CHAPTER III
RESEARCH METHOD

3.1 Research Design

The research design used in this research is survey. Survey design is one of the procedures in quantitative research procedure in which a researcher surveys the target sample and population to examine a population's attitudes, opinions, behavior, or characteristics. In survey research design, researchers collect data using questionnaire and interview. Interview were conducted to find out the causes of misconceptions that occurred in students. In the survey research design, the respondents were not given treatment by the researchers (Creswell, 2002). The research design used was in accordance with the objectives of this research, namely to identify the students' conception of diffusion and osmosis topics because the researchers did not provide treatment to the samples involved and the researchers only described the students' actual conception.

In this research, cross-sectional survey design types were used. Cross-sectional survey design means that the researcher only collects data from a sample at one time, even though all data collection lasts from days to several weeks (Creswell, 2002). Using a cross-sectional survey design, researchers can obtain information about current attitudes, beliefs, opinions, or practices in a brief period, such as the time needed to conduct the survey and gather the data. In this research, the questions used were four-tier test questions regarding diffusion and osmosis given to samples using the Google Forms. The instrument used in this research can be used to measure students' conceptions. The data collection process was carried out for one week. The data that has been obtained is then analyzed to find out how the students' conception regarding the topic of diffusion and osmosis.

3.2 Participants

Respondents involved in this research were students who learn in private junior high schools. The characteristics of students who became the research sample were students who studied diffusion and osmosis under the Cambridge Curriculum. All students in a school are the population of this research. There were 119 students

involved in this research consisting of 64 secondary three students and 55 secondary one students with an age range of 13-15 years. All the respondents involved came from a private junior high school in West Bandung Regency. The distribution of the research samples involved can be seen in Table 3.1.

Table 3.1 Participant Distribution

Grade	Number of Students	Percentage (%)
Secondary 1	55	46.2
Secondary 3	64	53.8
Total	119	100

The researcher chose participants from grade secondary one and grade secondary three because the participant requirements required students who had studied the concepts of osmosis and diffusion. This concept is taught in grade one, so the researcher intends to identify the extent to which students' understanding is consistent. In addition, researchers also want to investigate whether there are significant differences between the two classes. First graders are just learning this concept in the near future, while third graders have been studying this topic for a long time before the test.

Researchers used a convenience sampling technique to determine the sample involved in this research. This technique selects respondents based on ease of accessibility and closeness to the researcher. In convenience sampling, readily available and easy-to-reach respondents are included in the research without following a specific randomization process (Creswell, 2002). Respondents were collected until the researcher got the number of respondents needed to represent the population. Then the students' answer will be analyzed and categorized based on the rubric previously made in the research (Kiray & Simsek, 2021).

3.3 Research Instrument

To achieve the objectives of this research, a four-tier diagnostic test instrument was used on the concept of diffusion and osmosis. The instruments that have been

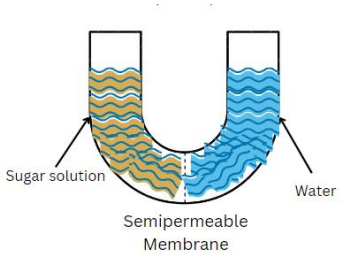
made are given to the research sample. Students are asked to work on the questions according to their abilities. A total of 14 four-tier multiple-choice questions are made. There are four concepts in the questions, namely the process of diffusion, the process of osmosis, concentration, and factors that influence the diffusion process. The distribution of questions and the concept can be seen in Table 3.2.

Table 3.2 Question Distribution

Topic	Sub Topic	Test Item
Diffusion and Osmosis	Process Of Diffusion	1, 7, 8, 9, 12
	Process Of Osmosis	2, 3, 11, 14
	Concentration	5, 6
	Factors That Influence Diffusion Process	4, 10, 13

Each question consists of four-tier questions. The first tier is a multiple-choice question to find out students' knowledge which has four options consisting of one correct answer option and three wrong answer options. The second tier is a multiple-choice question to find out the confidence level of students in answering questions in the first-tier questions, there are two multiple choice answer options, namely sure and not sure. The third tier is a question to determine the scientific reason for student answers. In the first tier of questions, four answer options contain one correct answer option and three wrong answer options. The fourth tier is a multiple-choice question to determine students' confidence level in answering questions. In the third-tier question, there are two multiple-choice answer options: sure and not sure. An example of the questions given to students can be seen in Table 3.3. Thirty questions have been made then will be tested and analyzed.

Table 3.3 Sample Question

No	Tier	Question
3	1st	In the process of osmosis, water moves from a lower concentration area to a higher concentration area. What will happen to the volume in the area of the sugar solution on the left side?  A. The volume will increase B. The volume will decrease C. The volume does not change D. The volume will be balanced
	2nd	Are you sure about your answer? A. Sure B. Not Sure
	3rd	What is your reason to choose that answer? A. Water molecules will move from a high concentration to a low concentration area B. Water molecules will move from a low concentration to a high concentration area C. There is no movement of molecules so the volume does not change D. The volume of solutions with high and low concentrations will always be balanced
	4th	Are you sure about your reason? A. Sure

No	Tier	Question
		B. Not Sure
	Scientific Concept	In osmosis process, solvent molecules (usually water) move across a selectively permeable membrane from an area of lower solute concentration to an area of higher solute concentration. So, the volume of sugar solution that have higher comcentration will be increase.

The question in Table 3.3 is given to students to determine their understanding of the movement of molecules during the osmosis process. In the question, students are given a picture to analyze whether there is a change in volume in one part of an area. In the osmosis process, solvent molecules (usually water) will move to a region with a higher concentration, namely sugar solution, because it has more solute (sugar) than water. So, the volume in the area containing the sugar solution will increase because the water molecules have moved. Students are then asked to select the reasoning option that supports their answer to the first-tier question. They are asked to choose whether they are confident or unsure in having the answer and reasoning in the first and third-tier questions.

Before the instrument is distributed to a predetermined sample, the instrument goes through two stages of development. There were 30 test items regarding diffusion and osmosis given to 32 students in the first development stage and 22 test items given to 72 students in the second development stage using the Google Form (see Appendix D.1) and written answer forms. The results of student answers were then analyzed using Microsoft Excel and SPSS applications to assess the validity of the test items and would be categorized as valid and invalid test items. Test items that are categorized as invalid will be eliminated from this research.

3.3.1 Validation Test

Before the instrument can be used in research, the instrument goes through the validation stage twice. The validation test is used to ensure that the instrument can measure what it is supposed to measure (Creswell, 2002). Both tier one and tier

three must be included in the valid category (Caleon & Subramaniam, 2010). The results of the validity test can be seen in Table 3.3. The validity test used is Item-analysis with Corrected item-total correlation. Item analysis through corrected total item correlation helps identify items that are uncorrelated on the overall test and measure different dispositions or traits. The results of the corrected item-total correlation analysis show that all items show a correlation with the overall test score (Coskun & Kara, 2019) .To find out the validity of the instrument to be used as a measuring tool, it can be done by comparing the value of the r-table (see Appendix B.5) and the calculated r-count (corrected item-total correlation). Suppose the value of corrected total item correlation (r-count) is higher than r-table. In that case, the test item significantly correlates with the total score, so the test items will be considered valid. Suppose the value of corrected total item correlation (r-count) is lower than the r-table. In that case, the item items are not significantly correlated with the total score, so the test item will be considered invalid. Furthermore, to determine the value of the r table, we can use the formula Df = n-2, where n is the number of samples used. For example, if the number of samples is 20 people, the df value will be 18. Furthermore, we can find the r table value corresponding to a df of 18 and a significance level of 5%.

Based on the information contained in Table 3.4. In the first testing phase, which involved 32 respondents, from 30 question that tested (see in Appendix A.1) four questions (Q15, Q16, Q21, and Q24) were declared valid. This means that the four questions meet the validity criteria so that they can be directly used for the real test and do not need to be included in the second testing stage (see in Appendix B.1). Then, four questions (Q1, Q11, Q19, and Q22) are eliminated or cannot be used in the second testing stage and real test, so they are not included in Table 3.3. This is due to the negative r value, indicating that the questions are invalid. Therefore, the four questions must be eliminated from the instrument to be used. In addition, the remaining questions were categorized as invalid. However, these questions were not eliminated immediately but were revised based on the suggestions of expert judgment by adding pictures to the questions, as well as

improving vocabulary and grammar in the questions and multiple choices. After this revision, the improved questions were distributed again to 72 respondents. In the second testing stage involving 72 respondents, the results were checked using a validation test (see Appendix B.2). From the second testing phase, 10 more question are valid. So, the total 14 questions were valid and suitable for use as instruments in the real test (see in Appendix A.2).

Table 3.4 The Result of Validation Test

Question	Tier	Test I (N=32)				Test II (N=72)			
		R Count	R Table (5%)	Interpretation	Decision	R Count	R Table (5%)	Interpretation	Decision
1	1	0.463	0.361	Valid	Rejected				
	3	-0.030	0.361	Not Valid					
2	1	0.223	0.361	Not Valid	Retest	0.372	0.235	Valid	Not Used
	3	0.000	0.361	Not Valid		-0.003	0.235	Not Valid	
3	1	0.359	0.361	Not Valid	Retest	0.394	0.235	Valid	Used
	3	0.434	0.361	Valid		0.421	0.235	Valid	
4	1	0.229	0.361	Not Valid	Retest	0.354	0.235	Valid	Not Used
	3	0.030	0.361	Not Valid		0.230	0.235	Not Valid	
5	1	0.125	0.361	Not Valid	Retest	0.425	0.235	Valid	Used
	3	0.342	0.361	Not Valid		0.280	0.235	Valid	
6	1	0.212	0.361	Not Valid	Retest	0.401	0.235	Valid	Not Used
	3	0.454	0.361	Valid		0.226	0.235	Not Valid	
7	1	0.507	0.361	Valid	Retest	0.100	0.235	Not Valid	Not Used
	3	0.038	0.361	Not Valid		0.404	0.235	Valid	
8	1	0.180	0.361	Not Valid	Retest	0.236	0.235	Valid	Used
	3	0.282	0.361	Not Valid		0.422	0.235	Valid	
9	1	0.268	0.361	Not Valid	Retest	0.340	0.235	Valid	Used
	3	0.691	0.361	Valid		0.266	0.235	Valid	
10	1	0.448	0.361	Valid	Retest	0.050	0.235	Not Valid	Not Used
	3	0.274	0.361	Not Valid		0.461	0.235	Valid	
11	1	-0.150	0.361	Not Valid	Rejected				

Question	Tier	Test I (N=32)				Test II (N=72)			
		R Count	R Table (5%)	Interpretation	Decision	R Count	R Table (5%)	Interpretation	Decision
	3	0.427	0.361	Valid					
12	1	0.101	0.361	Not Valid	Retest	-0.014	0.235	Not Valid	Not Used
	3	0.235	0.361	Not Valid		0.230	0.235	Not Valid	
13	1	0.234	0.361	Not Valid	Retest	0.351	0.235	Valid	Used
	3	0.390	0.361	Valid		0.537	0.235	Valid	
14	1	0.195	0.361	Not Valid	Retest	0.441	0.235	Valid	Used
	3	0.512	0.361	Valid		0.264	0.235	Valid	
15	1	0.541	0.361	Valid	Directly Used				
	3	0.644	0.361	Valid					
16	1	0.482	0.361	Valid	Directly Used				
	3	0.396	0.361	Valid					
17	1	0.469	0.361	Valid	Retest	0.288	0.235	Valid	Used
	3	0.183	0.361	Not Valid		0.243	0.235	Valid	
18	1	0.073	0.361	Not Valid	Retest	0.416	0.235	Valid	Not Used
	3	0.353	0.361	Not Valid		0.217	0.235	Not Valid	
19	1	-0.288	0.361	Not Valid	Rejected				
	3	0.083	0.361	Not Valid					
20	1	0.303	0.361	Not Valid	Retest	0.120	0.235	Not Valid	Not Used
	3	0.053	0.361	Not Valid		0.216	0.235	Not Valid	
21	1	0.476	0.361	Valid	Directly Used				
	3	0.396	0.361	Valid					
22	1	-0.091	0.361	Not Valid	Rejected				
	3	0.145	0.361	Not Valid					
23	1	0.313	0.361	Not Valid	Retest	0.178	0.235	Not Valid	Not Used
	3	0.393	0.361	Valid		0.144	0.235	Not Valid	
24	1	0.408	0.361	Valid	Directly Used				
	3	0.484	0.361	Valid					

Question	Tier	Test I (N=32)				Test II (N=72)			
		R Count	R Table (5%)	Interpretation	Decision	R Count	R Table (5%)	Interpretation	Decision
25	1	0.650	0.361	Valid	Retest	0.352	0.235	Valid	Used
	3	0.085	0.361	Not Valid		0.381	0.235	Valid	
26	1	0.087	0.361	Not Valid	Retest	0.371	0.235	Valid	Used
	3	0.169	0.361	Not Valid		0.431	0.235	Valid	
27	1	0.477	0.361	Valid	Retest	0.334	0.235	Valid	Not Used
	3	0.240	0.361	Not Valid		0.210	0.235	Not Valid	
28	1	0.216	0.361	Not Valid	Retest	0.197	0.235	Not Valid	Not Used
	3	0.388	0.361	Valid		0.232	0.235	Not Valid	
29	1	0.007	0.361	Not Valid	Retest	0.336	0.235	Valid	Not Used
	3	0.150	0.361	Not Valid		0.130	0.235	Not Valid	
30	1	0.163	0.361	Not Valid	Retest	0.304	0.235	Valid	Used
	3	0.216	0.361	Not Valid		0.342	0.235	Valid	

Before conducting the first test and second test, the instrument questions must be approved by expert judgment by providing feedback for each question and revising the questions based on the feedback provided (see Appendix C.1). Validity based on expert judgment involved three experts, namely two lecturers who were experts in the field of biology and one biology teacher. Determination of instrument validity for expert judgment uses the Aiken Index (see Appendix B.6). The content validity in this research was assessed by 3 Expert Judgments who were analyzed using Aiken's V formula (Azwar, 2015) Aiken's V formula is:

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

V= Rater's Fit Index

s= Average score – the lowest score in category

c= Number of Categories

n= Number of Raters

The V Aiken coefficient is then determined. To decide whether a test item is valid, criteria must be established. The parameters used to determine the validity of content are as follows: below 0.4, the validity is invalid; between 0.4 and 0.8, the validity is moderate; and beyond 0.8, the validity is high (Irawan & Wilujeng, 2020). The criteria for validity using the Aiken Index can be seen in Table 3.5.

Table 3.5 Criteria for Validity using the Aiken Index

Index Validity (V)	Interpretation
$V > 0.8$	High Validity
$0.4 < V \leq 0.8$	Moderate Validity
$V \leq 0.4$	Invalid

(Irawan & Wilujeng, 2020)

The Aiken's V Index ranges between 0 and 1, with a higher value indicating a higher level of agreement among validators regarding the relevance of the instrument items. A value closer to 0 suggests low agreement among validators, indicating that the item is not considered relevant or necessary for measuring the construct. A value closer to 1 indicates high agreement among validators, meaning that the item is considered essential and relevant for measuring the construct (Aiken, 1980).

Table 3.6 The Result of Aiken Test

Item Number	Tier I		Tier III	
	Aiken's Index (V)	Interpretation	Aiken's Index (V)	Interpretation
Q1	0.5	Moderate	0.5	Moderate
Q2	0.7	Moderate	0.7	Moderate
Q3	0.8	High	0.8	High
Q4	0.7	Moderate	0.7	Moderate

Item Number	Tier I		Tier III	
	Aiken's Index (V)	Interpretation	Aiken's Index (V)	Interpretation
Q5	0.7	Moderate	0.7	Moderate
Q6	0.7	Moderate	0.7	Moderate
Q7	0.8	High	0.8	High
Q8	0.8	High	0.8	High
Q9	0.7	Moderate	0.7	Moderate
Q10	0.7	Moderate	0.7	Moderate
Q11	0.8	High	0.8	High
Q12	0.8	High	0.8	High
Q13	0.7	Moderate	0.7	Moderate
Q14	0.7	Moderate	0.7	Moderate
Q15	0.8	High	0.8	High
Q16	0.8	High	0.8	High
Q17	0.7	Moderate	0.7	Moderate
Q18	0.8	High	0.8	High
Q19	0.8	High	0.8	High
Q20	0.8	High	0.8	High
Q21	0.8	High	0.8	High
Q22	0.8	High	0.8	High
Q23	1.0	High	1.0	High
Q24	0.8	High	0.8	High

Item Number	Tier I		Tier III	
	Aiken's Index (V)	Interpretation	Aiken's Index (V)	Interpretation
Q25	0.8	High	0.8	High
Q26	1.0	High	1.0	High
Q27	0.8	High	0.8	High
Q28	0.8	High	0.8	High
Q29	0.7	Moderate	0.7	Moderate
Q30	0.8	High	0.8	High
V Average	0.8	High	0.8	High

Table 3.6 above shows that the coefficient value V for each test item is in the range between 0.7 and 1, with an average value of 0.8 for all 30 questions on first tier and third tier, which is a high validity category. So, it can be concluded that this test item can be used to measure what should be measured, so this instrument is suitable to measure.

3.3.2 Reliability Test

Test items that have been declared valid in the first and second third tests, then undergo reliability testing. Reliability testing using Cronbach's alpha is a common method to assess the internal consistency of a scale or questionnaire (Creswell, 2002). In the first test, there are four test items with 0.618 and 0.549 Cronbach's Alpha value (see Appendix B.3) which are categorized as acceptable based on the results of research conducted by Taber (2018). After that, the revised instrument was then redistributed and a second reliability test was carried out. In the second test, as many as ten test items tested with 0.688 and 6.71 Cronbach's Alpha value (see Appendix B.4) were categorized as acceptable based on the findings of research

conducted by Taber (2018), so it can be concluded that the instrument can be used for research. The results of the overall reliability test can be seen in Table 3.7.

Table 3.7 Result of Reliability Test

Test	N of Items	Cronbach's Alpha	
		Tier 1	Tier 3
I	4	0.618	0.549
II	10	0.688	0.671

3.4 Data Analysis

The data that has been obtained will then be analyzed based on the answers given by each student. The answers to each question will be categorized to determine the students' conception. The answer categories are divided into Scientific Knowledge (SK), False Positive (FP), False Negative (FN), Misconceptions (M), and Lack of Knowledge (LK). Grouping student answers can be based on research Kiray & Simsek (2021), as seen in Table 3.6. Student answers can be categorized into the Scientific Knowledge category when students answer the first-tier question correctly, are sure of the answers they choose in the first-tier question, answer the right reasons to support the answers in the first-tier question, and believe in the reasons they choose in the third-tier question. The false positive category is when students answer the first-tier question correctly, are sure of the answer they chose in the first-tier question, but answer reasons that are not appropriate to support the answers to the first-tier question and believe these reasons are the right reasons. The false negative category is when the student is wrong in answering the first-tier question and believes the answer, then gives the right reasons for the third-tier question and believes the reason he answered. The misconception category is when students are wrong in answering the first-tier and third-tier questions, but students have confidence in answering the first and third-tier questions. Other combinations of answers from the above categories will be categorized as lack of knowledge.

Table 3.8 Comparison of Decision of Four-Tier Test

1st Tier	2nd Tier	3rd Tier	4th Tier	Decision of Four-Tier Test
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	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

1st Tier	2nd Tier	3rd Tier	4th Tier	Decision of Four-Tier Test
False	Not Confident	False	Confident	LK 11
False	Not Confident	False	Not Confident	LK 12

In analyzing student answers, researchers used the Microsoft Excel program by referring to the categorization of answers in Table 3.6. In addition, giving a score that will be used to analyze data on the four-tier test instrument is carried out by evaluating the correct or wrong answers and responses to the confidence tier together. In the assessment process, the correct answer to the first-tier and third-tier questions is coded "1," and the wrong answer is coded "0". In the confidence tier, namely the second tier and fourth tier, the question of the not confident level was encoded with 0 and the confidence level with 1. The scoring categories can be summarized as follows:

1. Scientific Knowledge: It is the category determined as a result of four-tier evaluation of the students' correct answers. When scoring through the correct answers, the correct answers in question and reason items in each question are coded as “1” and the wrong answers are coded as “0”. In this respect, if each of the question, reason and confidence tiers is coded with “1” with the sequence (1-1-1-1) the evaluation is made as scientific knowledge.
2. False Positive: Accompanied by high level of confidence, Students’ correct answer for the first tier and the wrong answer for the reason tier were coded as “1” and the others as “0” (1-1-0-1).
3. False Negative: Accompanied by high level of confidence, Students’ wrong answer for the first tier and the correct answer for the reason tier were coded as “1” and the others as “0” (0-1-1-1).

4. Misconception: It is the category determined for the students' wrong answers in the question and reason tiers. When evaluating misconceptions, the wrong answers in question and reason items in each question item are coded as “1” and the correct answers are coded as “0”. In the analysis for the wrong answers, if each of the question, reason and confidence tiers is coded with “1” with the sequence (1-0-1-0), evaluation is made as misconception.
5. Lack of Knowledge: Other answer categories apart from the categories described above were coded as “1” and the evaluation was made as lack of knowledge.

The frequency and percentage of each category and item are analyzed and calculated in the data analysis stage. In the first stage, the level of students' conceptions, such as scientific knowledge, false positives, false negatives, misconceptions, and lack of knowledge regarding diffusion and osmosis using as many as 14 test items, are counted and categorized based on Table 3.6. In the second stage, the students' conception category with the highest percentage, and the students' misconceptions category will be calculated and analyzed further.

3.5 Research Procedure

This research procedure was designed with the aim of systematically collecting relevant data which would provide an in-depth understanding of the phenomenon being studied. The research procedure carried out by the researcher is shown in Figure 3.1. There are three stages of the procedure in this research starting from the preparation stage, implementation stage, and completion stage. The stages in the research procedure will be explained further below.

1. Preparation Stage
    a. Identifying the problem and research question
    b. Formulating research objectives and benefit
    c. Analyzing the topic of diffusion and osmosis contained in the Cambridge curriculum.

d. Analyzing available instruments related to students' conception on the topic of diffusion and osmosis.
e. Making instruments in the form of 30 four tier questions regarding diffusion and osmosis which have been analyzed previously (see in Appendix A.1).
f. Distributing instruments to 32 students who are not research samples in research using written paper.
g. Conduct validation tests on student answers with research permit (see Appendix C.2 and Appendix C.3)
h. Revise instruments that are declared invalid and eliminate questions that cannot be used for further validation tests based on experts' judgment (2 lectures of biology education and 1 biology teacher).
i. Distributing instruments to 32 students who are not research samples in research using written paper.
j. Distributing the revised instrument to 72 students who were not research samples in the real research using written paper and google forms (see Appendix D.1).
k. Conduct reliability test test on the instrument.
l. Finalize the final instrument.

2. Implementation Stage

At this stage, four tier test instruments are given to students using the Google form. The instrument contains a total of 14 four tier test questions (see in Appendix A.2). In this research, researchers directly supervised respondents while they were working on the instruments that had been given. The instrument was worked on by grade 9 students on May 19 2023 and by grade 7 students on May 23 2023. Then interviews were conducted with 4 students to find out the causes of the misconceptions they experienced.

3. Completion Stage
   a. The data obtained were then analyzed statistically using the Microsoft Excel program.

b. Making a discussion of the findings obtained in the research and interpreting the data.
c. Make conclusions, implications, and recommendations based on the results of the data analysis that has been obtained.
d. Data for research results and discussion after carrying out thesis guidance then obtain permission from the supervisor for thesis defense (see Appendix C.4)
e. Create a research journal for this research which is then submitted to the Jurnal Pendidikan dan Pembelajaran Sinta 3 (See Appendix C.5)
f. Reporting research paper.

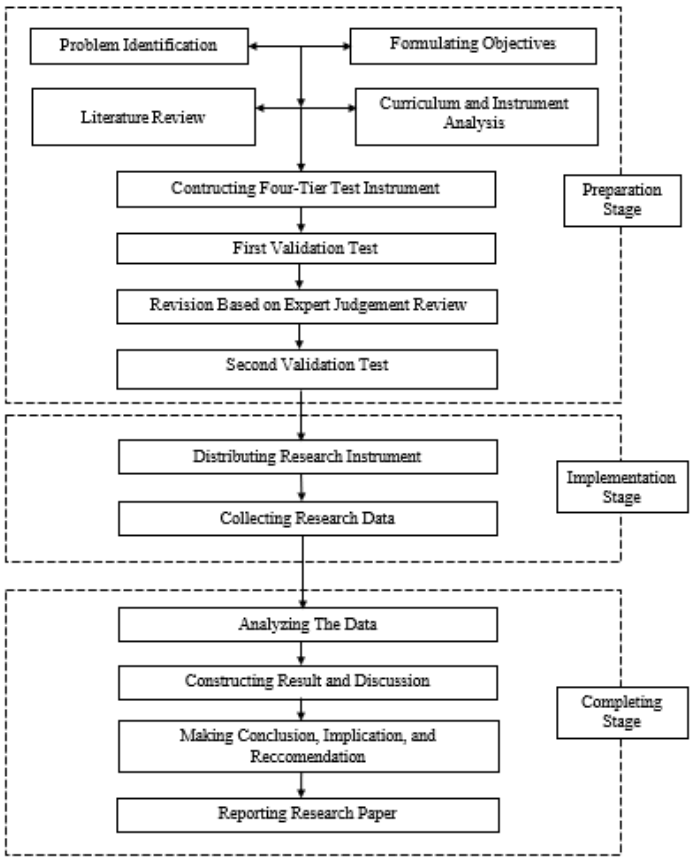


Figure 3.1 Research Procedure Flowchart