CHAPTER III
RESEARCH METHOD

3.1 Research Design

This research uses a quantitative approach with a cross-sectional survey design. This design was chosen because it is suitable for collecting data at a certain time to describe and explain the phenomena that occur in the field. Survey research is a method used to collect quantitative data from a population or sample through the use of instruments such as questionnaires or tests, with the aim of describing attitudes, opinions, behaviors, or certain characteristics (Creswell, 2015). In cross-sectional survey, data are collected at one particular point in time, and not longitudinally, with a view to provide a quick and efficient snapshot of the respondent's state or perception at that moment. This design was considered suitable to the research questions, which were focused on the measurement of students' conceptual knowledge and misconceptions about Earth and Solar System content. Under this strategy, the researcher assessed the trends of the students' responses within a relatively short period and investigated their interrelations of the other variables under investigation. With a four-level diagnostic test tool, the cross-sectional design facilitated the collection of precise information on students' knowledge, reasoning, and confidence to know.

However, it is important to note that cross-sectional surveys offer only a snapshot view of the population at a specific point in time, much like a photograph that captures what is present in a single moment but does not reveal changes or developments beyond that moment (Capili, 2021). Because all variables were measured simultaneously, the research was limited to identifying tendencies and associations among variables rather than establishing causal relationships (Wang & Cheng, 2020). Despite this limitation, the cross-sectional approach remained effective for capturing students' conceptual profiles and identifying the prevalence of misconceptions within the defined population at the time of data collection.

Zalfa Julia Rosyani, 2025
FOUR-TIER DIAGNOSTIC TEST AS A TOOL TO ASSESS STUDENTS MISCONCEPTION ABOUT EARTH AND SOLAR SYSTEM IN EAST BANDUNG REGION
Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

18

3.2 Participant

This research was conducted at a junior high school in East Bandung region. This research focused on eighth grade students who had previously studied about the Earth and the solar system. This research also included students who had taken science education under the Merdeka Curriculum. The division of research areas can be seen in Figure 3.1.
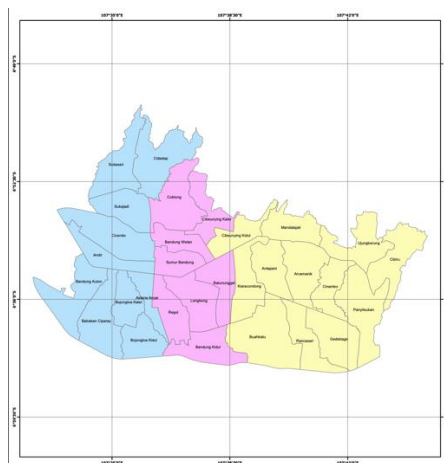

Figure 3.1: Region Maps of Bandung City

(Regrouped from Zonation Maps for PPDB 2019 in Bandung City, source:
https://ppdbkotabandung.wordpress.com/2019/04/20/4-zonasi-pada-ppdb-kota-bandung-2019)

The 2021 Assessment Research Division conducted a zoning classification of schools in Bandung City, resulting in three main areas, as shown in Figure 3.1. One of these areas is East Bandung, which has a total of 24 junior high schools. However, only 18 schools were selected as respondents in this study, as they met the established qualifications. The sample consisted of 547 junior high school students in the East Bandung area of West Java Province. Convenience sampling, as described by (Rahi, 2017), was used to select the sample, which involves collecting data from a population that is easily accessible to the researcher.

3.3 Research Instrument

The diagnostic test model questionnaire used in this investigation has four tier. It used a three-tier diagnostic test and was separated into two halves. The first focused on the students' thinking skills, while the second examined why they

answered in particular ways. To reduce the likelihood of students providing thoughtless responses or forgetting certain contexts and concepts, each question included four acceptable answers. These questions were initially developed in pilot research and subsequently converted into the diagnostic four-tier to better identify any student misconceptions. These questions were based on several of the sub-topic concepts listed in Table 3.1. This was done as part of a careful effort to make the assessment relevant.

Table 3.1 The Distribution of Questions

Sub-Topics	Concept	Item Question
Solar System	Eight Planets in the Solar System	Q1, Q2, Q3, Q4
	Other Celestial Bodies	Q5, Q6, Q7
Earth and Satellite	Earth's Movement in the Solar System	Q8, Q9
	The Moon as a Satellite of the Earth	Q10
	Earth's satellites other than the Moon	Q11
Get To Know The Sun More Closely	Characteristics of Sun	Q12
	Solar Eclipse	Q13
	The Role of the Sun in Human Life	Q14

This research focuses on eight concepts divided into three main subtopics: Solar System, Earth and Satellite, and Get to Know the Sun More Closely. At the initial stage, 14 questions were developed as part of the preliminary study. Furthermore, each question was developed using a four-tier diagnostic test. Examples of questions designed using the four-tier diagnostic test format can be seen in Table 3.2.
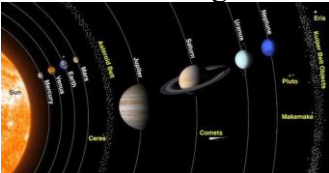
Table 3.2 A Sample Question

Item	Tier	Question
Q13	**1**	Take a look at the picture below!

		(sumber :https://www.cnnindonesia.com/gaya-hidup/20150106162342-269-22728/berwisata-sambil-melihat-gerhana-matahari-di-indonesia, 2016)
		What phenomenon occurs in the image?
		  A. Total solar eclipse
		  B. Total lunar eclipse
		  C. Partial lunar eclipse
		  D. Partial solar eclipse

	2	Are you sure of your answer?
		  A. Sure
		  B. Not Sure

	3	Why did you choose the answers in tier 1?
		  A. The sun is between the moon and the earth, blocking all sunlight from reaching the earth.
		  B. The moon is between the earth and the sun, blocking all of the sun's light from reaching the earth.
		  C. The moon moves in front of the sun, blocking some of the sun's light from reaching the moon.
		  D. The earth moves in front of the sun, blocking some of the sun's light from reaching the moon.

	4	Are you sure of your answer?
		  A. Sure
		  B. Not Sure

3.3.1 Preliminary Study

The preliminary study is the first step in this research process. This stage aims to analyze the problem as a reference in the implementation of subsequent research. On this topic, the preliminary study was conducted using open-ended questions in the form of two levels. The first level consists of questions to assess students' understanding of the topic, while the second level asks for reasons for the answers given. Examples of open-ended questions used in the preliminary study can be seen in Table 3.3.
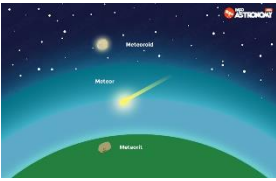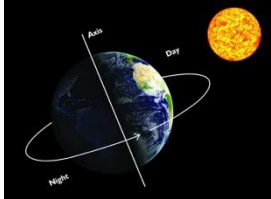
Table 3.3 Preliminary Question
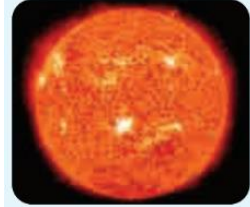
Item	Part	Question
Q1	1	Observe the image of the Milky Way galaxy below.

Item	Part	Question
		The image above shows the grouping of planets in the Milky Way galaxy. Scientists categorize these planets into two groups, namely inner planets and outer planets. What is outer planets and inner planets?
	2	Give a reason of your previous answer!
Q2	1	Observe the following image! Venus is the hottest planet in the solar system. Why is Venus considered the hottest planet?
	2	Give a reason of your previous answer!
Q3	1	The atmosphere is a layer of gas that forms the outermost layer of a planet. What is the dominant gas that makes up Earth's atmosphere?
	2	Give a reason of your previous answer!
Q4	1	All planets in the Milky Way galaxy revolve around the sun due to the sun's gravitational force. What is meant by gravitational force?
	2	Give a reason of your previous answer!
Q5	1	Take a look at the following picture below. (Source: Grid.id, 2021) The moon is a satellite of the earth. What is a satellite?
	2	Give a reason of your previous answer!
Q6	1	Take a look at the following image below. (Source: Astronomy Info, 2019) What is the difference between a meteoroid, meteor and meteorite?
	2	Give a reason of your previous answer!
Q7	1	Take a look at the picture below.

Item	Part	Question
		 (Source: Kompas.com, 2023) What is the difference between dwarf planets and terrestrial planets?
	2	Give a reason of your previous answer!
Q8	1	What is the rotational and revolutionary motion of the Earth?
	2	Give a reason of your previous answer!
Q9	1	Take a look at the picture below. (Source: Kumparan.com, 2020) Why is there an alternation between day and night?
	2	Give a reason of your previous answer!
Q10	1	Take a look at the image below! (source : solarsystem.nasa.gov/Bill Dunford, 2018) The picture shows the phases of the moon. What is a moon phase?
	2	Give a reason of your previous answer!
Q11	1	A satellite is any celestial body that revolves around a generally larger celestial body. The Moon, Titan, or Io are natural satellites that orbit planets. Apart from natural satellites, there are also artificial satellites. Explain what an artificial satellite is!
	2	Give a reason of your previous answer!
Q12	1	Take a look at the picture below! (sumber : Pearson Heinemann/Wendy Gorton, 2009)

Item	Part	Question
		The sun has several characteristics that can be observed with specialized solar telescopes from Earth, one of which is sunspots. What are sunspots?
	2	Give a reason of your previous answer!
Q13	1	Take a look at the picture below! (sumber : Pearson Heinemann/Wendy Gorton, 2009) The picture above shows the types of solar eclipses. Based on the picture above, how can a solar eclipse occur?
	2	Give a reason of your previous answer!
Q14	1	There is no creature on Earth that does not need the Sun, even creatures that even those that live in cold environments. The Sun plays a very important role in life. One of them is in fossil fuels. How does solar energy play a role in fossil fuels?
	2	Give a reason of your previous answer!

This preliminary study phase involved 33 grade 8 students who had studied material about the earth and solar system. From the 14 questions given, various answers were obtained, then generalized and grouped to become the basis for preparing multiple choice questions for the four tier diagnostic test. At this stage, some answers were found to contain misconceptions, even random answers. Some of the factors that cause suboptimal results include students who forget the concepts they have learned and students who are less serious in doing the test. The generalized forms of the questions and corresponding answers are provided in Appendix A.1.

3.3.2 Validation Test

After the questions are developed into a four-tier diagnostic test, the results are validated. The test used for validation test is using Ministep. In the validation process, Ministep aids in identifying how well each test item performs based on various criteria such as MNSQ (Mean-Square Fit Statistic), ZSTD (Z-Standardized Fit Statistic), and CORR (Correlation). By analyzing these parameters, educators

can determine whether the items are functioning as intended and whether they accurately measure students' knowledge or skills. This process is crucial for ensuring that assessments are valid and reliable, ultimately leading to better educational outcomes. Based on the results obtained, logical validity was assessed through expert judgment involving three assessors, consisting of two lecturers and one science teacher. The validation process employed the Aiken's V formula, as presented below:

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

With the description,

V = Rater's Fit Index

s = Average score – the lowest score in the category

c = Number of Categories

n = Number of Raters

According to Fajaruddin et al. (2021), the validity of a test can be classified as low or less valid if the Aiken's V value is less than or equal to 0.4 (0 < V ≤ 0.4), moderately valid if it falls between greater than 0.4 and up to 0.8 (0.4 < V ≤ 0.8), and highly valid if the value ranges from above 0.8 to less than 1 (0.8 < V < 1). The results of the test item validation, as determined through expert judgment, are presented in Table 3.4.

Table 3.4 Aiken Test for Tier 1 Validation Test 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	0	1	1	0	1	2	3	0.67	Moderate
Q4	1	0	1	1	0	1	2	3	0.67	Moderate
Q5	1	0	1	1	0	1	2	3	0.67	Moderate
Q6	1	1	1	1	1	1	3	3	1.00	High
Q7	1	1	1	1	1	1	3	3	1.00	High
Q8	1	0	1	1	0	1	2	3	0.67	Moderate
Q9	1	1	1	1	1	1	3	3	1.00	High

Q10	1	1	1	1	1	1	3	3	1.00	High
Q11	1	1	1	1	1	1	3	3	1.00	High
Q12	1	1	1	1	1	1	3	3	1.00	High
Q13	1	1	1	1	1	1	3	3	1.00	High
Q14	1	1	1	1	1	1	3	3	1.00	High

Based on Table 3.4, out of the 14 Tier 1 items developed, 10 items demonstrate high validity, while the remaining 4 fall under the category of moderate validity. This indicates that all Tier 1 items meet the criteria for validity and are therefore considered acceptable for use in the study. To further assess the validity of the Tier 3 items, refer to the data presented in Table 3.5.

Table 3.5 Aiken Test for Tier 3 Validation Test 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	0	1	1	0	1	2	3	0.67	Moderate
Q4	1	1	1	1	1	1	3	3	1.00	High
Q5	1	0	1	1	0	1	2	3	0.67	Moderate
Q6	1	1	1	1	1	1	3	3	1.00	High
Q7	1	1	1	1	1	1	3	3	1.00	High
Q8	1	0	1	1	0	1	2	3	0.67	Moderate
Q9	1	1	1	1	1	1	3	3	1.00	High
Q10	1	1	1	1	1	1	3	3	1.00	High
Q11	1	0	1	1	0	1	2	3	0.67	Moderate
Q12	1	1	1	1	1	1	3	3	1.00	High
Q13	1	1	1	1	1	1	3	3	1.00	High
Q14	1	1	1	1	1	1	3	3	1.00	High

Table 3.5 shows that all Tier 3 items are also classified as valid. Specifically, 10 items meet the high validity criteria, while 4 items fall into the moderate category. Therefore, based on the Aiken index analysis for both Tier 1 and Tier 3, it can be concluded that all test items are valid and suitable for identifying student misconceptions related to the solar system topic. The expert judgment forms from

the three experts are attached in Appendix A.3. Following the expert judgment process, 14 sets of instruments were finalized, as presented in Appendix A.4, and subsequently subjected to a validity test. Based on the data analysis conducted using Ministep, the findings are presented in Table 3.6.

Table 3.6 First Validation for Tier 1

Item Tier	MNSQ	ZSTD	CORR	Interpretation	Desicion
Q1	1.20	0.55	0.07	Valid	Used
Q2	0.80	-0.47	0.40	Valid	Used
Q3	0.77	-0.85	0.42	Valid	Used
Q4	1.34	1.23	-0.02	Valid	Used
Q5	1.05	0.28	0.27	Valid	Used
Q6	0.86	-0.81	0.42	Valid	Used
Q7	0.92	-0.61	0.43	Valid	Used
Q8	1.00	0.03	0.29	Valid	Used
Q9	1.14	1.27	0.15	Valid	Used
Q10	1.03	0.30	0.28	Valid	Used
Q11	0.92	-0.57	0.44	Not Valid	Revised
Q12	1.06	0.38	0.22	Valid	Used
Q13	1.02	0.15	0.28	Valid	Used
Q14	1.13	0.49	0.21	Valid	Used

Table above evaluates 14 items (Q1 to Q14) in Tier 1 based on three metrics: MNSQ (Mean-Square Fit Statistic), ZSTD (Z-Standardized Fit Statistic), and CORR (Correlation). Each item is interpreted as either "Valid" or "Not Valid," determining whether it is "Used" or "Not Used" in subsequent analyses. Items with appropriate MNSQ and ZSTD values and meaningful CORR values are marked "Valid" and "Used." Most items (Q1–Q10, Q12–Q14) are considered valid, while one item (Q11) is deemed "Not Valid" and is excluded. This indicates an overall strong fit of most items for the validation criteria. Furthermore, to examine the validity test results for Tier 3, refer to the data presented in Table 3.7.

Table 3.7 First Validation for Tier 3

Item Tier	MNSQ	ZSTD	CORR	Interpretation	Desicion
Q1	1.66	1.22	0.06	Not Valid	Not Used
Q2	0.70	-0.69	0.32	Valid	Used

Q3	0.90	-0.22	0.28	Valid	Used
Q4	1.07	0.36	0.13	Valid	Used
Q5	0.75	-1.45	0.47	Valid	Used
Q6	1.05	0.39	0.29	Valid	Used
Q7	1.20	1.78	0.09	Not Valid	Not Used
Q8	1.02	0.24	0.27	Valid	Used
Q9	1.05	0.50	0.28	Valid	Used
Q10	1.02	0.23	0.30	Valid	Used
Q11	0.94	-0.39	0.39	Valid	Used
Q12	0.94	-0.29	0.39	Not Valid	Revised
Q13	1.11	0.45	0.31	Valid	Used
Q14	0.88	-0.13	0.49	Not Valid	Revised

Table 3.7 shows the initial validation results for 14 items in Tier 3 (Q1–Q14) based on three parameters: MNSQ (Mean Square Error), ZSTD (Z-Standardized Score), and CORR (item-total correlation. However, some items (Q1, Q12, and Q14) were invalid due to excessively high or low MNSQ values and low CORR, and therefore were excluded from further analysis.

Table 3.8 Second Validation for Tier 1

Item Tier	MNSQ	ZSTD	CORR	Interpretation	Desicion
Q11	0.90	-0.58	0.70	Valid	Used
Q12	0.84	-1.17	0.73	Valid	Used
Q14	1.26	1.84	0.56	Valid	Used

Table 3.8 presents the second validation results for items Q11, Q12, and Q14 in Tier 1. Each item is evaluated using MNSQ, ZSTD, and CORR values. All three items meet the criteria for validity, as reflected in their "Valid" interpretation and "Used" decision. Notably, Q11, initially marked as "Not Valid" in the first validation, is now considered valid with improved correlation (0.70). Similarly, Q12 and Q14 maintain their valid status, with stronger correlation values of 0.73 and 0.56, respectively. This indicates these items are suitable for further use after validation.

Table 3.9 Second Validation for Tier 3

Item Tier	MNSQ	ZSTD	CORR	Interpretation	Desicion
Q11	1.07	0.43	0.58	Valid	Used
Q12	0.95	-0.30	0.65	Valid	Used
Q14	0.97	-0.06	0.68	Valid	Used

Table 3.9 shows the second validation values for items Q11, Q12, and Q14 in Tier 3. All items are "Valid" and "Used," and their MNSQ, ZSTD, and CORR values are all within acceptable values. The CORR values for these items are better than those of the first validation, with Q11 at 0.58, Q12 at 0.65, and Q14 at 0.68, which is indicative of greater correspondence with the validation criteria. This suggests that these items are now always appropriate for Tier 3 analyses.

3.3.3 Reliability Test

Reliability testing attempts to determine the stability of a measuring instrument, whether it produces consistent and reliable outcomes when the measurement is repeated (Hazlita et al., 2015). Reliability testing helps to screen and uncover any problems with the instrument before using it on a larger scale. If there is a problem on the stability or consistency side of test results, then the researcher can make necessary revisions or enhancements before the instrument is employed for complete research (Sullivan & Artino Jr, 2013). The result can be seen in table 3.10.

Table 3.10 The Reliability of Tier 1 and Tier 3

Tier	Item Reliability	Criteria
1	0.88	Very High
3	0.92	Very High

Table 3.10 presents the reliability results for Tier 1 and Tier 3, calculated using the Ministep method. The reliability values for both tiers are classified as "Very High," with Tier 1 achieving an item reliability score of 0.88 and Tier 3 scoring 0.92. These high reliability values indicate strong internal consistency and suggest that the items in both tiers are highly dependable for measuring the intended

constructs. The higher reliability in Tier 3 implies slightly greater stability and precision in its items compared to Tier 1.

3.4 Research Procedure

1. Preparation Stage

   a. Identify the research problem.

   b. Analyze the topic of the Earth and Solar System in the Grade 7 Science Textbook under the existing curriculum for preliminary study development.

   c. Develop open-ended questions focusing on scientific knowledge and reasoning, including representative questions for each concept.

   d. Validate the questions with a supervisor.

   e. Distribute the preliminary study instruments.

   f. Collect and summarize the results of the preliminary study.

   g. Develop a four-tier diagnostic test instrument based on the results of the preliminary study.

   h. Validate and revise the questions using feedback from the supervisor.

   i. Validate the instrument to three experts (two lecturers and one physics teacher).

   j. Revise the instrument based on validation given by expert.

   k. Conduct validity test to student whose already learn the topic.

   l. Analyze the result by using Ministep.

   m. Revise invalid instruments and eliminate questions unsuitable for further research.

   n. Conduct second validity test for revised question.

   o. Analyze the result by using Ministep.

   p. Perform correlation and reliability testing on the finalized instrument using Ministep.

2. Implementation Stage

   a. Contact the school in East Bandung region and provide a letter of permission to conduct the research. The letter provided and can be seen in Appendix C.1 and Appendix C.2

    b. Schedule a meeting regarding research procedures.

    c. Conduct research using google form as the primary instrument.

    d. Conduct interviews with teachers and students through chat to gain insights into the findings. The data collection documentation can be found in Appendix B.2.

3. Completion Stage

    a. Analyze the collected data statistically using Microsoft excel to diagnose misconceptions in each subtopic.

    b. Interpret the findings and create a discussion based on the research results.

    c. Draw conclusions and recommendations based on the data analysis.

    d. Validate the research results with the supervisor.

    e. Make the report the research paper thesis.

    f. Upload the result to Sinta 3 Journal, the proof can be seen in Appendix C.3
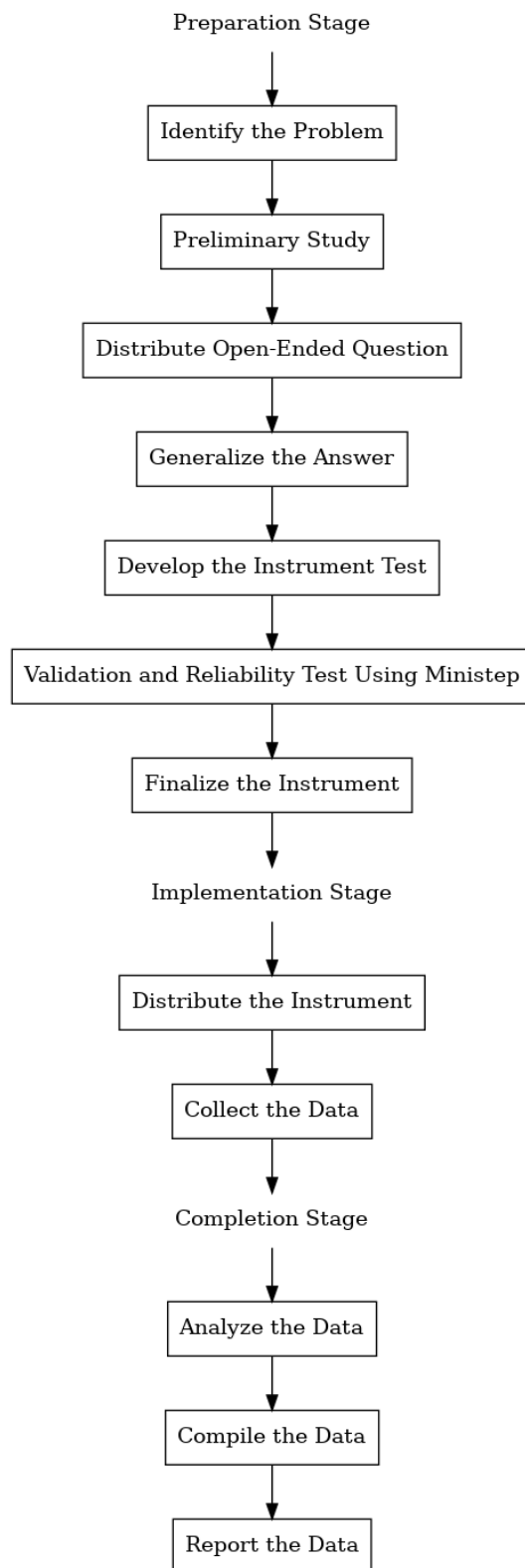
The research sequences can be seen in Figure 3.2.

Preparation Stage

↓

Identify the Problem

↓

Preliminary Study

↓

Distribute Open-Ended Question

↓

Generalize the Answer

↓

Develop the Instrument Test

↓

Validation and Reliability Test Using Ministep

↓

Finalize the Instrument

↓

Implementation Stage

↓

Distribute the Instrument

↓

Collect the Data

↓

Completion Stage

↓

Analyze the Data

↓

Compile the Data

↓

Report the Data

Figure 3.2: The Flowchart of the Research Procedure

3.5 Data Analysis

These answers fall into categories such as Scientific Knowledge (SK), False Positive (FP), False Negative (FN), Misconceptions (M), and Lack of Knowledge (LK). The case of Scientific Knowledge would be when the students derived an excellent understanding of both tiers and had confidence in their answers. In the case of False Positive, students confidently give wrong answers to any one of the first-tier answers but give correct answers to the third tier with confidence. False Negative: In the first tier, the student's answer was wrong, but at the third tier, the student answers correctly with full confidence. Misconception: The student was confidently wrong in the two tiers. Lack of Knowledge: The student had no confidence and did not know his or her answer to any of the questions in both tiers. The Analysis of student's conception is available in Appendix B.2. Table 3.11 presents the comparison of the decision scales for the four-tier test.

Table 3.11 The Decisions of Four-Tier Test (Kiray & Simsek, 2021)

Tier 1	Tier 2	Tier 3	Tier 4	Decicion
True	Confident	True	Confident	SK
True	Confident	False	Confident	FP
False	Confident	True	Confident	FN
False	Confident	False	Confident	M
True	Confident	False	Not Confident	LK 1
True	Not Confident	False	Confident	LK 2
True	Not Confident	False	Not Confident	LK 3
True	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	True	Not Confident	LK 10
False	Not Confident	True	Confident	LK 11
False	Not Confident	True	Not Confident	LK 12

In this study, a coding system is used to classify students' responses to the four-tier diagnostic test. The coding criteria are as follows:

1. Scientific Knowledge (SK): Both the first and third tiers have correct answers coded as "1," and the second and fourth tiers are answered confidently, also coded as "1." The code for SK is (1-1-1-1).

2. False Positive (FP): The first tier has a correct answer, but the third tier is incorrect. Despite this, the student confidently answered both tiers. The code for FP is (1-1-0-1).

3. False Negative (FN): The first tier is incorrect, but the third tier is correct. The student answered both tiers confidently, resulting in the code (0-1-1-1).

4. Misconception (M): Both the first and third tiers have incorrect answers, and the student answered confidently for both. The code for M is (0-1-0-1).

5. Lack of Knowledge (LK): If a student's response does not match any of the other categories, it is classified as lack of knowledge.