CHAPTER III
RESEARCH METHODOLOGY

3.1 Research Design

A cross-sectional survey design was used for this quantitative research. Quantitative methods emphasize collecting data in the form of exact numbers through instruments and explaining it with statistical tests. Research that uses quantitative methods has the characteristic of being more systematic, planned, structured, clear from beginning to end of research and not influenced by conditions in the field (Ahyar et al., 2020). A survey design in the world of education is a test that is used without any special treatment of the sample, manipulation of data or conditions during the research. It cannot explain cause and effect relationships or comprehensive explanations, but rather trends in the data (Creswell, 2012). This design is used to describe the characteristics of a population (Fraenkel et al., 2011).

While cross-sectional surveys have limitations in establishing cause and effect, their strengths in capturing trends and collecting data efficiently make them a popular choice for the initial phase of research. By using cross-sectional surveys, researchers can gain a valuable picture of the current educational landscape, identify potential areas for attention, and formulate more specific hypotheses for further investigation (Kesmodel, 2018). In essence, cross-sectional surveys serve as a springboard for more in-depth research efforts. Cross-sectional survey design is a method where relevant information or data collection is analyzed at the same time (Creswell, 2012).

Some methods used to analyze quantitative data are verification research which functions to determine the influence of each variable; comparative causal to determine the cause-and-effect relationship of each variable; inferential to determine the relationship between variables; experimental to determine possible causes and effects; and descriptive by describing each variable. This research was analyzed using descriptive methods. It is a method used to collect information on a phenomenon when research is carried out with the character of exploring and clarifying a phenomenon through variable descriptions and does not test existing hypotheses but rather describes the research results as they are (Zellatifanny & Mudjiyanto, 2018).

3.2 Participants

This research sample focused on junior high school students who had studied the circulatory system. Apart from that, the sample targeted in this research were students who studied science with the latest curriculum, namely the *Kurikulum Merdeka*. 233 samples with an eighth-grade background in an age range of 13-14 years were collected from several junior high schools in Bandung Regency, West Java Province. The large number of samples used to support a more varied representation of information. In determining the sample, researchers used convenience sampling. Convenience sampling is a sample-selected technique regarding the willingness and availability of sample to take part of the research (Wang & Cheng, 2020). Additional information, it is a non-probability sampling technique which requires little effort, cost, time, and simple operation (Golzar & Noor, 2022).

3.3 Research Instrument

The instrument used in this research is a questionnaire containing questions with a four-level diagnostic test model. This test is a development and modification of a test from a three-level diagnostic test by changing the test into two parts, questions that function to measure students' cognitive understanding and another part that is the reason for answering the questions. The questions are made in the form of four options with the aim of limiting students' opportunities to answer carelessly and not according to the existing context and concept. To diagnose misconceptions in students, this question was initially developed through a preliminary study which was then modified into a four-level diagnostic test. The researcher used several concepts developed from the sub-topics which can be seen in Table 3.1.

Table 3.1 The Distribution of Questions

Sub-Topics	Concepts	Item Questions
Structure and Function of Blood Circulatory Organs	Heart	Q1. Q2, Q17, Q21
	Blood Vessels	Q3, Q4, Q19, Q26
	Blood	Q5, Q6, Q16, Q20
Diseases related to the circulatory system	Atherosclerosis and Heart Attack	Q7, Q8, Q15, Q27

Sub-Topics	Concepts	Item Questions
	Stroke	Q9, Q1O, Q17, Q24
	Hypertension	Q11. Q12, Q25, Q28
	How to maintain the health of our circulatory system	Q13, Q14, Q18, Q22

Seven concepts were tested in this research. This concept is divided into two large subtopics, namely the structure and function of the circulatory system, as well as matters related to diseases of the circulatory system. Initially the researcher used 14 questions with two questions representing each concept in the preliminary study. Then the questions were developed into four questions for each concept. Examples of questions developed using a four-tier diagnostic test structure can be seen in Table 3.2.

Table 3.2 A Sample Question

Item	Tier	Question
Q4	**1**	Why do blood vessels have to be different sizes?
		A. To maintain and adjust blood pressure
		B. To avoid the decreasing oxygen in the blood
		C. To adjust the size of passing particles
		D. To increase the elasticity of flowing blood
	2	Are you sure of your answer?
		A. Sure B. Not Sure
	3	What is the scientific reason of your answer?
		A. Size responds equally to contraction
		B. Size supports optimal ox8ygen exchange
		C. Size maintains the continuity of blood flow
		D. Size flows substances efficiently
	4	Are you sure of your answer?
		A. Sure B. Not Sure

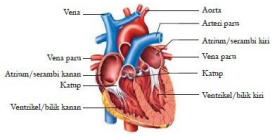
In the four-tier diagnostic test structure, the first tier is to measure students' understanding of the topic, which is then included in the level of student confidence at

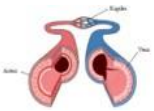
the second tier. The third tier is the reason behind the student's answer at the first tier which is confirmed by the student's level of confidence at the fourth tier. Before conducting real research, the questions created are validated by experts and have gone through a validity test first.

3.3.1 Preliminary Study

A preliminary study is the initial stage in research and development of a research model. Preliminary studies can be used to analyze problems as an illustration for conducting further research. In this topic, a preliminary study is used in the form of open questions in a two-level form. The first level is in the form of questions used to measure students' understanding of the topic, and the second level is the reason for the answer to the question. The open-ended questions –used in the preliminary study can be seen in Table 3.3.

Table 3.3 Preliminary Questions

Item	Part	Question
1	**1**	Valves are the separation between the atria and ventricles that help the heart work. What are the fatal consequences when the valve does not function and blood flows back to the heart?
	2	Give a brief explanation of your previous answer!
2	**1**	Pay attention to the figure of structure of the heart below!  source: shutterstock.com/ilusmedical In the structure of the heart, the ventricles have stronger and thicker muscles than the atria. What happens if the ventricle muscle becomes thin?
	2	Give a brief explanation of your previous answer!
3	**1**	Pay attention to the figure of the blood vessels!

Item	Part	Question
		
		source: shuttershock.com/Blamb
		What happens if one of the blood vessels becomes blocked?
	2	Give a brief explanation of your previous answer!
4	**1**	Veins, arteries and capillaries are blood vessels that have different sizes. Why do you think blood vessels should be different sizes?
	2	Give a brief explanation of your previous answer!
5	**1**	Apart from plasma, what are blood cells that carry carbon dioxide?
	2	Give a brief explanation of your previous answer!
6	**1**	Why do white blood cells have a larger nucleus than red blood cells?
	2	Give a brief explanation of your previous answer!
7	**1**	What is the definition of atherosclerosis?
	2	Give a brief explanation of your previous answer!
8	**1**	One of the consequences of atherosclerosis is a heart attack. How can this happen?
	2	Give a brief explanation of your previous answer!
9	**1**	What is the main cause of stroke?
	2	Give a brief explanation of your previous answer!
10	**1**	One of the fatal consequences of a stroke is death, why does this happen?
	2	Give a brief explanation of your previous answer!
11	**1**	Before being checked, how can someone be diagnosed as having hypertension?
	2	Give a brief explanation of your previous answer!

Item	Part	Question
12	**1**	Hypertension cannot be considered trivial because it can attack several organs. What organs can be attacked?
	2	Give a brief explanation of your previous answer!
13	**1**	What foods should be avoided to maintain the circulatory system?
	2	Give a brief explanation of your previous answer!
14	**1**	Pay attention to the warnings on the following cigarettes!  source: www.kaskus.co.id How can inhaling cigarettes cause a heart attack?
	2	Give a brief explanation of your previous answer!

This preliminary study stage involved 33 grade 9 students who had studied the circulatory system. From the 14 questions, various answers were obtained which were then generalized and grouped to be used as the basis for creating multiple-choice questions for the four-tier diagnostic test. There were answers that contained misconceptions, even random ones. So, the development of the diagnostic test questions used alternatives from other sources. Several things that caused sub-optimal results are the allocation of time in taking the test, students who have forgotten the existing concepts, and students who do not take the test seriously. Generalized questions and answers can be seen in Appendix A.1.

3.3.2 Validation Test

After being developed into a four-tier diagnostic test structure, the results were validated. Validation is used to state the degree of accuracy of a measuring instrument, as well as its suitability for function and correct results (Sanaky, 2021). In validity test, the resulting score will be a reference in making decisions about test items (Muzaffar, 2016). In this research, logical validity and empirical validity were measured. Logical

validity is a test based on expert thinking or reasoning, while empirical validity is based on real evidence or experience carried out in research (Riyani et al., 2017). Based on the results gained, logistical validity was carried out through expert assessment by 3 experts (lecturers and a teacher). The test used for logistic validity is the Aiken Test with the following formula:

$$V = \frac{\Sigma s}{n(C-1)}$$

With the description,

V is the agreement index from the experts' rate

Σs is the average of the score gained from the rater minus the lowest score can be chosen ($s = r\text{-}l_o$)

n is the total or number of experts

C is the number of categories can be chosen.

Test validity can be categorized as low or less valid if the test value is less or equal to 0.4 ($0 < V \leq 0.4$), moderate or quite valid if the value is above the range 0.4 to less or equal to 0.8 ($0.4 < V \leq 0.8$), and high or very valid if the value is above the range 0.8 to 1 ($0.8 < V < 1$) (Fajaruddin et al., 2021). Based on the expert judgment, the validation of test item questions can be seen in Table 3.4.

Table 3.4 Aiken Test for Tier 1 Validation Result

Item	Expert			S_1	S_2	S_3	$\sum s$	n(c-1)	V	Validity
	I	II	III							
Q1	1	1	1	1	1	1	3	3	1.00	High
Q2	1	1	1	1	1	1	3	3	1.00	High
Q3	1	1	1	1	1	1	3	3	1.00	High
Q4	1	1	1	1	1	1	3	3	1.00	High
Q5	0	1	1	0	1	1	2	3	0.67	Moderate
Q6	1	0	1	1	1	1	3	3	1.00	High
Q7	1	1	1	0	0	1	1	3	0.33	Low
Q8	1	0	0	1	0	0	1	3	0.33	Low
Q9	1	0	1	1	0	1	2	3	0.67	Moderate

Item	Expert			S_1	S_2	S_3	$\sum s$	n(c-1)	V	Validity
	I	II	III							
Q10	1	0	1	1	0	1	2	3	0.67	Moderate
Q11	1	0	1	1	0	1	2	3	0.67	Moderate
Q12	1	0	1	1	0	1	2	3	0.67	Moderate
Q13	1	1	1	1	1	1	3	3	1.00	High
Q14	1	1	1	1	1	1	3	3	1.00	High
Q15	1	1	1	1	1	1	3	3	1.00	High
Q16	1	0	1	1	0	1	2	3	0.67	Moderate
Q17	0	0	1	0	0	1	1	3	0.33	Low
Q18	1	0	1	1	0	1	2	3	0.67	Moderate
Q19	1	1	1	1	1	1	3	3	1.00	High
Q20	1	1	1	1	1	1	3	3	1.00	High
Q21	1	0	0	1	0	0	1	3	0.33	Low
Q22	1	0	1	1	0	1	2	3	0.67	Moderate
Q23	1	0	1	1	0	1	2	3	0.67	Moderate
Q24	1	1	0	1	1	0	2	3	0.67	Moderate
Q25	1	0	0	1	0	0	1	3	0.33	Low
Q26	1	1	1	1	1	1	3	3	1.00	High
Q27	1	1	0	1	1	0	2	3	0.67	Moderate
Q28	1	0	0	1	0	0	1	3	0.33	Low

Based on this table, it was found that the average content validity value of the 28 questions was 0.73, which indicates moderate validity (Fajaruddin et al., 2021). This result shows that there is a strong correlation between scores and shows that it can be used as a valid measuring tool (Slamet & Wahyuningsih, 2022). To create a more accurate instrument, the validity of the four-tier diagnostic questions also needs to be measured at tier 3 which can be seen in Table 3.5.

Table 3.5 Aiken Test for Tier 3 Validation Result

Item	Expert			S_1	S_2	S_3	∑s	n(c-1)	V	Validity
	I	II	III							
Q1	1	1	1	1	1	1	3	3	1.00	High
Q2	1	1	1	1	1	1	3	3	1.00	High
Q3	1	1	1	1	1	1	3	3	1.00	High
Q4	1	1	1	1	1	1	3	3	1.00	High
Q5	0	1	1	0	1	1	2	3	0.67	Moderate
Q6	1	0	1	1	1	1	3	3	1.00	High
Q7	1	1	1	0	0	1	1	3	0.33	Low
Q8	1	0	0	1	0	0	1	3	0.33	Low
Q9	1	0	1	1	0	1	2	3	0.67	Moderate
Q10	1	0	1	1	0	1	2	3	0.67	Moderate
Q11	1	0	1	1	0	1	2	3	0.67	Moderate
Q12	1	0	1	1	0	1	2	3	0.67	Moderate
Q13	1	1	1	1	1	1	3	3	1.00	High
Q14	1	1	1	1	1	1	3	3	1.00	High
Q15	1	1	1	1	1	1	3	3	1.00	High
Q16	1	0	1	1	0	1	2	3	0.67	Moderate
Q17	0	0	1	0	0	1	1	3	0.33	Low
Q18	1	0	1	1	0	1	2	3	0.67	Moderate
Q19	1	1	1	1	1	1	3	3	1.00	High
Q20	1	1	1	1	1	1	3	3	1.00	High
Q21	1	0	0	1	0	0	1	3	0.33	Low
Q22	1	0	1	1	0	1	2	3	0.67	Moderate
Q23	1	0	1	1	0	1	2	3	0.67	Moderate
Q24	1	1	0	1	1	0	2	3	0.67	Moderate
Q25	1	0	0	1	0	0	1	3	0.33	Low
Q26	1	1	1	1	1	1	3	3	1.00	High
Q27	1	1	0	1	1	0	2	3	0.67	Moderate

Item	Expert			S_1	S_2	S_3	$\sum s$	n(c-1)	V	Validity
	I	II	III							
Q28	1	0	0	1	0	0	1	3	0.33	Low

In Table 3.5, it can also be seen that the average of content validity of Tier 3 is 0.73, which indicates moderately valid. The expert judgment forms from the three experts are attached in Appendix A.3. The validity of the questions was assessed through a correlation test to ensure their suitability. This test examines the relationship between two variables without examining the influenced variables (Astuti, 2017). The questions were distributed to grade 8 students twice due to low validity values in the first test involving 170 studentsThe questions were revised to enhance their validity by providing more information about the challenges faced by students. After revise the questions, the questions were distributed to 83 students in different schools which can be seen in Appendix A.5. The result of validity which used the questions as can be seen in Appendix A.2 held by the students is described in Table 3.6.

Table 3.6 Result of Validity Test

Item	Tier	Test I			Test II		
		Pearson Correlation	Interpretation	Decision	Pearson Correlation	Interpretation	Decision
1	1	-0.043	Not Valid	Revised and Retest	0.628	Valid	Used
	3	-0.099	Not Valid		0.246	Valid	
2	1	-0.026	Not Valid		0.288	Valid	
	3	0.02	Not Valid		0.471	Valid	
3	1	0.13	Not Valid		0.421	Valid	
	3	-0.051	Not Valid		0.413	Valid	
4	1	-0.047	Not Valid		0.535	Valid	
	3	0.082	Not Valid		0.681	Valid	
5	1	0.108	Not Valid		0.332	Valid	
	3	0.183	Valid		0.543	Valid	
6	1	0.233	Valid		0.343	Valid	
	3	0.136	Not Valid		0.399	Valid	
7	1	0.144	Not Valid		0.438	Valid	
	3	0.034	Not Valid		0.282	Valid	

Item	Tier	Test I			Test II		
		Pearson Correlation	Interpretation	Decision	Pearson Correlation	Interpretation	Decision
8	1	0.088	Not Valid		0.226	Valid	
	3	-0.052	Not Valid		0.423	Valid	
9	1	0.08	Not Valid		0.336	Valid	
	3	-0.004	Not Valid		0.509	Valid	
10	1	0.175	Valid		0.541	Valid	
	3	0.185	Valid		0.251	Valid	
11	1	0.023	Not Valid		0.275	Valid	
	3	0.062	Not Valid		0.573	Valid	
12	1	-0.032	Not Valid		0.501	Valid	
	3	0.055	Not Valid		0.352	Valid	
13	1	0.105	Not Valid		0.429	Valid	
	3	0.043	Not Valid		0.387	Valid	
14	1	0.054	Not Valid		0.703	Valid	
	3	0.024	Not Valid		0.687	Valid	
15	1	-0.02	Not Valid		0.525	Valid	
	3	0.014	Not Valid		0.748	Valid	
16	1	0.225	Valid		0.452	Valid	
	3	0.037	Not Valid		0.535	Valid	
17	1	-0.015	Not Valid		0.559	Valid	
	3	0.018	Not Valid		0.442	Valid	
18	1	0.121	Not Valid		0.528	Valid	
	3	0.121	Not Valid		0.375	Valid	
19	1	-0.107	Not Valid		0.593	Valid	
	3	0.007	Not Valid		0.586	Valid	
20	1	0.254	Valid		0.355	Valid	
	3	0.063	Not Valid		0.605	Valid	
21	1	0.06	Not Valid		0.46	Valid	
	3	0.105	Not Valid		-0.248	Not Valid	
22	1	-0.005	Not Valid		-0.099	Not Valid	Not Used
	3	-0.046	Not Valid		0.433	Valid	
23	1	0.05	Not Valid		0.017	Not Valid	
	3	0.09	Not Valid		0.663	Valid	

Item	Tier	Test I			Test II		
		Pearson Correlation	Interpretation	Decision	Pearson Correlation	Interpretation	Decision
24	1	0.026	Not Valid		0.607	Valid	Used
	3	0.123	Not Valid		0.672	Valid	
25	1	0.252	Valid		0.421	Valid	Not used
	3	0.067	Not Valid		-0.174	Not Valid	
26	1	0.141	Not Valid		0.477	Valid	
	3	0.009	Not Valid		0.496	Valid	
27	1	0.067	Not Valid		0.452	Valid	Used
	3	0.016	Not Valid		0.609	Valid	
28	1	-0.006	Not Valid		0.448	Valid	
	3	-0.101			0.538		

In the first test, question item number 10 was the only valid question in the validity test. This certainly shows that the question is not feasible. Apart from that, question number 10 has low reliability, so it still needs to be revised. A low reliability value indicates that the question cannot be relied on for consistency when the test is carried out (Sanaky, 2021). Then the second test was carried out. Of the 28 questions tested, 24 were valid. The questions then were used as the real instrument tool to assess misconceptions on human circulation, and can be seen detailed in Appendix B.1. The test is said to be valid when the R-Count value is greater than the R-Table value. Table R is a table used to determine the correlation coefficient from data analysis that has significant values (Aiken, 1980). The result of validity test is more detail can be found in Appendix A.7 which compared to the value of r-Table also can be seen in Appendix A.4.

3.3.3 Reliability Test

Apart from validity testing, it is necessary to carry out a reliability test before conducting actual research. Reliability testing is carried out to see how consistently the test can be carried out. High test reliability can be seen through the consistency of the results produced through several tests to find out how much variability occurs when there is an error in measuring the score as it should be (Setiyawan, 2014). The results of the reliability test can be seen in Table 3.7

Table 3.7 Reliability Test on Second Validity Test Result

Test	Tier	N of Item	Cronbach's Alpha
II	1	24	0.882
	3	24	0.897

Reliability test is only carried out on valid questions. This is in accordance with the consistency and stability of the questions which have question validity criteria, where if the questions are not valid, then the consistency of the questions is considered to be reduced. In the second validity test, the reliability value obtained at tier 1 was 0.882, and at tier 3 it was 0.897. To see the more detail of the reliability results, it can be seen in Appendix A.6. Instruments in these two tiers are said to have very high correlation coefficient values because they are in the range of $0.80<n \leq 1.00$ (Riyani et al., 2017). Thus, it can be said that this instrument has high consistency of questions so that it is suitable for use in real research as a tool to assess or diagnose misconceptions on the topic of the human circulatory system.

3.4 Research Procedure

To complete this research, the researcher divided the steps into three parts, namely the preparation stage which is carried out before the real research is carried out, the implementation stage when the real instrument is used, and the completion stage which is used to complete the data based on the results obtained at the implementation stage.

1. Preparation Stage
   a. Identify research problem
   b. Analyze the topic of Human Circulatory System contained in Science Textbook for Grade 8 in *Kurikulum Merdeka* to be developed in preliminary study
   c. Make the open-ended questions in scientific knowledge and reason form for preliminary study with two representative questions of each concept
   d. Validate the questions to the supervisor
   e. Distribute the preliminary study instrument to students whose have learnt the topic
   f. Collect and generalize the results of preliminary study

g. Make and develop the four-tier diagnostic test instrument from the preliminary study results
h. Validate and revise the questions based on feedback given by supervisor
i. Validate and revise the questions by 3 expert judgements (2 lecturers and 1 biology teacher)
j. Distribute the four-tier diagnostic test instruments to students whose have learnt the topic
k. Conduct the validity and reliability test based on students' answer by using SPSS
l. Revise instruments that are declared invalid and eliminate the questions cannot be used for further research
m. Final the validation of instrument whose have studied the topic of the human circulatory system
n. Check the correlation and reliability test of final validation test by using SPSS.

2. Implementation stage
   a. Contact the school and providing a letter of permission to conduct research from the campus, the letters were provided in Appendix C.1 and Appendix C.2.
   b. Make an appointment regarding research procurement
   c. Conduct research on students with the main instrument being an online form and providing print-outs for students who cannot access the internet which was conducted in March 2024, the documentation can be seen in Appendix B.2.
   d. Conduct interviews as additional research through chat with teachers regarding the research results.
3. Completion stage
   a. The data obtained were analyzed statistically to diagnose the misconceptions using the certainty answer index can be seen in Appendix D.1 and Appendix D.2.
   b. Made the discussion based on the findings obtained in the research and interpret the data
   c. Drew conclusions, implications, and recommendations based on the results of the data analysis
   d. Validate the data results to the supervisor
   e. Create and report the research paper thesis
   f. Upload the research to Sinta 3 and can be seen in Appendix C.3.

The research sequence can be seen in Figure 3.1.

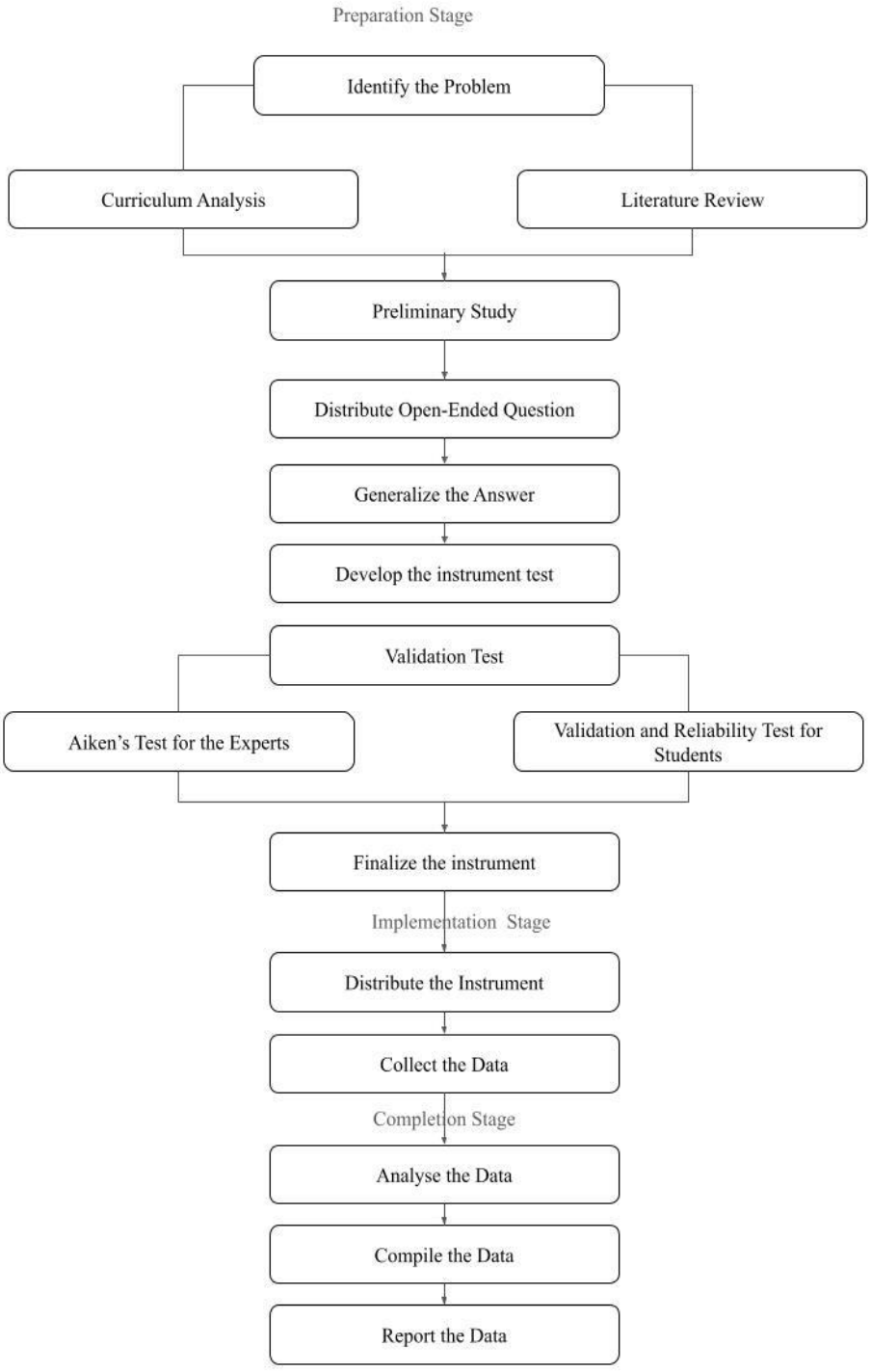


Figure 3.1 The Flowchart of the Research Procedure

3.5 Data Analysis

The answers are categorized as Scientific Knowledge (SK), False Positive (FP), False Negative (FN), Misconceptions (M), and Lack of Knowledge (LK). Scientific knowledge is the condition that student have a good understanding both tiers and they sure about their answers. The false positive is the condition when students are confident about their answer to the first-tier, but they have confidence in a false answer at the third-tier. In reverse, a false negative is the condition when the students have confidence but answer incorrectly in first-tier, while they answer the third-tier correctly with confidence. The misconception is the condition when they have confidence in wrong answers for tiers. Lack of knowledge is the condition when they are not confident in answering the first or the third-tiers, or even both tiers. The consideration scale comparison of decisions of four-tier test is given in Table 3.8.

Table 3.8 The Decisions of Four-Tier Test (Kiray&Simsek, 2021)

Tier 1	Tier 2	Tier 3	Tier 4	Decision
True	Confident	True	Confident	SK
True	Confident	False	Confident	FP
False	Confident	True	Confident	FN
False	Confident	False	Confident	M
True	Confident	True	Not Confident	LK 1
True	Not Confident	True	Confident	LK 2
True	Not Confident	True	Not Confident	LK 3
True	Not Confident	False	Not Confident	LK 4
True	Not Confident	False	Confident	LK 5
True	Not Confident	False	Not Confident	LK 6
False	Confident	True	Not Confident	LK 7
False	Not Confident	True	Confident	LK 8
False	Not Confident	True	Not Confident	LK 9
False	Confident	False	Not Confident	LK 10
False	Not Confident	False	Confident	LK 11

Tier 1	Tier 2	Tier 3	Tier 4	Decision
False	Not Confident	False	Not Confident	LK 12

In analyzing students' answer, Microsoft Excel is used. Answers are coded "1" as the score for a correct answer, and "0" for a false answer. If the students are sure or confident with the answer, it will be scored as "1", and if not confident will be "0". So, the code of scientific knowledge will be (1-1-1-1); "False Positive", the first-tier will be coded as "1", while the third-tier will be coded as "0". So, the code of false positive will be (1-1-0-1); "False Negative", the first-tier will be coded as "0", while the third-tier will be coded as "1". So, the code of false positive will be (0-1-1-1); "Misconception", is the false answer of first and third-tier. So, the code will be (1-0-1-0); "Lack of Knowledge", the sequence of the code is based on the answer. The true answer will be coded as "1" and the false will be "0".