII)

5. Distinguishing power

According to Novalia and Muhammad Syazali (2024), discriminant analysis is the process of assessing test items based on their ability to distinguish between students with low and high abilities. The determination of discriminant power in the test item analysis program can

be seen in the discriminant power table in the DP (%) column. Items with a discrimination index ≥ 0.30 are categorized as good, while items with an index below 0.30 are considered poor. (Putri et al., 2025)

Discrimination analysis is useful for improving the quality of each item based on empirical data and for determining the extent to which an item can distinguish between students who have understood the material and those who have not.

6. Table 3.12 Distinguishing power categories

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

Source: (Nurhalimah, Hidayati, Rosidi, Madura, & Soal, 2022)

Testing research questions requires a balance in the level of difficulty of test items, so that no item is too easy or too difficult. Based on the calculation of the difficulty level, a value of 42% was obtained. According to the applicable criteria, this value falls into the medium category, because the range of 30% to 70% is classified as a medium level of difficulty. Thus, these questions are appropriate and suitable for use in research. The most commonly used formula is:

$$DP = \frac{U-L}{N}$$

N

Description:

DP = Question discriminating power.

U = Number of correct answers from high score group test takers
 L = Number of correct answers from low score group test takers
 N = Number of students in each group

$$DP = \frac{29-13}{42}$$

42

$$DP = 38,1\%$$

7. Table 3.13 Interpretation of the Distinguishing power test

T-test	Interpretation of results
$0,70 \leq DP \leq 1,00$	Very good
$0,40 \leq DP \leq 0,69$	Good
If $+0,20 < DP < 0,39$	Sufficient
If $0 < DP \leq 0,19$	Weak

Source: (Penggunaan & Dan, 2021)

Based on the criteria in the discriminant power table above, the results of this practice test show that there are 3 items in the “very good” category with a score range above 0.728, while the “good” category dominates. There

are also 3 items in the “sufficient” category with a score range between 0.227 and 0.303. Finally, there were two test items in the "weak" category with scores below 0.130. the data can be viewed in the table below :

8. Table.3.14 Distinguishing power Result

No item	Combrach Item total correlation	Interpretation
1	0,481	Good
2	0,684	Good
3	0,657	Good
4	0,130	Waek
5	0,484	Good
6	0,424	Good
7	0,568	Good
8	0,527	Good
9	0,303	Sufficient
10	0,484	Good
11	0,499	Good
12	0,657	Good
13	0,500	Good
14	0,544	Good
15	0,231	Sufficient
16	0,227	Sufficient
17	0,529	Good
18	0,495	Good

19	0,484	Good
20	0,074	Weak
21	0,626	Good
22	0,588	Good
23	0,585	Good
24	0,570	Good
25	0,588	Good
26	0,677	Good
27	0,848	Very good
28	0,783	Very good
29	0,648	Good
30	0,728	Very good

(See Appendixs 11 ,P.CXXXIX)

9. Effectiveness of the Power deception

When analyzing multiple-choice test items, the use of distractors is also considered. According to Kaune (2025), distractors commonly referred to as “distractors” are a method for measuring how effective a distractor is within an assessment instrument such as a multiple-choice test. Distractors are used not only to provide varied answer choices but also to trap students who do not fully understand the test material. Distractors are designed to resemble the correct answer as closely as possible.(Kaune & Majid, 2025)

Based on the above definition, the researcher concludes that distractors are an effort to trap students so that they provide more varied answers. A distractor can be considered effective if at least some students are misled by it. In interpreting the effectiveness of distractors for each test item, the following Likert scale criteria are used:

- The effectiveness of distractors is considered very good if all three distractors function effectively.
- The effectiveness of distractors is considered good if two distractors function effectively.
- The effectiveness of distractors is considered less than good if one distractor functions effectively.
- The effectiveness of distractors is considered poor if none of the distractors function effectively.

Tabel 3.15 power Deception Result

No	Effectiveness of distractors	No item	Total	percentage
1	Very High	1,3,6,7,8,10, 12,13,14,18, 19,21,22,23, 24,25,27,28, 29,30	20	66,6%
2	High	2,5,9,11,17, 16,20,26	8	26,6%

3	Weak	4 and 15	2	0,6%
4	Not High	0	0	0%

(See Appendixs12 ,P.CXLI)

Based on the table above, items in the “very high” and “high” categories can be used in the study, while items in the “weak” and “not high” categories cannot be used.

J. Test Of Basic Assumptions

1. Normality test

Normality test according to Ananda (2025) the normality test can be carried out to find out whether the data obtained by the researcher is normally distributed or not (Kamtini & Apriani, 2023). This test is very important to ensure that the research tool works properly and that the research findings are acceptable. Researchers usually use programs such as SPSS to conduct normality tests. In addition, when assessing whether the data is normal.

There are several criteria that must be considered in this test:

- Sig Value (P Value) < 0,05 concluded that the data was not normally distributed.
- Sig Value (P Value) > 0,05 conclude that the data is normally distributed.
-

2. Homogeneity test

The homogeneity test is used to determine whether some population variants are equal. The basic assumption of analysis of variance (Anova) is that As a condition for both independent sample t test and Anova analysis, this test is performed. The population variances are equal. The test of equality of two variances checks whether the distribution of data is homogeneous by comparing the two variances. If two or more groups of data have the same level of variance, then the homogeneity test is no longer necessary (Materi et al., 2026). Use the following principles to determine homogeneity:

- Significance test (α) = 0.05
- If $\text{sig} > \alpha$, then the variance of each sample is the same (homogeneity).
- If $\text{sig} < \alpha$, then the variance of each sample is the not same (not homogeneity) to determine it the researcher uses SPSS.

K. Test Of Hypothesis

1. T-test

T-test is a statistical test used to test the truth or falsity of the null hypothesis. To check for significant differences between two groups with small samples (n is usually less than 30) and non-normally distributed data,

the t-test uses the t-statistic value resulting from the difference between the means of the two groups and the difference between the sample size and variability of each group (Syafriani et al., 2023) and the formula to be used is as follows.

- The independent variable affects the dependent variable uniquely if the significance level is less than 0.05.
- The independent variable does not affect the dependent variable if the significance level is greater than (0.05).

2. Test Effectiveness N-Gain Score

The N-Gains score test can be used when there is a significant difference between the mean post-test scores of the experimental group and the control group, as determined by an independent samples t-test. The N-Gain score is calculated using the following formula:

$$NGain = \frac{Posttest\ score - Pretest\ score}{Ideal\ score - Pretest\ score}$$

The N-Gain score categories, expressed as percentages, can be found in the following table:

Table. 3.16 Interpretation of N-Gain

Presentense	Interpretation
< 40	Ineffective
40 – 55	Less effective

56 – 75	Moderately effective
> 76	Effective

Source (Hake,R.R,1999)

