CHAPTER III

RESEARCH METHODOLOGY

3.1 Research Design

This study uses a survey research design. Survey design is one of procedures in quantitative descriptive method that after which the findings will be processed and examined to draw a conclusion. This indicates that the research is heavily focused on numbers (Cohen, 2007). The variables that being examined are things that an individual already exists, and the researcher does not modify the research variables, this study is non experimental.

The design of this study used a cross-sectional survey design. In a cross-sectional survey design, information is gathered all at once (Creswell, 2012). The advantage of this design is that it can gauge current beliefs or behaviors. Additionally, it offers information in a brief period of time, such as the time needed to conduct the survey and gather the data. The instruments that have been filled in are then collected and analyzed by researchers to find a profile of students' misconceptions.

3.2 Population and Sample

1. Population

The population of this research are all 7^{th} grade students in a Public Junior High School that are located in Bogor, that have already learned about acid and base topic using 2013 national curriculum.

2. Sample

The sample of this research are 121 students of 7^{th} students in Bogor City, West Java Province, that have already learned about acid and base topic using 2013 national curriculum. The distribution of students shown in Table 3.1. The sample was considered for the study is convenience sampling was employed in this research, which involves selecting samples from easily accessible individuals (Wu & Thompson, 2020). The sample was chosen based on convenience and availability, and data collection continued until the required sample size was attained. This sampling method was particularly suitable for the research due to the

need for a large number of participants within a limited timeframe. Based on the students' list in Appendix A.1, the distribution of participant can see in Table 3.1.

Table 3.1

Participant Distribution

Gender	Number of Students	Percentage (%)
Male	51	42.14
Female	70	57.86
Total	121	100

3.3 Research Instrument

A four-tier diagnostic test on acid and base will be used as the research instrument in this study that was constructed using a list of learning indicators for the acid and base topic from the 2013 national curriculum. Questionnaires were used as the form of this research. There are 13 sets of questions constructed with three sub-topics which are characteristics of acid and base, identification of acid and base, and determination scale on acidity and alkalinity. The distribution and the indicators can be seen in Table 3.2.

Table 3.2

Test Item Concept Indicator and Distribution

Sub-Concept	Indicator of Question	Number of Question
Characteristics of Acid and Base	Properties of Acid and Base	Q3
	Properties of Acid and Base	Q4
	Role Acid and Base in Daily Life	Q13
Identification of Acid and Base Solution	Acid and Base Solution	Q6
	Neutralization Reaction	Q7
	Base Indicator use Universal Indicator	Q8
	Acid Indicator use Universal Indicator	Q12
Determination on Scale Acidity and	Base Indicator on pH Diagram	Q1
	Acid Indicator on pH Diagram	Q2

Sub-Concept	Indicator of Question	Number of Question
Alkalinity	Measure pH of Acid and Base Solution	Q5
	Measure pH of Acid and Base Solution	Q9
	Acid Indicator on pH Diagram	Q10
	Measure pH of Acid and Base Solution	Q11

The questions are organized into four tiers. The first tier consists of multiple-choice questions with four options. The second tier focuses on the confidence level of the answer provided in the first tier, offering two options: "sure" and "not sure." Moving to the third tier, students are asked to provide a scientific reason for their answer in the first tier, with four options to choose from. Lastly, the fourth tier evaluates the confidence level of the answer provided in the third tier. An example question from this four-tier structure is illustrated in Table 3.3. The package all of question can be seen in Appendix A.2.

Table 3.3

Example of Four-Tier Diagnostic Test on Acid and Base Topic

Tier	Question
1	The following is the diagram of the pH of solutions A, B, C, and D according to color of universal litmus paper used. The order of the solutions containing H^+ ions from smallest to largest is…  Source: Dok. JanSanConsulting pH scale D, C, B, A C, D, B, A A, B, C, D C, A, B, D

Tier	Question
2	Are you sure with your answer? Sure b. Not sure The reason you choose the answer is… The number of H^+ ions appropriate to the concentration of acid and base The number of H^+ ions is inversely proportional to the concentration of acid and base The smallest number of H^+ ions is found in neutral compounds The smallest number of H^+ ions is found in the strongest compounds
3	
4	Are you sure with your answer? Sure b. Not sure

In Table 3.3, the question is to see students' conception on one of the identification acid and base solution which is identify pH value of the solution. Acid solution has pH value smaller than 7 while base have pH value bigger than 7. The students then are asked what the order of pH value of the solution from the smaller. Students are also asked to choose their reason and the confidence level both of the first and the third tiers.

The instrument undergoes a two-stage development process. In the first stage, a group of 30 students participates, and they are presented with 30 sets of questions through a Google Form. Subsequently, the obtained results are examined for validity and reliability using SPSS software. Moving on to the second stage of development, another group of 30 students is involved. This time, 13 sets of questions are tested (can be see in Appendix A.3), which have been revised based on the feedback received from the previous stage as some of the initial questions were found to be invalid.

3.3.1 Validity Test

Validity and reliability tests must be conducted on the instrument being utilized. The result of first test of validity and reliability can see in Appendix A.4.

For a question to be deemed valid, both tier two and tier three must be valid. The instrument test will be run using SPSS by the Pearson correlation coefficient approach to strengthen the instrument's validity. The result of the correlation test is shown in Table 3.4. Pearson's correlation coefficient (*r*) is a measure of the linear association of two variables (Fraenkel, 2012).

To assess the instrument's validity as a measurement tool, a comparison is made between the calculated corrected total item correlation (*r*-count) and the value in the r-table. If the r-count is higher than the r-table value, it indicates that the test item is significantly correlated with the total score, and therefore, the test items are considered valid. Conversely, if the r-count is lower than the r-table value, the test items are not significantly correlated with the total score, making them invalid.

To find the value in the r-table corresponding to a given degrees of freedom (df), we can use the formula Df = n - 2, where 'n' is the number of samples used. The value of r count with r table can seen in Appendix A.5. The Pearson correlation coefficient for 30 respondents, 2 tailed test, in 0.05 significance level is *r* (28) = 0.374 (Pearson Education, 2017). The question is considered valid and capable of measuring the desired variable when the coefficient exceeds 0.374 for both the first and second tests. The result of correlation test shown in Table 3.4.

Table 3.4

The Result of Validity Test

Question	Tier	Pearson Correlation	Test I Interpretation	Decision	Person Correlation	Test II Interpretation	Decision
Q1	1	.116	Not Valid	Not Used			
	3	-.087	Not Valid				
Q2	1	.157	Not Valid	Not Used			
	3	-.198	Not Valid				
Q3	1	.169	Not Valid	Retest	.384	Valid	Directly Used
	3	.155	Not Valid		.377	Valid	
Q4	1	.144	Not Valid	Not Used			
	3	-.013	Not Valid				
Q5	1	.291	Not Valid	Retest	.374	Valid	Directly Used
	3	.184	Not Valid		.429	Valid	
Q6	1	.043	Not Valid	Retest	.397	Valid	Directly Used
	3	.380	Valid		.374	Valid	
Q7	1	.382	Valid	Directly Used			
	3	.374	Valid				
Q8	1	.111	Not Valid	Retest	.282	Not	Not Used

Question	Tier	Pearson Correlation	Test I Interpretation	Decision	Person Correlation	Test II Interpretation	Decision
	3	.328	Not Valid		-.065	Valid Not Valid	
Q9	1	.309	Not Valid	Not Used			
	3	-.321	Not Valid				
Q10	1	.384	Valid	Retest	.390	Valid	Directly Used
	3	.287	Not Valid		.379	Valid	
Q11	1	-.110	Not Valid	Not Used			
	3	.000	Not Valid				
Q12	1	.379	Valid	Retest	.440	Valid	Directly Used
	3	.210	Not Valid		.384	Valid	
Q13	1	.415	Valid	Directly Used			
	3	.375	Valid				
Q14	1	-.291	Not Valid	Not Used			
	3	.064	Not Valid				
Q15	1	.374	Valid	Retest	-.178	Not Valid	Not Used
	3	.202	Not Valid		.157	Not Valid	
Q16	1	.229	Not Valid	Retest	.434	Valid	Directly Used
	3	.316	Not Valid		.397	Valid	
Q17	1	-.046	Not Valid	Not Used			
	3	.115	Not Valid				
Q18	1	.376	Valid	Retest	.429	Valid	Directly Used
	3	.259	Not Valid		.377	Valid	
Q19	1	-.065	Not Valid	Not Used			
	3	.192	Not Valid				
Q20	1	-.178	Not Valid	Not Used			
	3	.116	Not Valid				
Q21	1	.309	Not Valid	Retest	.397	Valid	Directly Used
	3	.159	Not Valid		.374	Valid	
Q22	1	.446	Valid	Directly Used			
	3	.387	Valid				
Q23	1	-.0.35	Not Valid	Not Used			
	3	-.0,70	Not Valid				
Q24	1	.377	Valid	Retest	.144	Not Valid	Not Used
	3	.056	Not Valid		-.157	Not Valid	
Q25	1	.208	Not Valid	Not Used			
	3	-.038	Not Valid				
Q26	1	.397	Valid	Retest	.291	Not Valid	Not Used
	3	.062	Not Valid		.111	Not Valid	
Q27	1	.213	Not Valid	Not Used			
	3	-.146	Not Valid				
Q28	1	-.195	Not Valid	Not Used			
	3	-.295	Not Valid				
Q29	1	.429	Valid	Retest	.471	Valid	Directly Used
	3	.031	Not Valid		.434	Valid	
Q30	1	.251	Not Valid	Retest	.379	Valid	Directly Used
	3	.377	Valid		.384	Valid	

Based on the information provided in Table 3.5, it can be deduced that only Q7, Q13, and Q22 are considered valid questions in the first test. Following the identification of invalid questions in the previous test, revisions were made to address the issues. Various improvements were implemented, such as modifying certain lines within the questions, incorporating images, and making changes to certain options. These revisions aimed to enhance the validity and effectiveness of the questions.

In this study, the content validity was assessed through three expert judgments, comprising two chemistry expert lecturers and one chemistry teacher (Can be see in Appendix A.6). The content validity of the instrument involves an analysis of how well the questions represent the measurable abilities, and this can be determined by obtaining agreement from the experts. To gauge this agreement, the validity index introduced by Aiken (1980) can be utilized as follows:

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

The rater agreement index (V) regarding item validity is calculated using the formula: V = (s / (n * c - lo)). Here, 's' represents the score determined by each rater minus the lowest score in the category used (s = r – lo, where r = the score in the rater's chosen category; lo = the lowest score in the scoring category); 'n' is the number of raters, and 'c' is the number of categories the rater can choose.

Based on the calculated V index result, each item can be categorized accordingly. The content validity results using Aiken's approach can be found in Appendix A.7. The categorization of the instrument's content validity is based on the criteria presented in Table 3.5.

Table 3.5 Validity Index Rating Scale

Aiken's Index (V)	Validity Criteria
$0 \leq V \leq 0.3$	Small
$0.3 < V \leq 0.7$	Moderate
$0.7 < V \leq 1.0$	Large

(Hsu, et al., 2015)

The validity index also falls within the range of 0 to 1. A value closer to 0 suggests a lower level of agreement among validators concerning the relevance of the instrument item. Conversely, if the validity coefficient approaches 1, it indicates a higher level of agreement among validators regarding the instrument items' relevance to their respective indicators (Retnawati, 2016). The Aiken test result on the instrument shown in Table 3.6.

Table 3.6 Aiken Test Result

Test Item	Tier I		Tier 3	
	Aiken's Index (V)	**Validity Criteria**	**Aiken's Index (V)**	**Validity Criteria**
Q1	0.8	Large	1.0	Large
Q2	1.0	Large	1.0	Large
Q3	1.0	Large	0.8	Large
Q4	0.7	Moderate	0.7	Moderate
Q5	1.0	Large	1.0	Large
Q6	0.8	Large	1.0	Large
Q7	1.0	Large	0.8	Large
Q8	0.7	Moderate	0.7	Moderate
Q9	0.8	Large	1.0	Large
Q10	1.0	Large	0.8	Large
Q11	0.8	Large	0.8	Large
Q12	0.8	Large	1.0	Large
Q13	1.0	Large	1.0	Large
Q14	0.8	Large	0.8	Large
Q15	1.0	Large	0.7	Moderate
Q16	1.0	Large	1.0	Large
Q17	1.0	Large	0.7	Moderate
Q18	1.0	Large	0.8	Large
Q19	0.7	Moderate	0.7	Moderate
Q20	1.0	Large	1.0	Large
Q21	0.8	Large	0.8	Large

Test Item	Tier I		Tier 3	
	Aiken's Index (V)	Validity Criteria	Aiken's Index (V)	Validity Criteria
Q22	1.0	Large	1.0	Large
Q23	1.0	Large	1.0	Large
Q24	0.7	Moderate	0.8	Large
Q25	1.0	Large	1.0	Large
Q26	0.8	Large	0.8	Large
Q27	0.7	Moderate	0.7	Moderate
Q28	1.0	Large	1.0	Large
Q29	1.0	Large	1.0	Large
Q30	1.0	Large	0.8	Large
$V_{average}$	0.9	Large	0.9	Large

According to the findings presented in Table 3.6, the results of the content validity for the 30 questions indicated that all of the questions were determined to be valid with the average of Aiken's Index is 0.9 for both tiers which shows that the set of questions is in the category of high validity. Specifically, the analysis revealed that 83% of the questions were classified in the high category, while the remaining 17% were classified into moderate category.

3.3.2 Reliability Test

After being re-administered to another group of 30 students, the question items underwent a second round of validity and reliability testing. This time, all 13 questions were deemed valid. During this second validation phase, 13 questions successfully met the criteria and were considered valuable items, as shown in Table 3.4. These 13 questions were meticulously selected as the final instrument for analyzing students' misconceptions related to the topic of acids and bases. The valid questions then underwent a reliability test. Cronbach's alpha is a measure of reliability and, more specifically, internal consistency (Creswell, 2012). the result of reliability test can be see in Appendix A.8.

The reliable questions from both the first and second tests were

subsequently subjected to a reliability assessment. The 13 questions in the first tier demonstrated a reliability score of 0.720, as indicated by the Cronbach's Alpha Score. Similarly, the 13 questions in the second tier exhibited a reliability score of 0.705 according to the Cronbach's Alpha Score. With these scores, the instrument is deemed acceptable based on the guidelines established by Taber in the publication "The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education" (2018). Detailed information regarding the Cronbach's Alpha Scores can be found in Table 3.7. In total, there are 13 sets of questions that are considered suitable for diagnosing students' understanding of the acid and base topic.

Table 3.7

Result of Reliability Index on Test Item

N of Items	Tier	Cronbach's Alpha
13	1	0.720
13	3	0.705

3.4 Research Procedure

1. Preparation Stage
   a. Identify the research problem and elaborate it into several research questions
   b. Literature review about students' conception
   c. Designing instrument for making 30 question set four-tier diagnostics test of acid and base topic base on 2013 curriculum.
   d. Conducting expert judgement by 3 experts in chemistry topic and revising instrument.
   e.First administering the instrument to students with Google Form (https://bit.ly/InstrumentAcidandBase)
   f. Running the validity and reliability of the instrument checked by using SPSS with Pearson Correlation test and Cronbach's Alpha formula.
   g. Revision the 10 invalid questions.
   h.Second administering of 10 invalid questions to students

(https://bit.ly/InstrumentAcidandBasePart2)

i. Running the validity and reliability of the instrument checked by using SPSS with Pearson Correlation test and Cronbach's Alpha formula.

2. Implementation Stage

In this stage, the four-tier diagnostics test is administered to students with google form. The test consists of 13 sets of questions in total. Students filled in the question in the class with the presence of the researcher. The link of google form (https://bit.ly/*InstrumenTesAsamBasa*). The link of questions was administered on May 4th 2023. Permission letter from faculty and school can be seen in Appendix B.1 and Appendix B.2.

3. Completion Stage

a. Analyzing the collected data quantitatively, then constructing the result based on analysis data. After that, making conclusion from the result.

b. Constructing the discussion based on data analysis and interpretation.

c. Constructing conclusion and recommendation based on the result and discussion.

d. Completing the research paper by approval of supervisor (Can be see in Appendix B.3) and also already submit journal to SINTA 3 that can be see in Appendix B.4.

The schema of the research procedure shown on Figure 3.1 below.

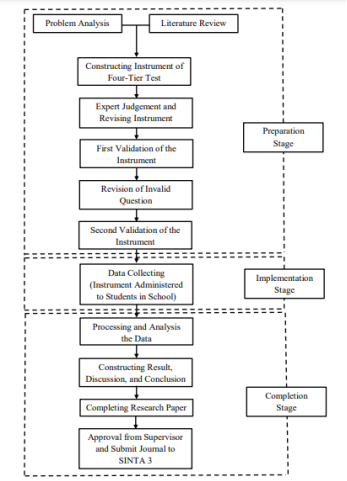

Figure 3.1 Schema of Research Procedure

3.5 Data Analysis

This research got the data by giving instruments to 7th grade students. Answer for each question will be categorized into Scientific Knowledge (SK), False Positive (FP), False Negative (FN), Misconception (M), and Lack of Knowledge (LK). The categories can be seen in Table (Gurel, 2015) for the detail. Scientific knowledge is characterized by students correctly answering the first-tier, expressing confidence in their response to the first-tier, providing a correct response to the third-tier, and displaying confidence in their answer to the third-tier. False positives occur when students answer the first-tier correctly, exhibit certainty in their response to the first-tier, but provide an incorrect response to the third-tier while remaining confident in their answer. On the other hand, false negatives transpire when students answer the first-tier incorrectly, yet express certainty in their response to the first-tier, while correctly answering the third-tier and displaying confidence in their response. Misconceptions arise when students answer both the first- and third-tier incorrectly, but remain confident in their answers to both tiers. Any combinations that do not fall into these categories are classified as a lack of knowledge. The combination answer and decision on the instrument test can seen in Table 3.8.

Table 3.8

Combination Answer and Decision on Four-tier Test

Tier 1	Tier 2	Tier 3	Tier 4	Decision of Four-Tier
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	LK1
True	Not Confident	True	Confident	LK2
True	Not Confident	True	Not Confident	LK3

Tier 1	Tier 2	Tier 3	Tier 4	Decision of Four-Tier
True	Confident	False	Not Confident	LK4
True	Not Confident	False	Confident	LK5
True	Not Confident	False	Not Confident	LK6
False	Confident	True	Not Confident	LK7
False	Not Confident	True	Confident	LK8
False	Not Confident	True	Not Confident	LK9
False	Confident	False	Not Confident	LK10
False	Not Confident	False	Confident	LK11
False	Not Confident	False	Not Confident	LK12

(Kiray & Simsek, 2021)

SK scientific knowledge, *FP* false positive, *FN* false negative, *M* misconception, *LK* lack of knowledge.

To analyze the data, the Excel program was utilized, incorporating the combinations and decisions specified in Table 3.8. A scoring system was implemented, assigning a score of "1" to responses that were correct in both the first and third tiers, and a score of "0" to incorrect answers. In the case of the second and fourth tiers, the selection of the "sure" option was scored as "1," while the "not sure" option was scored as "0". This scoring system enabled a structured evaluation of the data, facilitating further analysis and interpretation.

According to the scoring criteria used, responses that exhibit a sequence of "1-1-1-1" were classified as scientific knowledge (SK). These answers demonstrate

a correct understanding of the subject matter. On the other hand, responses showing a sequence of "1-1-0-1" were categorized as false positive (FP). These responses may falsely indicate scientific knowledge despite containing some incorrect information. Conversely, a sequence of "0-1-1-1" represented false negative (FN). These responses may lack scientific knowledge even though they contain some correct information. Lastly, answers demonstrating the sequence "0-1-0-1" were classified as misconceptions (M). These responses contain incorrect information and demonstrate a misunderstanding or misconception about the subject. Any responses that do not match these specific sequences are categorized as a lack of knowledge (LK). These answers indicate a general absence of understanding or knowledge regarding the subject matter being assessed.

During the data analysis, frequencies and percentages were computed to examine the results. In the first stage, the percentages of scientific knowledge (SK), false positive (FP), false negative (FN), misconception (M), and lack of knowledge (LK) were analyzed. In the second stage, particular attention was given to questions that exhibited a misconception rate exceeding 45%. A deeper analysis was conducted on the option combinations of student answers to those questions. Further frequencies and percentages were calculated to provide a comprehensive understanding of the data. The formula used for calculated the percentages is:

$$P = \frac{F}{N} \times 100\%$$

Description:

P: Percentage of students' conceptual level

F: Frequency of student's conceptual level

N: Total number of students