CHAPTER III
RESEARCH METHODOLOGY

3.1 Research Design

This research used quantitative methods specially, a cross-sectional survey design. Quantitative methods emphasize collecting data in the form of exact numbers through instruments and explaining it with statistical tests. Research that used planned structures, that are clear from beginning to the end of research and not influenced by conditions in the field. A cross-sectional survey design is procedure in quantitative research in which researchers collect data by using questioners or interviews to sample of the population to be studied (Creswell, 2015).

The purpose of this survey research is to described the attitudes, opinions, behaviors, or characteristic of the student related to the research topic namely to diagnose students' conception of the students related to topic classification of living things. There was no intervention or treatment given to the participant in this research, so the researchers only observe an describe the students' conceptions in data collected.

3.2 Population and Sample

The respondent of this research is the student on Public School eighth-grade student aged 13 – 14 years from the East Bandung Region, West Java. The student who had studied the classification of living things in new system of the national curriculum. In Research Division, this study divided a Bandung Region into three regions. There are West Bandung, Central Bandung, and East Bandung. Based on the agreement assessment division, this study got East Bandung to taken a data instrument. As part of the data collection process, the research sites by compiling a list of junior high schools located in the East Bandung Regency as presented in Appendix B.1. To complement the quantitative data, documentation of research

activities was also carried out during the implementation phase as documented in Appendix B.2. The division can be seen in Figure 3.1.

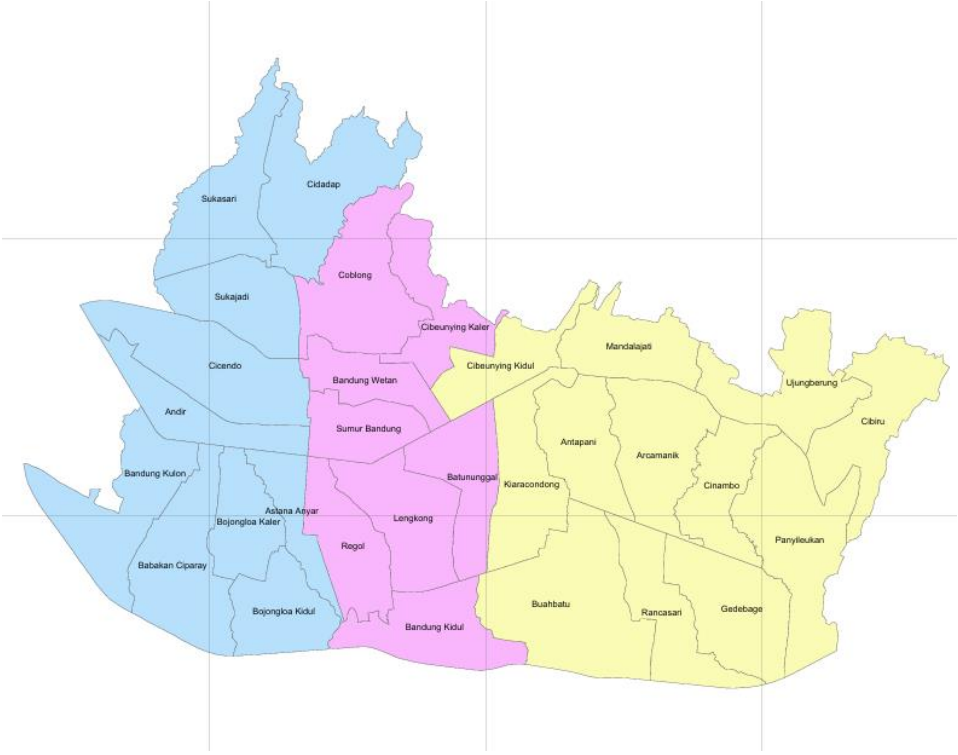

Figure 3. 1: Bandung City Region Maps

The researchers conducted used Indonesian language to comply with school regulations and effective communication. This researcher used convenience sampling. Convenience sampling is sample-selected technique regarding the willingness and availability of sample to take part of the research. It's a nonprobability sampling technique which requires little effort, cost, time, and simple operation (Golzar & Noor, 2022).

3.3 Research Instrument

This instrument used in this research is a questionnaire containing four-tier diagnostic test. This test items are made in the form of four options with the aims of limiting students' opportunities to answer carelessly and not according to the existing context and concepts. This test items consist 14 multiple-choice questions from a preliminary study, which underwent a through validation phase. Eleven questions were selected after validation to make up the final set of question for the

real test instrument. The researcher used several concepts developed from the subtopics can be seen in Table 3.1.

Table 3. 1 The Distribution of Question

No	Sub-Topics	Item Questions
1	Living or Non-Living Things?	Q1, Q2, Q3, Q4, Q5
2	Why are the Living Thing categorized?	Q6, Q7, Q8, Q9
3	Living Things are Diverse	Q10, Q11

These concepts were tested in this research as they were taught to the students as part of their learning curriculum. The purpose of the four-tier diagnostic test question is determining how well the applicants understand these concepts. By using trap solutions- answer that seem plausible to student but are different from the correct answer, the researcher can accurately determine the misconceptions held by the students. Responses from students are expected to differ according to the level of understanding and any misconceptions. Researcher can analyze the frequency of particular misunderstandings and determine the rationale behind students' incorrect answer. An example question from the four-tier diagnostic test instruments can be seen in Table 3.2.

Table 3. 2 A Sample Question

Item	Tier	Question
Q7	1	Based on how living things are classified, which of the following statements is correct? A. In the past, jellyfish were classified as fish because they can swim and their habitat is in the sea. However, now jellyfish are classified as vertebrates because they can swim and do not have a spine. B. In the past, whales were classified as fish because they can swim and their habitat is the sea. However, now whales are classified as mammals because in addition to being able to swim and also give birth (breastfeeding). C. In the past, algae were classified as plants because they can photosynthesize. However, now algae are classified as Monera because in addition to photosynthesizing, but there are no roots, stems and leaves.

Item	Tier	Question
		D. In the past, fungi were classified as plants because they have cell walls. However, now fungi are classified as protists because in addition to having cell walls, they cannot produce their own food.
	2	Are you sure about your answer? A. Yes B. No
	3	What is the reason for your answer? A. Because there was a mistake in the previous taxonomy (classification system of living things). B. Because there are always new characteristics of a species, so taxonomy (the classification system of living things) must be updated. C. Because taxonomists do not have an agreement in taxonomy (classification system of living things). D. Because of the development of technology used taxonomy (classification system of living things) allows the classification of living things based on the characteristics - more specific characteristics.
	4	Are you sure about your answer? A. Yes B. No

In the four-tier diagnostic test structure, the first-tier is to measure students understanding of the topic, while the third-tier is to choose the best reasoning for their answer in the first-tier question. In the second-tier and fourth tier is measure their confidence of their answer in the first-tier and third-tier questions.

3.3.1 Preliminary Study

A preliminary study is the initial stage in research and development of research model. Data from the preliminary is used to verify the validity and reliability of the questionnaire items prior to commencing the main research. This topic, a preliminary study used in the form of open questions in a two-tier form. The first-tier consists of questions aimed at evaluating students' misunderstanding

of the topic. The second-tier consists the reasoning behind the given replies. The open-ended questions used in the preliminary study can be seen in Table 3.3

Table 3. 3 Preliminary Question

Item	Tier	Question
Q1	1	Unlike humans and animals, as living things, the movement of plants cannot be clearly observed by us. How do plants move?
	2	What is the reason for your answer?
Q2	1	Take a look at some sample images below.  Sumber: blogspot.com  Sumber: blog.studyhouse.id Based on the two pictures above, what characteristics of living things are shown?
	2	What is the reason for your answer?
Q3	1	Write down the possible ways of reproduction by living things and their definitions briefly!
	2	What is the reason for your answer?
Q4	1	Write down what is meant by living things being sensitive to stimuli and their examples in everyday life!
	2	What is the reason for your answer?
Q5	1	One of the characteristics of living things is that they can take and use energy. How do humans and animals differ from plants in obtaining energy for their needs?
	2	What is the reason for your answer?
Q6	1	Explain the following statement: “All living things breathe.” Is this statement true?
	2	What is the reason for your answer?
Q7	1	What are the remains of living things?
	2	What is the reason for your answer?
Q8	1	Observe the following picture!

Item	Tier	Question
		 Based on the picture above, which is the correct difference between animal cells and plant cells, below?
	2	What is the reason for your answer?
Q9	1	Name the order of levels in classifying living things from smallest to largest!
	2	What is the reason for your answer?
Q10	1	How is the order of classification (from broadest to narrowest category) used to identify organisms? Does the order of classification always correctly identify organisms?
	2	What is the reason for your answer?
Q11	1	Make a dichotomy key to identify the following animal species: 1. Cat 2. Dog 3. Cow 4. Goat 5. Chicken
	2	What is the reason for your answer?
Q12	1	Make a table format determination key to identify the following plant species: 1. Mango 2. Apple 3. Water guava 4. Coconut 5. Rice
	2	What is the reason for your answer?
Q13	1	Give an example for each taxon level in plants and animals!
	2	What is the reason for your answer?
Q14	1	Give an example of each kingdom and explain its characteristics!
	2	What is the reason for your answer?

In this stage, there 35 students at eighth grade had studied the classification of living things topic. From the 14 questions given, a variety of answer were collected then simplified and grouped to become the basis for preparing multiple-choice questions in a four-tier diagnostic test. The preliminary study instruments used for this research as listed in Appendix A.1

3.3.2 Validation Test

Validation is assessed by examining the alignment between the test items and the underlying construct, ensure that the items measure the specific concepts intended by the research (Avinç & Doğan, 2024). Validation is used to state the degree of accuracy of a measuring instrument. The Rasch Model provides several indicators of construct validity, that assess whether individual items behave as expected within the model's framework (Stone et al., 2022).

The formula for calculating content validity is the Aiken V coefficient, which is determined by giving a number from 0 to 1 (not representative to very representative) (Rahmawati & Harun, 2019). The following is the Aiken V formula:

$$V = \frac{\sum s}{n\,(c-1)}$$

V = Rater's Fit Index

s = Average Score – The Lowest Score in The Category

c = Number of Categories

n = Total or Number of Experts

The validation result of the Aiken Formula was used to analysis each item from three expert (2 lecturer and a teacher) to determine the content validity of an advanced multiple-choice instrument designed to assess students' critical thinking skills. The expert judgement forms from the three experts can be seen in Appendix A.3. This formula calculates the content validity coefficient based on the number of validators and the extent to which an item represents the intended indicator (Bhakti et al., 2024).

Table 3. 4 Description for Validity using The Aiken Index

Interval	Description
$V > 0.8$	High Validity
$0.4 < V \leq 0.8$	Moderate Validity
$V \leq 0.4$	Invalid

(Irawan & Wilujeng, 2020)

The item score is considered valid if the analysis the analysis results show that each question given has a value greater than V coefficient value, which ranges from 0.4 to 1 according to the number of validators. If the data can be considered valid, it can be used. Meanwhile, the data considered invalid, it cannot be used. Therefore, it is crucial to re-study and even recreate the items in each evaluation component and carry-out in-depth analysis and validation of the material (Hidayah, 2023). The result of the validity of the instrument using Aiken Test can be seen in Tabel 3.5 and 3.6. For detail, the instrument test that validity test as provided in Appendix A.2.

Table 3. 5 Aiken Test for Tier 1 Validation Result

Item	Expert			S1	S2	S3	∑s	n(c-1)	V	Validity
	I	II	III							
Q1	1	1	1	1	1	1	3	3	1.00	High
Q2	0	1	1	0	1	1	2	3	0.67	Moderate
Q3	0	1	1	0	1	1	2	3	0.67	Moderate
Q4	0	1	1	0	1	1	2	3	0.67	Moderate
Q5	0	1	1	0	1	1	2	3	0.67	Moderate
Q6	0	1	1	0	1	1	2	3	0.67	Moderate
Q7	0	1	1	0	1	1	2	3	0.67	Moderate
Q8	0	1	1	0	1	1	2	3	0.67	Moderate
Q9	0	1	1	0	1	1	2	3	0.67	Moderate
Q10	0	1	1	0	1	1	2	3	0.67	Moderate
Q11	1	1	1	1	1	1	3	3	1.00	High
Q12	0	1	1	0	1	1	2	3	0.67	Moderate
Q13	0	1	1	0	1	1	2	3	0.67	Moderate
Q14	0	1	1	0	1	1	2	3	0.67	Moderate

In Table 3.5 in above. The V coefficient value for each test item ranges from 0.4 to 1, with an average value of 0.71 for 14 questions at the first and third tiers, indicating good indicating. These instruments test is suitable for measuring the desired results.

Table 3. 6 Aiken Test for Tier 3 Validation Result

Item	Expert			S1	S2	S3	∑s	n(c-1)	V	Validity
	I	II	III							
Q1	1	1	1	1	1	1	3	3	1.00	High
Q2	0	1	1	0	1	1	2	3	0.67	Moderate
Q3	0	1	1	0	1	1	2	3	0.67	Moderate

Item	Expert			S1	S2	S3	∑s	n(c-1)	V	Validity
	I	II	III							
Q4	0	1	1	0	1	1	2	3	0.67	Moderate
Q5	0	1	1	0	1	1	2	3	0.67	Moderate
Q6	0	1	1	0	1	1	2	3	0.67	Moderate
Q7	0	1	1	0	1	1	2	3	0.67	Moderate
Q8	0	1	1	0	1	1	2	3	0.67	Moderate
Q9	0	1	1	0	1	1	2	3	0.67	Moderate
Q10	0	1	1	0	1	1	2	3	0.67	Moderate
Q11	1	1	1	1	1	1	3	3	1.00	High
Q12	0	1	1	0	1	1	2	3	0.67	Moderate
Q13	0	1	1	0	1	1	2	3	0.67	Moderate
Q14	0	1	1	0	1	1	2	3	0.67	Moderate

In Table 3.6 in above. The V coefficient value for each test item ranges from 0.4 to 1, with an average value of 0.71 for 14 questions at the first and third tiers, indicating good indicating. These instruments test is suitable for measuring the desired results.

The research test instrument needs to be validated to get the accuracy between the learning outcome score data and the learning objectives the Rasch Model (Darmana et al., 2021). There is condition that the instrument is valid can be seen in Table 3.7.

Table 3. 7 Description Value Rasch Model

Description	Interval
Outfit Mean Square (MNSQ)	0.5 <MNSQ <1.5
Outfit z-standardized (ZTSD)	-0.2 <ZTSD <+2.0
Correlation (CORR)	0.4 <CORR <0.85

(Bond & Fox, 2013)

The validity test is carried out so that researchers know the instruments used can measure the variables to be measured. This study delivered to the 49 students at eight-grade. Test the validity of this instrument using Ministep Rasch Model software computer. The result of the validity of the instrument carried out by student can be seen in Table 3.8 and 3.9.

Table 3. 8 The First Tier 1 Validation Result

Item	MNSQ	ZSTD	CORR	Interpretation	Decision
Q1	1.13	0.53	0.23	Valid	Used

Item	MNSQ	ZSTD	CORR	Interpretation	Decision
Q2	1.11	0.49	0.40	Valid	Used
Q3	1.18	0.84	0.32	Valid	Used
Q4	0.65	-1.48	0.57	Valid	Used
Q5	1.06	0.30	0.37	Valid	Used
Q6	1.03	0.23	0.44	Valid	Used
Q7	0.90	-0.40	0.43	Valid	Used
Q8	0.78	-0.48	0.33	Not Valid	Not Used
Q9	0.77	-0.78	0.47	Valid	Used
Q10	1.02	0.16	0.41	Valid	Used
Q11	1.12	0.40	0.28	Valid	Used
Q12	0.97	-0.15	0.42	Valid	Used
Q13	1.16	0.77	0.28	Valid	Used
Q14	1.02	0.20	0.17	Valid	Used

In Table 3.8 in above. There were 14 items tested, of which 13 items were valid and declared feasible to be use, while one item (Q8) was declared invalid because the correlation value was below the standard.

Table 3. 9 The First Tier 3 Validation Result

Item	MNSQ	ZSTD	CORR	Interpretation	Decision
Q1	1.00	0.04	0.33	Valid	Used
Q2	0.95	-0.03	0.22	Valid	Used
Q3	1.01	0.10	0.32	Valid	Used
Q4	1.04	0.22	0.28	Valid	Used
Q5	0.45	-0.99	0.44	Not Valid	Not Used
Q6	1.58	1.12	-0.09	Not Valid	Not Used
Q7	0.76	-1.27	0.53	Valid	Used
Q8	0.82	-0.58	0.43	Not Valid	Not Used
Q9	1.12	0.64	0.21	Valid	Used
Q10	0.92	-0.63	0.41	Valid	Used
Q11	1.12	0.99	0.21	Valid	Used
Q12	1.05	0.34	0.23	Valid	Used
Q13	0.96	-0.20	0.39	Not Valid	Revised and Retest
Q14	0.95	-0.30	0.37	Not Valid	Revised and Retest

In Table 3.9 in above. There were 14 items tested, of which 9 items were valid and declared feasible to be use, while 5 item was declared invalid because the correlation value was below the standard. Because the researcher focusses on

subtopics where each subtopic should be minimum valid is 3 questions. The targeted four-tier diagnostic test that used still less, then for 2 items (Q13 and Q14) revised and then conducting to distribute the four-tier diagnostic test to students. The result of the validity of the instrument carried out by student can be seen in Table 3.10 and 3.11.

Table 3. 10 The Second Tier 1 Validation Result

Item	MNSQ	ZSTD	CORR	Interpretation	Decision
Q13	1.00	0.15	0.84	Valid	Used
Q14	1.00	0.15	0.55	Valid	Used

In Table 3.10 in above. The results of Tier 1 validation using the Ministep Rasch Model, two items (Q13 and Q14) have been analyzed to determine their validity.

Table 3. 11 The Second Tier 3 Validation Result

Item	MNSQ	ZSTD	CORR	Interpretation	Decision
Q13	1.00	0.00	0.77	Valid	Used
Q14	1.00	0.00	0.77	Valid	Used

In Table 3.11 in above. The results of Tier 3 validation using the Ministep Rasch Model, two items (Q13 and Q14) have been analyzed to determine their validity. The analysis results show that both items are valid and suitable for use in the assessment instrument. This validity indicates that the items have met the set criteria and can be used in the next stage to accurately measure student understanding. The question then was as the real instrument tool to assess misconception can be found in Appendix A.4

3.3.3 Reliability Test

Reliability test focuses on evaluating the consistency and accuracy of data in the context of measuring models, such as tests or questionnaires. Person reliability measurers the consistency of responses from individuals, while item

reliability evaluates the internal consistency of the test items themselves (Stone et al., 2022). Reliability test carried out that researchers obtain information about the consistency of items in providing measurement results.

These statistics provide insight into the ability of the test items to distinguish between different levels of respondent abilities and the reproducibility of item calibration across different samples. The reliability values have range between 0 and 1, with higher values indicate greater reliability. High reliability values (≥ 0,8) indicate that the test consistently measures the intended construct. There is condition that the instrument is valid can be seen in Table 3.12.

Table 3. 12 Description Value Rasch Model

Category	Interval
Very Low	< 0.200
Low	0.200 – 0.399
Moderate	0.400 – 0.599
High	0.600 – 0.799

(Bond & Fox, 2013)

Test reliability only used for valid instruments. If the question is invalid, its consistency will be considered less. The result of the reliability test on valid instrument can be seen in Table 3.13 and 3.14.

Table 3. 13 Person Reliability Result

Tier	Person Reliability	Category
1	0.33	Low
3	0.00	Very Low

In Table 3.13 in above. The results of the person reliability scores for both tiers were categorized as very low. This shows that the consistency of answers between respondents is still weak, so there may be high variability in understanding or interpretation of the questions given.

Table 3. 14 Item Reliability Result

Tier	Item Reliability	Category
1	0.91	Very High
3	0.85	Very High

In Table 3.14 in above. The results show that the reliability of the items in both categories is very high. These results indicate that the instrument used has good item quality and is consistent in measuring the concepts tested, although the level of reliability from the respondent side still needs to be improved.

In the Rasch Model, the probability of an individual answering an item correctly depends on individual ability and item difficulty. Higher ability individuals have the highest probability of answering all items, while easier items have a high probability of getting correct answers from all respondents when the difference between item difficulty and individual ability is monitored (Bond & Fox, 2013).

3.4 Research Procedure

The researcher divided the steps into three parts to complete this research. There is 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 the research problem.
    b. Analyse the topic of Classification of Living Things that contained in Science Textbook for Grade 7 in new system of the national curriculum to be developed in preliminary study.
    c. Make an open-ended question in scientific knowledge form for preliminary study.
    d. Make an open-ended question in scientific knowledge form for preliminary study.
    e. Validate the question of preliminary study to the supervisor.
    f. Distribute the preliminary study instrument to students that have already learned the topic.
    g. Collect and generalize the results of preliminary study.

h. Make a multiple-choice question on scientific knowledge and open-ended questions in reason from for preliminary study.
i. Make and develop the four-tier diagnostic test instrument form the preliminary study results.
j. Validate and revise the questions based on feedback given by supervisor.
k. Validate and revise the questions by 3 expert judgements (2 lecturers and 1 science teacher).
l. Distribute the four-tier diagnostic test instruments to students who have already learn the topic.
m. Conduct the validity and reliability tests based on students' answers by using Ministep Rasch.
n. Revise instruments items that are declared invalid and eliminate the questions that cannot be used for further research.
o. Final the validation of instrument that have already learn the topic.
p. Check the validation and reliability test of final validation test by using Ministep Rasch

2. Implementation Stage
    a. Contact the school and provide letter of permission to conduct research from the campus to conduct research. Appendix C.1 and Appendix C.2 provided the letters.
    b. Make an appointment regarding research procurement.
    c. Conduct research on students with the main instrument being an online form.
    d. Collect and analyze the data to determine the conception of the student and determine the concept of classification of living things topic.
3. Completion Stage
    a. The data are analyzed statistically to diagnose the misconceptions of students.
    b. Make a discussion based on the findings the result of the data.

c. Make a conclusions, implications, and recommendations based on the results of the data analysis.
d. Validate the data result to the supervisor.
e. Create and report the research paper thesis.
f. Put all the data evidence in the Appendix.
g. Upload the journal based on the research paper thesis to Sinta 3.

From the preparation stage to the completion phase, a systematic process is followed to strengthen the accuracy of the instrument. This process includes pilot testing, interviews, and statistical evaluations, all of which contribute not only to the validation of the instrument but also to a deeper analysis of students' conceptual grasp. The study's findings are expected to play a crucial role in science education, assisting teachers in formulating effective teaching strategies that enhance students' understanding especially of biology.

Detailed research documentation assures a precise report of the methods and authorization procedure of the research. Results and methodology can be replicated with assessment and utilitarian application possible within educational frameworks and areas.

In the last stage, proof of publication in a journal is shown in Appendix D.4. Moreover, before submission of the thesis, a plagiarism test is taken, and the similarity index report is attached in Appendix D.5.

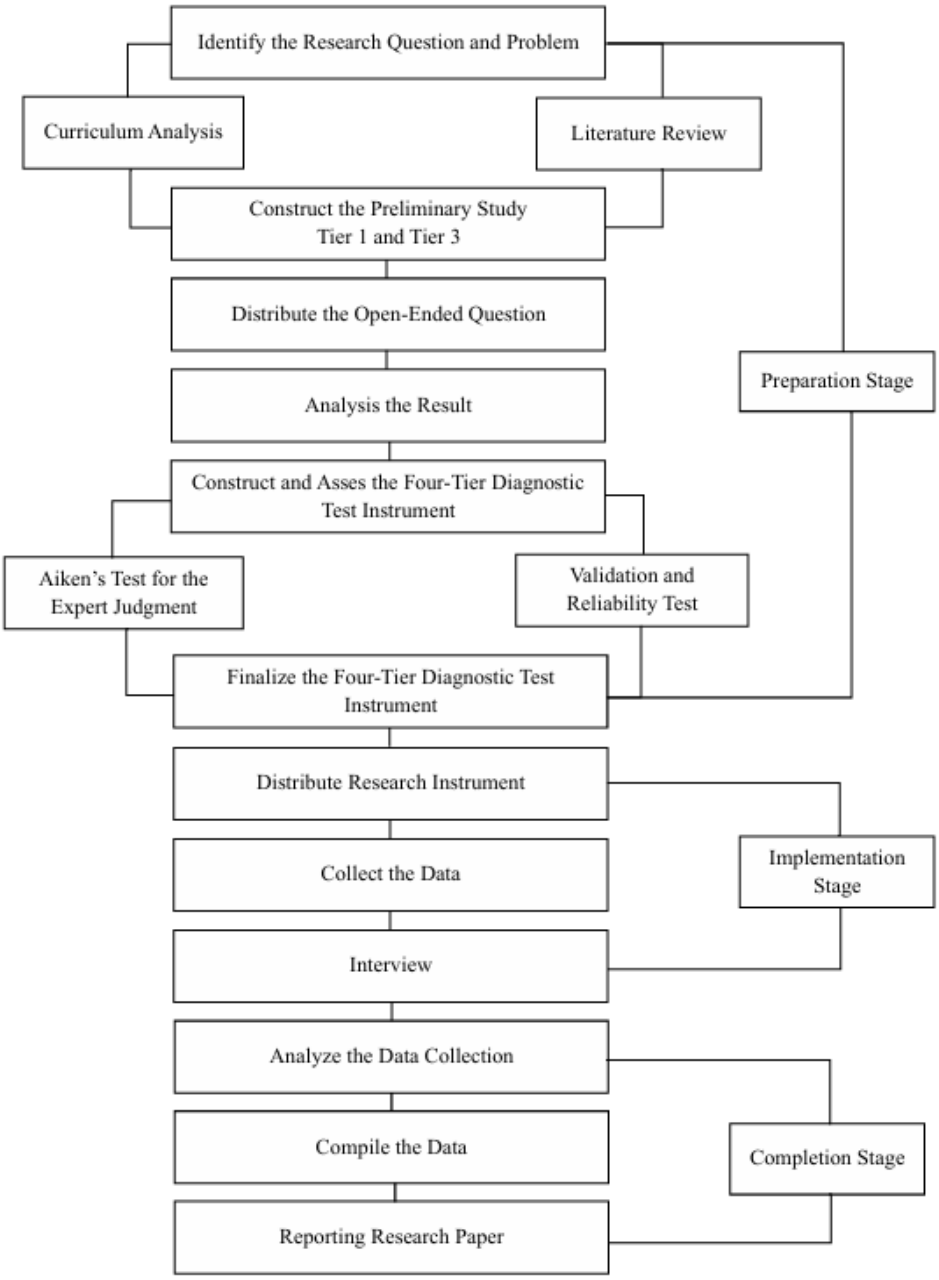

Figure 3. 2: The Research Stage

3.5 Data Analysis

These techniques include the examination of test instruments by material expert and evaluation experts' judgement (Negoro & Karina, 2019). 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 where a student's answers show a correct understanding both tiers with confidence. The false positive is the condition where a student's answers show a confidence their correctly in first-tier, but the student answer confidence their incorrectly in third-tier. The false tier is the condition where student's answers show a confidence their incorrectly in first-tier, but where a student answer show a confidence their correctly in third-tier. The misconception is the condition where a student's answers show a confidence their incorrectly both tiers. The lack of knowledge is the condition where a student's answer shows a not confidence in first-tier or third-tier, or even both tiers. Specific answers will result in various concepts decisions, so the data can be analyzed easily (Kiray & Simsek, 2021) can be seen Table 3.15.

Table 3. 15 Comparison of Decision of Four-Tier Diagnostic Test

Tier 1	Tier 2	Tier 3	Tier 4	Decision of Four-Tier Diagnostic Test
True	Confident	True	Confidence	SK
True	Confident	False	Confident	FP
False	Confident	True	Confident	FN
False	Confident	False	Confident	M
True	Not Confident	True	Not Confident	LK 1
True	Not Confident	False	Confident	LK 2
True	Not Confident	False	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
False	Not Confident	False	Not Confident	LK 12

(Kiray & Simsek, 2021)

Student responses in four-tier diagnostic test can be easily coded. Answers are coded “1” as the score for a correct answer, and “0” for the incorrect answer. If the students are confident answer, is scored as “1”, and if not confident, answer will be “0”. The coded system is following below:

1. Scientific Knowledge: Correctly answer in first-tier and third-tier and both of confident answer in second-tier and fourth tier. The code will be (1-1-1-1).
2. False Positive: Correctly answer in first-tier, but incorrectly answer in third-tier. Both of confident answer in second-tier and fourth tier. The code will be (1-1-0-1).
3. False Negative: Incorrectly answer in first-tier, but correctly answer in third-tier. Both of confident answer in second-tier and fourth tier. The code will be (0-1-1-1).
4. Misconceptions: Incorrectly answer in first-tier and third-tier both of confident answer in second-tier and fourth tier. The code will be (0-1-0-1).
5. Lack of Knowledge: The sequence of the code is based on the answer. The answer of the students that’s not fit into previous code and it categorize into lack of knowledge.