CHAPTER III
RESEARCH METHODOLOGY

This chapter details the methodology used to conduct the research. It specifies the chosen design, how the data was collected, and the techniques used for analysis, concluding with a clear schedule that maps out the study's timeline.

3.1 Research Design

This study uses a quantitative approach by applying a correlational design. This approach is used given its relevance to the research questions, which aim to examine the relationship between two variables using numerical data systematically (Waruwu et al., 2025). Meanwhile, the correlational design is considered appropriate as one of the study designs for predicting potential outcomes between two or more variables by examining the degree of their relationship (Nurhaswinda et al., 2025).

3.2 Research Variables

A variable refers to any attribute or characteristic that can be observed and studied by researchers to obtain information that can later be analyzed for drawing conclusions (Sugiyono, 2013). In this study, there are two variables: reading fluency as the independent variable and reading comprehension as the dependent variable. The independent variable, also commonly called the antecedent variable, is a variable that is assumed to cause or influence changes in another variable and is commonly represented by X. Meanwhile, the dependent variable represents the outcome that is affected by one or more independent variables and is commonly denoted as Y (see Figure 6).

Figure 6 Research Variables

3.3 Population and Sample

3.3.1 Population

The population refers to the entire group of individuals relevant to a study, all of whom meet the research criteria (Willie, 2024). The population of this study was the sixth-semester students from the English Education Department. There were approximately 163 students in total. The rationale for electing this demographic is because they have already undergone instructional treatment to improve their decoding capabilities through several courses, such as Phonetics and Phonology and Introduction to English Pronunciation. Students during the Phonetics and Phonology course learn to recognize letter-sound correspondences, whereas in Introduction to English Pronunciation, they are taught to read English words accurately, word by word, in sentence form. Moreover, courses like Intensive Reading and Extensive Reading most frequently expose students to English texts, which enables them to improve their reading fluency. Additionally, when compared to the fourth-semester students and the eighth-semester students who have undergone the same instructional treatment, the sixth-semester students tend to exhibit more exposure to the English text; meanwhile, the eighth-semester students are generally more focused on completing their final projects, which introduces potential confounding variables, such as different learning priorities, that could make it harder to carry out the study. Thus, the sixth-semester students were chosen for this study because they had more exposure to the English text. It would support this study to seek a relationship between EFL students' reading fluency and comprehension.

3.3.2 Sample

The researcher employed simple random sampling, ensuring that each student had an equal opportunity to be selected. To determine the sample size, this study utilized the Harry King Nomogram (Sugiyono, 2013). By drawing a line from the total population through the predetermined error level on the Harry King Nomogram, the specific percentage of the population required for the study's sample was identified.

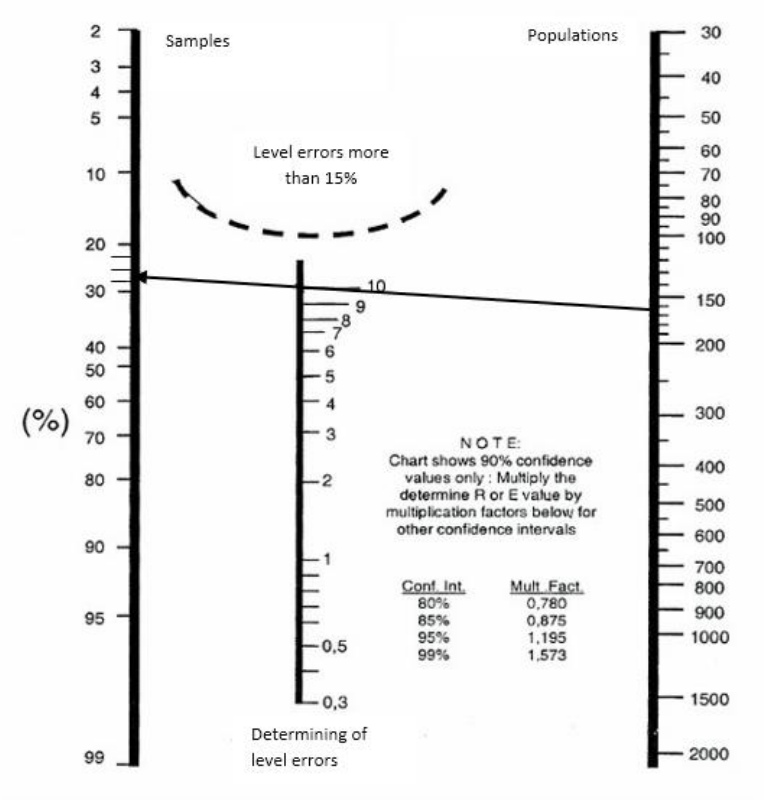

Figure 7 Harry King Nomogram

Based on Figure 7, the predetermined error level for this study was set at 10%, which corresponded to a 90% confidence level as per the standard parameters of the Harry King Nomogram. Consequently, for a population of 163 students, the nomogram indicated a sampling proportion of 26%. The conversion of this percentage yielded a value of 42.38, which was rounded up to 43. Therefore, the final sample for this research consisted of 43 participants. The following section provides an overview of the participants' characteristics based on the collected data, which have been processed using descriptive statistical analysis.

3.3.2.1 Distribution of Sample by Gender

The total number of participants in this study is 43 EFL students. Within this group, there are 13 male participants, while female participants represent a higher proportion with a total of 30 individuals. The detailed distribution of the participants based on gender is presented in Table 3.

Table 3

Distribution of Participants by Gender

Gender	N	Percentage
Male	13	30.2%
Female	30	69.8%
Total	43	100%

3.3.2.2 Distribution of Sample by Class

Regarding the distribution by class, the participants are categorized into four groups. The majority of the EFL students are from Class B, while Class A represents the smaller portion of the sample. The detailed distribution of the participants based on their respective classes is presented in Table 4.

Table 4

Distribution of Participants by Class

Group	N	Percentage
Class A	2	4.6%
Class B	22	51.2%
Class C	11	25.6%
Class D	8	18.6%
Total	43	100%

3.4 Data Collection

3.4.1 Validity and Reliability

Before conducting the test, the researcher should take attention on the validity and reliability of the test used.

3.4.1.1 Validity

3.4.1.1.1 Validity of Reading Fluency Instrument

Reading fluency consists of two components: automaticity and prosody. Prosody plays a vital role in determining reading competence and identifying reading fluency (Schreiber, 1991). However, according to Valencia et al. (2010), assessing prosody is more difficult compared to speed and accuracy in automaticity. Prosody involves reading by paying attention to many elements like intonations,

stress, etc. The development of rubrics to assess prosody began with Allington (1983), followed by Zutell & Rasinski (1991), which dealt with three dimensions of phrasing, smoothness, and pace, with each organized on four levels. The rubric was later updated by Rasinski (2004) to include a fourth dimension, expression and volume. This update was completed to allow a separate assessment of the four basic features included in prosody. This, known as the Multidimensional Fluency Scale (MFS), has become the most commonly used rubric to assess prosody (Arıkan et al., 2022). It was shown to be among the best scales to assess prosody (Benjamin et al., 2013) in many studies using the rubrics. Meanwhile, automaticity is assessed through Reading Curriculum-Based Measurement (R-CBM), which is recognized as an effective instrument for measuring student automaticity using texts integrated into the reading curriculum (Rasinski, 2004). In this study, the material utilized is an analytical exposition text titled "Global Warming", from More Reading Power (Mikulecky & Jeffries, 1996, p. 271), the primary textbook used in the students' reading course. This passage was recommended by the Intensive Reading course instructor for this research, with a readability level measured at 63 on the Flesch Reading Ease scale. This score indicated that the passage is written in plain English (see Enclosure 7), making it both readable and appropriate for EFL college students as it falls within the standard range of 60 – 70 (Flesch & Ferry, 1948).

Both MFS dan R-CBM are measurement tools recommended by Rasinski (2004) in his book, *Assessing Reading Fluency*.

3.4.1.1.2 Validity of Reading Comprehension Instrument

The reading comprehension test comprises of 30 multiple-choice questions, developed using AI-assisted platform Monsha.ai and grounded in Barrett's Taxonomy framework. To ensure consistency across the assessment materials, these questions are based on the same passage used for the reading fluency test. The test design follows the format established by readingsoft.com to effectively evaluate the students' understanding of the material they have read.

To validate the instrument, experts reviewed each item against established reading comprehension indicators. The experts, who comprised the researcher's two supervisors, rated the relevance of each question on a scale of 1 to 5. Their

ratings were then be analyzed using the Aiken's V formula. According to L. R. Aiken (1985) distribution table, for a panel of two experts, the V-value must reach a maximum of 1.00 to achieve statistical significance ($p < 0.05$). Consequently, only questions achieving a V-value of 1.00 were categorized as valid (see Enclosure 5). On this basis, any items that did not meet this threshold were discarded.

Table 5

Content Validity Results for Multiple-Choice Questions

Number of Items	Indicators	Rater 1	Rater 2	S1	S2	ΣS	V	Judgment
Multiple-Choice Questions								
1	[1]	5	5	1	1	2	1.00	Valid
2	[2]	5	5	1	1	2	1.00	Valid
3	[1]	5	5	1	1	2	1.00	Valid
4	[2]	5	5	1	1	2	1.00	Valid
5	[1]	5	5	1	1	2	1.00	Valid
6	[2]	5	5	1	1	2	1.00	Valid
7	[1]	5	5	1	1	2	1.00	Valid
8	[2]	5	5	1	1	2	1.00	Valid
9	[1]	5	5	1	1	2	1.00	Valid
10	[2]	5	5	1	1	2	1.00	Valid
11		5	5	1	1	2	1.00	Valid
12	[3]	5	5	1	1	2	1.00	Valid
13		5	5	1	1	2	1.00	Valid
14		5	5	1	1	2	1.00	Valid
15		5	5	1	1	2	1.00	Valid
16		5	5	1	1	2	1.00	Valid
17		5	5	1	1	2	1.00	Valid
18	[4]	5	5	1	1	2	1.00	Valid
19		5	5	1	1	2	1.00	Valid
20		5	5	1	1	2	1.00	Valid
21		5	5	1	1	2	1.00	Valid

Number of Items	Indicators	Rater 1	Rater 2	S1	S2	ΣS	V	Judgment
Multiple-Choice Questions								
22		5	5	1	1	2	1.00	Valid
23		5	5	1	1	2	1.00	Valid
24		5	5	1	1	2	1.00	Valid
25		5	5	1	1	2	1.00	Valid
26		5	5	1	1	2	1.00	Valid
27	[5]	5	5	1	1	2	1.00	Valid
28		5	5	1	1	2	1.00	Valid
29		5	5	1	1	2	1.00	Valid
30		5	5	1	1	2	1.00	Valid

As presented in Table 5, all multiple-choice questions obtained a V-value of 1.00, confirming their content validity. In addition to content validity, an item validity test was also conducted. Item validity was performed to verify the instrument's effectiveness through a pilot test before the primary data collection. A multiple-choice question was categorized as valid if the r-value exceeded the r-table value. Conversely, questions with an r-value lower than the r-table were categorized as invalid. For a sample size of 43 students (N = 43) at significance level of 0.05, the r-table value is 0.301 (see Enclosure 6). The results of the item validity test are presented in Table 6.

Table 6

Item Validity Results for Multiple-Choice Questions

Number of Items	r-value	r-table	Judgment	Number of Items	r-value	r-table	Judgment
1	0.793	0.301	Valid	16	0.615	0.301	Valid
2	0.615	0.301	Valid	17	0.451	0.301	Valid
3	0.646	0.301	Valid	18	0.793	0.301	Valid
4	0.132	0.301	Invalid	19	0.000	0.301	Invalid
5	0.110	0.301	Invalid	20	0.438	0.301	Valid
6	c	0.301	Invalid	21	0.511	0.301	Valid

Number of Items	r-value	r-table	Judgment	Number of Items	r-value	r-table	Judgment
7	0.061	0.301	Invalid	22	0.548	0.301	Valid
8	0.110	0.301	Invalid	23	0.088	0.301	Invalid
9	0.793	0.301	Valid	24	0.619	0.301	Valid
10	c	0.301	Invalid	25	0.548	0.301	Valid
11	c	0.301	Invalid	26	c	0.301	Invalid
12	0.225	0.301	Invalid	27	c	0.301	Invalid
13	0.621	0.301	Valid	28	0.793	0.301	Valid
14	0.387	0.301	Valid	29	0.193	0.301	Invalid
15	0.702	0.301	Valid	30	-0.044	0.301	Invalid

Based on Table 6, a total of 16 items were categorized as valid, while the remaining 14 items were identified as invalid. Consequently, the final reading comprehension test consisted of 16 valid multiple-choice questions. The valid items were numbers 1, 2, 3, 9, 13, 14, 15, 16, 17, 18, 20, 21, 22, 24, 25, and 28. Conversely, the invalid items consisted of numbers 4, 5, 6, 7, 8, 10, 11, 12, 19, 23, 26, 27, 29, and 30, which were excluded from the final reading comprehension test.

3.4.1.2 Reliability

The test should be reliable as a measuring instrument. Waruwu et al. (2025) states, reliability refers to the extent to which an instrument produces consistent and stable results when used repeatedly under the same conditions. Furthermore, a reliable instrument will yield similar results if tested again at different times or by different researchers. In this study, a reliability test was conducted for both reading fluency and reading comprehension, as explained below.

3.4.1.2.1 Reliability of Reading Fluency Instrument

In this study, data on EFL students' reading fluency were collected through Reading Curriculum-Based Measurement (R-CBM) and Multidimensional Fluency Scale (MFS). The reliability of these instruments is established as follows.

1) Reading Curriculum-Based Measurement

The Reading Curriculum-Based Measurement (R-CBM) reliability is maintained through procedural consistency and objective scoring. As R-CBM provides a direct measure of student performance through Words Correct Per

Minute (WCPM), the researcher ensure reliability by strictly adhering to standardized administration guidelines for all participants (see Enclosure 2).

2) Multidimensional Fluency Scale

The Multidimensional Fluency Scale (MFS) reliability is established as reliable to assess prosody, with an inter-rater reliability coefficient ranging from 0.91 to 0.94 (Smith & Paige, 2019), suggesting excellent consistency between raters. Others studies in the research have confirmed that valid and reliable measures were obtained by rating using this rubric (Moser et al., 2014; Rasinski et al., 2009, 2017; Valencia et al., 2010; Zimmerman et al., 2019).

3.4.1.2.2 Reliability of Reading Comprehension Instrument

Drawing from the results of the pilot test conducted, the reading comprehension instrument demonstrated a Cronbach's Alpha reliability coefficient of 0.777 (see Table 7). This result gains significant weight when viewed through the lens of Kerlinger & Lee (2000), who maintain that reliability scores falling within the 0.7 to 0.8 range are adequate for the requirements of basic research. Given these established parameters, it is evident that the multiple-choice format of the instrument successfully fulfils the criteria for reliability.

Table 7
Cronbach's Alpha Coefficient for Reading Comprehension Test Reliability

	Cronbach's Alpha	N of Items
Multiple-Choice Questions	0.777	30

3.4.2 Instrument

This study employs several instruments to secure accurate and comprehensive data for the variables examined. The following section details the instruments used for each variable.

1) Reading Fluency

Data on reading fluency were collected through an oral reading test assessed using the Reading Curriculum-Based Measurement (R-CBM) and Multidimensional Fluency Scale (MFS) instruments. To ensure scoring accuracy and facilitate repeated verification, each performance was audio-recorded as data were collected through an oral reading test. Each student was asked to read the

chosen passage aloud at their natural pace, with timing commencing immediately upon the start of the reading. While they completed the entire text, the automaticity assessment focused exclusively on the first minute. During that initial minute period, uncorrected errors such as mispronunciations, omissions, or hesitations exceeding three seconds were marked to calculate the Words Correct Per Minute (WCPM). Conversely, the prosody assessment, based on Rasinski (2004) scale, spanned the full duration of the reading, with the final fluency score derived from the summation of all assessed dimensions, including expression and volume, phrasing, smoothness, and pace.

The selection of the R-CBM and MFS instruments in this study was based on the reading fluency assessment framework proposed by Rasinski (2004), considering that there was no single reading fluency assessment that evaluated all aspects of fluency. R-CBM measured speed and accuracy, while MFS measured prosody. Hasbrouck & Tindal (2006) noted that R-CBM could be administered efficiently. Meanwhile, the MFS offered a broad scoring range (up to 16 points), which allowed for a more accurate assessment of participants' prosody based on their reading performance (Jihye, 2021).

2) Reading Comprehension

Data on reading comprehension were collected through a multiple-choice test consisting of 16 questions. The distribution of the items can be seen below.

Table 8

Blue Print of Reading Comprehension Test

No	Indicators	Number of Items	Total Items
1	Determining the main idea	2	1
2	Finding specific information in the text	5, 6, 7, 8, 9	5
3	Finding reference	1, 3, 4	3
4	Finding inference	10, 11, 12, 13, 14, 15	6
5	Identifying vocabulary	16	1

Multiple-choice tests have long been recognized as a practical and effective tool for assessing reading comprehension. As noted by Butler (2018), it may be the most commonly used format in educational assessment due to its scalability and

objectivity. However, its reliability depends on how the questions are designed to reflect key aspects of comprehension. It also can provide high validity, especially when the text remains available during the test, allowing students to engage in deeper processing (Sedlmayr & Weissenbacher, 2025). Well-designed multiple-choice items can effectively measure comprehension components such as inference, vocabulary in context, and logical reasoning (Saswati, 2021).

3.5 Data Analysis

The research data comprises two primary variables: reading fluency and reading comprehension. Fluency data was gathered through the Reading Curriculum-Based Measurement (R-CBM) and the Multidimensional Fluency Scale (MFS), while comprehension data was obtained via a 16-item multiple-choice test (MCQ). All datasets were processed using SPSS 24, which was also utilized to investigate the correlation between the two variables. Consequently, the following analysis is organized into three sections: the analysis of students' reading fluency achievement, the analysis of students' reading comprehension achievement, and the evaluation of the relationship between the two.

3.5.1 Analysis Data on Reading Fluency

The analysis of reading fluency data in this study was conducted through the following stages.

1) Descriptive statistical analysis was performed on the R-CBM and MFS scores to determine the maximum and minimum scores, mean, and standard deviation. Simultaneously, the scores from both instruments were standardized to account for their differing rating scales, allowing them to be integrated into a unified reading fluency metric.
2) The standardized R-CBM and MFS scores were then aggregated into a single composite score using the 'Transform' then 'Compute Variable', function in SPSS. This integration is based on the conceptualization of reading fluency as a single, multi-faceted construct.
3) A descriptive statistical analysis was subsequently conducted on this single reading fluency score to identify its maximum, minimum, mean, and standard

deviation.

4) Visual Binning was applied to the composite reading fluency scores to categorize student achievement into three levels: low, average, and high. The results from this visual binning were then converted into percentages through frequency analysis, providing a clear distribution of students across each achievement category.

3.5.2 Analysis Data on Reading Comprehension

The analysis of reading comprehension data in this study was conducted through the following stage.

1) The raw data collected from the multiple-choice questions (MCQ) were converted into a standardized scale with a maximum score of 100, using the formula: (score earned/total points) x 100.
2) Subsequently, the converted data were analyzed using descriptive statistics to provide a general overview of the maximum and minimum scores, mean, and standard deviation.
3) Visual Binning was applied to the standardized scores to categorize students' reading comprehension achievement into three levels: low, average, and high. The results from the visual binning were then converted into percentages through frequency analysis, enabling a clear distribution of students across each achievement category in reading comprehension.
4) To analyze students' performance across specific reading comprehension indicators, an item-based analysis was performed. The data for each indicator, including finding main idea, specific information, reference, inference, and vocabulary, were processed to identify the Mean Item Score.
5) The Mean for each indicator was calculated using the 'Descriptive Statistics' function. This total mean was then divided by the number of items assigned to each indicator to produce the Mean Item Score, providing a precise measurement of mastery for each sub-skill.

3.5.3 The Correlation between Reading Fluency and Reading Comprehension

To investigate whether a relationship exists between EFL students' reading fluency and their reading comprehension, this study used the Pearson Product-Moment correlation. This approach was chosen because the data followed a normal distribution (see Enclosure 3.5), which is the main requirement for using this kind of statistical test. The formula utilized to compute this correlation is provided below.

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

Where:

r = correlation coefficient value
N = number of pairs of scores
Σxy = sum of the products of paired scores
Σx = sum of x scores
Σy = sum of y scores
Σx^2 = sum of squared x scores
Σy^2 = sum of squared y scores

The correlational analysis was systematically executed through the following stages.

1) The relationship between the two variables was calculated by using the 'Analyze' menu in SPSS 24, moving to 'Correlate', and then selecting the 'Bivariate' option. This process links the fluency scores and comprehension scores to see how they move together.
2) After the software processed the data, the resulting table showed a value known as the correlation coefficient (*r*). This r-value is found in the row labeled 'Pearson Correlation' where the two variables meet in the final table.
3) To understand what the r-value actually means, the categories suggested by Cohen (2013), a value between 0.10 and 0.29 is considered a weak connection, 0.30 to 0.49 is a moderate one, and anything from 0.50 to 1.00 shows a strong relationship.
4) The last step involved looking at the results to see if the correlation was

significant.

3.6 Research Schedule

The research time in question is the time span for the implementation of the research as outlined in the form of a table, as follows.

Description	Sept/ 2025	Oct/ 2025	Nov/ 2025	Dec/ 2025	Jan/ 2026	Feb/ 2026	Mar/ 2026	Apr/ 2026
Research Proposal writing								
Research Proposal examination								
Data Collection								
Data Analysis								
Report								
Thesis Result Seminar								
Thesis Examination								