CHAPTER III
RESEARCH METHODOLOGY

3.1 Research Design

A quantitative approach with a quasi-experimental design was employed to examine the impact of the THIEVES reading comprehension strategy on students. Although a control group was included in the study design, it was not fully able to eliminate external variables that were unrelated to the main research focus and could potentially influence the outcomes. This quasi-experimental method involved the use of a non-equivalent control group. (Sugiyono, 2013). Participants in an experimental group receive treatment(s) from the researcher but do not receive treatment(s) from the researcher if they are placed into a control group.

Table 3. 1 Non-equivalent Control class Design

Classes	Pre-test	Treatment	Post-test
Experimental Class	O_1	X	O_2
Control Class	O_3		O_4

Where:

O_1 = Pretest score experimental class

O_2 = Posttest score experimental class

O_3 = Pretest score control class

O_4 = Posttest score control class

If there is no significant difference between the experimental and control groups in the pre-test scores, it shows that the experimental group scored equivalently as well. To understand the impact of treatment, take the difference in the mean post-test score (O2 - O1)-(O4-O3).

In this research, two classes were included, one serving as the experimental class and the other as the control class. The experimental class was instructed to use THIEVES, which is a new method for teaching/learning; however, the control class used traditional methods of teaching/learning. To pre-test each class before any treatment was initiated and to determine what the students already knew about reading comprehension, they were given a pre-test; both classes received an identical pre-test before any instructional treatment. Following this, the experimental class received instructional treatment by using THIEVES as a teaching method; however, the control class received instructional treatment by following traditional teaching methods.

3.2 Population and Sample

3.2.1 Population

Population is a broad classification of objects or subjects that share common traits or attributes as determined by researchers to conduct research and make conclusions (Sugiyono, 2013, p. 80). The research population comprised 332 students of SMA Muhammadiyah 8 Cerme Gresik.

3.2.2 Sample

A purposive sample was chosen for the research. Sample selection using purposive sampling is a selection of participants that were selected based on predetermined criteria (Sugiyono 2013, p. 82). The researcher used predetermined criteria for selecting which class would be included in the sample. For instance, the class had to be on the same grade level and have the equivalent ability to demonstrate reading comprehension according to prior reading assessments conducted by the participants' teacher. The sample was determined by ensuring that the 10th-grade classes had the same level of reading comprehension ability. Information regarding this is presented in the table below:

Table 3. 2 Students Previous Test-Score

No	Classes	Test score	Rounded Score
1	X-1	82,27419355	82
2	X-2	83,66129032	83
3	X-3	83,65625	83
4	X-4	89,4375	89

Using the average test results, the researcher determined that X-2 would serve as the control class and X-3 as the experimental class.

3.2.3 Homogeneity Test

If the data groups samples of interest are homogeneous with regard to variances, it's possible to assess their variances (Sugiyono, 2013, p. 199). A test for homogeneity is used to identify whether the variances of two groups (control versus experimental) are equal or not. Homogeneous samples are said to be composed of similar subjects when they exhibit similar distributions or characteristics.

3.3 Research Instrument

The researcher utilized this assessment as a tool for their research by performing a number of testing procedures such as administering pre-tests prior to intervention commencement and

post-tests afterwards. Brown (2004, p. 207) lists several samples open-ended questions used as study instruments as follows:

Table 3. 3 Indicators of the Instrument

No	Indicator	Pre-test	Post-test
1	Students are able to determine the main idea of the text	What is the central idea of the text?	What is the central idea of the text?
2	Students can identify the author's purpose	What is the author's purpose in writing this text?	What is the author's purpose in writing this text?
3	Students are able to interpret the message conveyed in the text	What lesson or message can be drawn from the text?	What lesson or message can be drawn from the text?
4	Students can evaluate the relevance and applicability of the text to personal needs, goals, or interests.	Which part of this text is most relevant to your personal interests or future plans?	Which part of this text is most relevant to your personal interests or future plans?
5	Students can conclude from indirect statements	What can be inferred about the author's opinion regarding the topic discussed in the text?	What can be inferred about the author's opinion regarding the topic discussed in the text?

3.3.1 Rubric Scoring

A scoring rubric assesses the following aspects to measure students' critical reading comprehension. In this research, the researcher adopted and used the scoring rubric from (Leist et al., 2012) as follows:

Table 3. 4 Rubric Scoring

Standard and Elements	1	2	3	4	Points
ACCURACY: Identifying the main purpose and/or key ideas in the reading	Completely inaccurate; no correct identification of purpose or ideas	Low level of accuracy; purpose or ideas are incorrectly stated	Moderately accurate; includes correct purpose and ideas but with minor errors	Fully accurate; purpose and key ideas are clearly and correctly identified	

CLARITY: Understanding supporting details such as facts, data, or examples	No reference to facts, data, or examples from the text	Minimal or incorrect use of supporting information	Some correct use of facts, data, or examples	Consistent and accurate use of supporting details from the text	
PRECISION: Use of subject-specific vocabulary from the text	No use of academic or content-specific vocabulary	Limited and incorrect use of key vocabulary	Some appropriate use of vocabulary, with occasional correct paraphrasing	Accurate and frequent use of content-specific terms, often with correct paraphrasing	
DEPTH: Understanding relationships between ideas (purpose, concepts, support)	No understanding of relationships among ideas	Limited awareness of how ideas are connected	General understanding of relationships between concepts and support	Deep and well-developed understanding of how ideas are interconnected	
RELEVANCE: Drawing conclusions and relating ideas personally or contextually	No conclusions or relevance identified	Weak or basic conclusions with little connection	Some relevant conclusions but limited personal connection	Strongly relevant conclusions with clear personal or contextual connections	

LOGIC: Applying ideas to broader contexts	No application of ideas beyond the text	Poor or incorrect application of concepts	Some application using general or broad ideas	Clear and effective application of concepts to wider contexts with relevant examples and details	
Total of the Points					

3.3.2 Validity of the Instrument

Content valid instruments are those that accurately assess what they are meant to assess. The researcher in this instance performed content validity testing to determine if the instrument he used to obtain the data was indeed valid. Content validity testing compares the instrument's content to the content area of previously instructed subject matter (Sugiyono, 2013). The instrument used consisted of 5 critical thinking items and was given at both baseline (pre-test) and after the program (post-test).

Aiken (1985) used expert reviewer ratings of individual items to calculate a content validity coefficient (Aiken's V) based on the agreement among several experts regarding how well an item reflects the meaning it was designed to represent. He also established a set of guidelines to assess

- A validity coefficient of 0.60–0.89 indicates a high level of validity
- A validity coefficient of 0.40–0.59 indicates a moderate level of validity
- A validity coefficient of 0.20–0.39 indicates a low level of validity
- A validity coefficient of 0.00–0.19 indicates a very low level of validity

Table 3. 5 Validity Aiken's

Questions	Expert			S1	S2	S3	$\sum s$	n(c-1)	V	Ket
	I	II	III							
Q1	4	4	4	3	3	3	9	9	1	Very High
Q2	4	3	3	3	2	2	7	9	0.7777777778	High
Q3	4	4	4	3	3	3	9	9	1	Very High
Q4	4	3	3	3	2	2	7	9	0.7777777778	High
Q5	4	4	3	3	3	2	8	9	0.8888888889	Very High
Total	20	18	17	15	13	12	40	45	0.8888888889	Very High

Based on Aiken's V formula on table 3.5, it can be showed that the result is 0,88889 which means that the instrument of this research is valid.

3.3.3 Reliability of the Instrument

It is crucial for research instruments to exhibit reliability, which indicates

their ability to exhibit both consistency and dependability over a particular period. The reliability of an instrument can be evaluated using the internal or external consistency methods suggested by Sugiyono (2013) on page 130. In the case of this study, the internal consistency method was selected for use by utilizing the Cronbach Alpha formula. As indicated by Sugiyono (2013), the reliability of an instrument is evidenced by a Cronbach Alpha score of .60 or higher. The data provided above is criteria to determine if the selected instrument for research is actually reliable:

- An instrument is considered reliable when the reliability coefficient exceeds 0.6.
- Conversely, if the reliability coefficient is below 0.6, the instrument is regarded as not reliable.

The results of the **Cronbach's Alpha test** are presented in the table below.

Reliability Statistics

Cronbach's Alpha	N of Items
.600	3

Based on Table 3.6, the Cronbach's Alpha value obtained is 0.600. This indicates that the research instrument can be considered reliable.

3.4 Data Collection

This study used a pre-test and post-test design to collect data from the selected classes. At the beginning, a pre-test was conducted to measure the initial equivalence of the samples and to determine the baseline performance between the experimental and control groups. After this stage, the students were assigned to their respective groups. The experimental class received instruction through the THIEVES strategy, while the control class was taught using conventional teaching methods without any additional treatment.

a) Pre-test

The following describes how the data was collected from the participants: a pre-test was given during the first meeting which was used to determine whether or not there was any homogeneity among the sample as well as to establish a baseline difference in critical reading comprehension scores between the two class groups prior to receiving their respective treatments.

b) Treatment

The experimental group of participants received a pre-test followed by three preview meetings prior to taking a comprehension test on the fourth meeting. The students in the experimental group used the THIEVES strategy as a method of previewing the text prior to reading it; while students in the control group had the same number of preview meetings, they did not preview the text using the THIEVES strategy, they previewed using a conventional previewing strategy, as they normally would. All four of the meetings for the experimental and control groups contained identical materials. In meetings one through three, participants in the experimental group used the

THIEVES strategy in the following manner:

1) The teacher explains to the students about what the THIEVES strategy is
2) The students are given the THIEVES practice paper to
3) The students are given some exercises after using the THIEVES strategy

Meanwhile, the control class was taught using the conventional strategy usually used.

c) Post-test

The researcher will perform a post-test after applying the treatment during the last meeting. This post-test will be conducted on two classes: the experimental class and the control class. The goal of the post-test is to see if there is a difference between how well students can read critically after having been taught using the THIEVES method compared to being taught using regular teaching strategies. The researcher will analyze the results of the post- test once it has been completed.

3.5 Data Analysis

After collecting the data, the researcher must analyze the data itself. The researcher uses the technique of data analysis as follows:

3.5.1 Descriptive Statistic

3.5.1.1 Mean Score

The mean score is often used to represent the data. To know the mean score, the researcher used the descriptive statistic on SPSS 26

3.5.1.2 Standard Deviation

In analyzing the data, the researcher needs to measure the data to determine the variation in the data distribution. To know the standard deviation of each class, the researcher used the descriptive statistic on SPSS 26

3.5.1.3 Normality Test

The normality test is used to determine whether the data follows a normal distribution. According to Widodo et al. (2023, p. 116), this test can be carried out either manually or with the assistance of SPSS software. In SPSS, normality can be examined using the Kolmogorov-Smirnov and Shapiro-Wilk tests. Meanwhile, manual procedures may involve the Chi-Square and Lilliefors tests.

In this study, the researcher applied the Shapiro-Wilk test to assess data normality since the sample size was fewer than 100 participants. The decision criteria are as follows:

- If the significance (Sig.) value is less than 0.05, the data distribution is considered not normal.
- If the significance (Sig.) value is equal to or greater than 0.05, the data distribution is considered normal.

3.5.1.4 Hypothesis Test

To determine the effectiveness of the treatment, the collected data will be examined using an independent samples t-test. This statistical analysis compares the performance of students in both the experimental and control groups based on the following hypotheses:

- Null Hypothesis (H0): The experimental and control groups do not differ significantly in their critical reading comprehension performance.
- Alternative Hypothesis (H1): There is a significant difference in critical reading comprehension performance between the experimental and control groups.

The conclusions of this study are based on the following significance rules:

- When the 2-tailed significance value is below 0.05, the alternative hypothesis (H1) is supported while the null hypothesis (H0) is rejected.
- When the 2-tailed significance value exceeds 0.05, the null hypothesis (H0) is accepted and the alternative hypothesis (H1) is rejected.