CHAPTER III
RESEARCH METHODOLOGY

3.1 Research Design

Aligned with the aim of the research to diagnose students' misconception on work and simple machine topics, this research used survey research design. According to Creswell (2012), survey research designs are procedures in quantitative research in which investigators administer surveys using questionnaires or interviews to gather numerical data from a sample or the entire population. The aim is to understand the attitudes, opinions, behaviors, or characteristics of the population. The collected data is then statistically analyzed to identify trends in responses and test research questions. Additionally, researchers interpret the data by comparing it to previous studies. Different from experimental research, survey research does not involve a treatment given to participants by the researcher.

This research is also identified as a cross-sectional survey design research. This design offers the advantage of assessing current attitudes or behaviors. In a cross-sectional survey design, the researcher gathers data at a single moment in time. It provides information in a brief period of time due to the efficient process of survey administration and data collection (Creswell, 2012). The instrument used in this research consists of a four-tier test on work and simple machine topics that can diagnose misconceptions. The data were analyzed and interpreted to identify the misconceptions on work and simple machine topics experienced by the sample. In this research, the researcher simply portrays students' conception as what it is, no intervention is made in this research.

3.2 Participant

The participants involved in this research were junior high school students. The criteria for selecting participants were students who have already learned work and simple machine topics as outlined in the 2013 national curriculum. All the participants came from one public school. The population of this research are all 8^{th} and 9^{th} grade students in a public junior high school that is located in Tasikmalaya,

West Java. The 110 students of 8th and 9th grade students from this school participated as the sample.

Sample of convenience is used in this research, according to Creswell (2012) it is a method in which the researcher chooses participants based on their willingness and availability to be included in the study. The respondents are chosen for the sample based on conveniences and availability. The respondents are continued to be searched by obtaining data from accessible samples until the required sample size has been achieved. This sampling technique is very suitable to use in this research because researchers have limited time and require many participants.

3.3 Research Instrument

The instrument is required to gather data from the participants. In this research, a four-tier diagnostic test on work and simple machine topics is used to diagnose students' conception on the topics. The instrument is developed based on a list of indicators from the previous research with several changes (Ponidi, 2011; Maison et al., 2020; Yuberti et al., 2020, Nazura et al., 2021; Taqwim, 2021; Christina et al., 2022). There are 15 sets of questions constructed. There are three subtopics which are work, the types of simple machines, and the working principle of a simple machine in a movement system. The subtopic and concept of each question can be seen in Table. 3.1 below.

Table 3.1 Test Item Number Distribution

Subtopic	Concept	Test Item
Work	Work concept	Q1
	Magnitude of work	Q2
Types of simple machine	Example of simple machine	Q3
	Example of simple machine	Q4
	Example of simple machine	Q5
	Mechanical advantage of lever	Q6
	Mechanical advantage of pulley	Q7
	Mechanical advantage of pulley	Q8
	Mechanical advantage of inclined plane	Q9
	Principle of simple machine on movement system	Q10
	Principle of simple machine on movement system	Q11

Subtopic	Concept	Test Item
Simple machine in human movement system	Principle of simple machine on movement system	Q12
	Principle of simple machine on movement system	Q13
	Principle of simple machine on movement system	Q14
	Principle of simple machine on movement system	Q15

Every question set comprises four tiers. In the first tier, questions are presented as multiple-choice items with four options. The second tier measures the respondent's confidence level in regard to their answer in the first tier, offering two choices: "sure" and "not sure." For the third tier, students are prompted to provide the scientific reason for their answer to the first-tier question, four options will be available. The fourth tier measures the confidence level associated with the answer provided in the third tier. A sample question is displayed in Table 3.2.

Table 3.2 Sample Question

No	Tier	Question
2	1	Note the following statement. 1) Nana pushes a chair with 10N force within 6 meters. 2) Rara pulls a chair with 12N force within 5 meters. Which statement produces the biggest work? A. The first activity done the bigger magnitude of work B. The second activity done the bigger magnitude of work C. Both statements done the same magnitude of work D. None of both statements done work
	2	Are you sure with your answer? A. Sure B. Not sure
	3	Why did you choose that answer? A. Work done only when the force direction is same with the movement direction B. Work done only when the force direction is perpendicular to the movement direction C. Work is calculated by multiplying force and displacement D. The force done was not enough to produce the work.
	4	Are you sure with your answer? A. Sure B. Not sure

Displayed in Table 3.2 is a question aimed at assessing students' conception of the work subtopic. Alongside selecting their answers, students are also prompted to indicate their reasoning and express their confidence levels for both the first and the third tiers. At first 30 sets of questions are constructed then the instrument undergoes expert judgement and two stages of development. The first development stage engages 30 students, who are administered these 30 sets of questions via a Google form. Subsequently, the outcome is subjected to validity and reliability tests through the utilization of SPSS. As a result of this stage, 8 sets of questions are considered as valid. The second phase of development engages 32 students. Within this stage 15 question sets, revised from the previous stage, undergo a retest. The results then undergo validity and reliability tests. The second development resulting total of 7 test items considered as valid. So, in total there are 15 test items for the real research. The test item for validation test can be seen in Appendix A.1. The test item for real research can be seen in Appendix A.2.

3.3.1 Validity Test

Once the instrument is administered to students, it undergoes the validity test. Validity refers to the degree to which the evidence supports the inferences made by a researcher based on the data collected using a specific instrument (Fraenkel et al., 2011). Validity test is used to show how accurate the instrument is. Content validity relates to the ability of the assessment tool to measure the content it should be.

In this research, content validity was carried out with the help of three expert judgements consisting of two expert lecturers in the field of physics and one science teacher. The expert judgement form can be seen in Appendix A.3. The content validity of an instrument is an analysis of the representation of questions with the ability to be measured that can be determined by using the expert agreement. To find out this agreement, the validity index introduced by Aiken (1980) can be used as follows:

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

where V is the rater agreement index regarding item validity; s is the score determined by each rater minus the lowest score in the category used ($s = r - l_o$,

where r = score in the rater's chosen category; l_o = the lowest score in the scoring category); n is number of raters; and c is the number of categories the rater can choose). From the calculation of the V index result, an item can be categorized based on its index. The categorization of the content validity of the instrument is based on Table 3.3.

Table 3.3 Validity Index Rating Scale

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

(Hsu et al., 2015)

If the V value is high (i.e., >0.70), experts concur that the instrument is suitable to be included. Conversely, when Aiken's V is low (i.e., <0.3), experts consider the instrument unsuitable to be included. Moderate values (>0.3 and <0.7) indicate limited appropriateness of the instrument (Hsu et al., 2015; Yu, 1993). Based on the Table 3.3, the range of this index is also between 0 and 1. A value closer to 0 indicates a lower level of agreement among validators regarding the instrument item. On the other hand, if the validity coefficient approaches 1, it signifies a higher level of agreement among validators regarding the instrument items' relevance to their indicators (Retnawati, 2016). The aiken test result can be seen in Table 3.4.

Table 3.4 Aiken Test Result

Test Item	Tier I		Tier 3	
	Aiken's Index (V)	Validity Criteria	Aiken's Index (V)	Validity Criteria
Q1	1.0	High	1.0	High
Q2	1.0	High	1.0	High
Q3	0.8	High	0.8	High
Q4	0.7	Moderate	0.7	Moderate
Q5	1.0	High	1.0	High
Q6	1.0	High	1.0	High

Test Item	Tier I		Tier 3	
	Aiken's Index (V)	Validity Criteria	Aiken's Index (V)	Validity Criteria
Q7	0.7	Moderate	0.7	Moderate
Q8	0.8	High	0.8	High
Q9	0.8	High	1.0	High
Q10	1.0	High	0.8	High
Q11	0.5	Moderate	0.7	Moderate
Q12	0.7	Moderate	0.7	Moderate
Q13	0.8	High	1.0	High
Q14	0.8	High	0.8	High
Q15	1.0	High	1.0	High
Q16	1.0	High	1.0	High
Q17	1.0	High	0.7	Moderate
Q18	1.0	High	0.7	Moderate
Q19	1.0	High	0.7	Moderate
Q20	1.0	High	1.0	High
Q21	0.8	High	0.8	High
Q22	1.0	High	1.0	High
Q23	1.0	High	1.0	High
Q24	1.0	High	0.8	High
Q25	1.0	High	1.0	High
Q26	0.7	Moderate	0.7	Moderate
Q27	1.0	High	0.8	High
Q28	1.0	High	1.8	High
Q29	1.0	High	1.0	High
Q30	0.7	Moderate	0.8	High
$V_{average}$	0.9	High	0.9	High

According to the findings presented in Table 3.4, 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 77% of the questions were classified in the high category, while the remaining 23% were classified into moderate category.

Validity test is also done with corrected item-total correlation. Item analysis through corrected-item total correlation helps to identify the items that do not correlate with the overall test and measure different dispositions or traits. Corrected item-total correlation analysis results showed that all the items exhibited a correlation with the overall test score (Coskun & Kara, 2019; Nunnally &Bernstein, 1994). According to Widiyanto (2010) the corrected-item total correlation technique is theoretically based on the correlation formula concerning spurious overlap effects. The decision-making foundation in validity testing is as follows. If the value of Corrected-Item Total Correlation [R_{count}] > R_{table}, then the questionnaire item or statement is significantly correlated with the total score (meaning the questionnaire item is considered valid). If the value of Corrected-Item Total Correlation [R_{count}] < R_{table}, then the questionnaire item or statement is not significantly correlated with the total score (meaning the questionnaire item is considered invalid).

The result of the correlation test is shown in Table 3.5. For 30 respondents, 2 tailed tests, in 0.05 significance level is R_{table} (30)=0.361 while for 32 respondents is R_{table} (32)= 0.349 (Pearson Education, 2017). The question is valid and can measure the variable wanted when the coefficient is more than 0.361 for the first test and 0.349 for the second test. From table 3.5, it can be concluded that in the first test there are 8 valid questions which are Q8, Q21, Q22, Q25, Q26, Q27, Q28, and Q29. The invalid questions from the previous test are then revised. There are improvements to the questions. some lines in the questions are changed, images are added, and some options are changed. After being re-administered to 32 students, the question items undergo correlation and reliability tests for the second time. From 15 questions tested in the second test, 7 questions are valid which are Q3, Q11, Q15, Q16, Q19, Q23, and Q30. So, the total question that is valid from the first test and second test is 15 questions. The SPSS result for the validity test can be seen in Appendix A.4.

Table 3.5 The Result of Validity Test

Question	Tier	Test I			Test II		
		R_{count}	Interpretation	Decision	R_{count}	Interpretation	Decision
1	1	-.059	Not Valid	Rejected			
	3	.390	Valid				
2	1	.125	Not Valid	Retest	-.013	Not Valid	Rejected
	3	.450	Valid		.188	Not Valid	
3	1	.309	Not Valid	Retest	.377	Valid	Used
	3	.471	Valid		.409	Valid	
4	1	-.094	Not Valid	Rejected			
	3	.642	Valid				
5	1	.356	Not Valid	Retest	.279	Not Valid	Rejected
	3	.410	Valid		.436	Valid	
6	1	.158	Not Valid	Rejected			
	3	-.109	Not Valid				
7	1	.357	Not Valid	Retest	.152	Not Valid	Rejected
	3	.173	Not Valid		.218	Not Valid	
8	1	.515	Valid	Directly			
	3	.521	Valid	Used			
9	1	.285	Not Valid	Retest	.409	Valid	Rejected
	3	.435	Valid		.302	Not Valid	
10	1	.204	Not Valid	Retest	-.099	Not Valid	Rejected
	3	.142	Not Valid		.376	Valid	
11	1	.593	Valid	Retest	.504	Valid	Used
	3	.292	Not Valid		.402	Valid	
12	1	.338	Not Valid	Retest	.114	Not Valid	Rejected
	3	.434	Valid		.216	Not Valid	Rejected
13	1	.564	Valid	Retest	.259	Not Valid	Rejected
	3	.191	Not Valid		.274	Not Valid	
14	1	-.241	Not Valid	Rejected			
	3	.239	Not Valid				
15	1	.294	Not Valid	Retest	.580	Valid	Used
	3	.317	Not Valid		.432	Valid	
16	1	.350	Not Valid	Retest	.610	Valid	Used
	3	.236	Not Valid		.427	Valid	
17	1	.583	Valid	Rejected			
	3	-.105	Not Valid				
18	1	.086	Not Valid	Retest	.276	Not Valid	
	3	.204	Not Valid		.113	Not Valid	
19	1	.203	Not Valid	Retest	.380	Valid	Used
	3	.067	Not Valid		.650	Valid	
20	1	.209	Not Valid	Rejected			
	3	-.343	Not Valid				

Question	Tier	Test I			Test II		
		R_{count}	Interpretation	Decision	R_{count}	Interpretation	Decision
21	1	.511	Valid	Directly			
	3	.471	Valid	Used			
22	1	.453	Valid	Directly			
	3	.495	Valid	Used			
23	1	.601	Valid	Retest	.480	Valid	Used
	3	.351	Not Valid		.588	Valid	
24	1	.536	Valid	Rejected			
	3	-.158	Not Valid				
25	1	.547	Valid	Directly			
	3	.565	Valid	Used			
26	1	.574	Valid	Directly			
	3	.571	Valid	Used			
27	1	.581	Valid	Directly			
	3	.695	Valid	Used			
28	1	.601	Valid	Directly			
	3	.467	Valid	Used			
29	1	.513	Valid	Directly			
	3	.420	Valid	Used			
30	1	.357	Not Valid	Retest	.612	Valid	Used
	3	.356	Not Valid		.602	Valid	

3.3.2 Reliability Test

The valid test items from the first test then underwent a reliability test. Reliability refers to the consistency of scores or answers from one administration of an instrument to another, and from one set of items to another. The alpha coefficient, commonly known as Cronbach's alpha, is another method used to assess internal consistency and estimate the reliability of an instrument (Fraenkel et al., 2011).

The test items are reliable with 0.829 and 0.850 Cronbach's Alpha Score for tier 1 and tier 3 respectively. With that score, the instrument is considered acceptable and fairly high. The second reliability test was done for the revised questions. The 7 test items, consisting of tier 1 and tier 3, are also reliable with 0.778 and 785 Cronbach's Alpha score. This score is considered acceptable, acceptable reliability for each scale alpha is 0.70 or above (Taber, 2018). The Cronbach's Alpha score can be seen in Table 3.6. In total, there are 15 sets of questions that are feasible to diagnose students' conception on work and simple machine topics. The SPSS result for the reliability test can be seen in Appendix A.5

Table 3.6 The Result of Reliability Test

Test	Tier	N of Items	Cronbach's Alpha
I	1	8	0.829
	3	8	0.850
II	1	7	0.778
	3	7	0.785

3.4 Data Analysis

The gathered data underwent analysis, with responses to each question being categorized into distinct classifications: scientific knowledge (SK), lack of knowledge (LK), misconception (M), false negative (FN), and false positive (FP), as detailed in Table 3.7 (Kiray & Simsek, 2020). Scientific knowledge pertains to conditions where students answer the first-tier question correctly, express certainty in their second-tier response, answer the third-tier question correctly, and maintain confidence in their fourth-tier response. False positive is conditions when students correctly answer the first-tier question and are certain about their response, yet answer the third-tier question incorrectly while remaining confident in their fourth-tier response. Conversely, a false negative condition is when students incorrectly answer the first-tier question but exhibit confidence in their response and then proceed to answer the third-tier question correctly with confidence in their response. Misconceptions occur when students respond inaccurately to the first and third tiers while remaining confident in their answers to both. Any other combinations beyond those specified earlier are classified as a lack of knowledge.

Table 3.7 Categories of Conception

1. Tier	2. Tier	3. Tier	4. Tier	Category
True	Confident	True	Confident	SK
True	Confident	False	Confident	FP
False	Confident	True	Confident	NP
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

1. Tier	2. Tier	3. Tier	4. Tier	Category
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	LK9
False	Confident	False	Not confident	LK 10
False	Not confident	False	Confident	LK 11
False	Not confident	False	Not confident	LK 12

(Kiray & Simsek, 2020)

Data analysis was done using the MS Excel program, referencing the combinations and classifications presented in Table 3.7. Correct responses for the first and third tiers were assigned a score of "1," while incorrect answers were given a score of "0." Regarding the second and fourth tiers, selecting the "sure" option was assigned a score of "1," while the "not sure" option received a score of "0." Utilizing these scores, responses featuring a sequence of "1-1-1-1" were categorized under the classification of scientific knowledge (SK). A parallel coding approach was adopted for false negative, false positive, and misconception categories. Responses following a "1-1-0-1" sequence were categorized as false positive (FP), while a "0-1-1-1" sequence was linked to false negative (FP), and a "0-1-0-1" sequence was categorized as misconception (M). Responses different from these sequences were categorized as a lack of knowledge.

The data analysis involved the calculation of frequencies and percentages. During the first stage, the percentages of scientific knowledge (SK), false positive (FP), false negative (FN), misconception (M), and lack of knowledge (LK) were calculated.. The calculation is as follows:

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

Description:

P : Percentage of each level

s : Total number of responses for each level

N : Total number of responses from all level

In the second stage, questions with misconception percentages exceeding 50% were subjected to further analysis. The combinations of answer options chosen by students for these questions were subjected to a more comprehensive analysis. Frequencies and percentages were once again calculated in this context.

3.5 Research Procedure

1. Preparation Stage
   a. Identify the research problem and elaborate it into several research questions.
   b. Analyzing the work and simple machine topics on the 2013 National Curriculum.
   c. Analyzing the existing instruments related to diagnosing students' conception on the work and simple machine topics.
   d. Constructing the instrument from the concepts of work and simple machine topics that have been analyzed.
   e. Conducting the expert judgement and revise the instrument
   f. Administering the instrument to students who had been taught the work and simple machines topics with Google Form on the link https://forms.gle/1bbJaq3Kj4wmEUKt5
   g. Running the validity and reliability test on the result.
   h. Revising the 15 invalid sets of questions.
   i. Administering the15 invalid set of questions to students who had been taught the work and simple machines topics with Google Form on the link https://forms.gle/QGhbVGrAXY8vqVzU9
   j. Running the validity reliability test on the result.
2. Implementation stage

In this stage, the four-tier diagnostic test is administered to students with Google Form on the link https://forms.gle/cgFJtrxM8RkvLzkJ7. The test consists of 15 sets of questions in total. The data collection was conducted over a span of two days, specifically on the April 10th and April 11th 2023. The test carried out within each classroom under the direct supervision of the researchers. Furthermore, short interviews with teacher and students as a sample were conducted. Permission letter

from the faculty can be seen in Appendix B.1 and permission letter from the school can be seen in Appendix B.2. Documentation during this stage can be seen in Appendix C.1.

3. Completion Stage
    a. Analyzing the data that has been collected statistically. To diagnose students' conception, MS-Excel program and SPSS are used.
    b. Constructing the discussion based on data analysis and interpretation.
    c. Constructing conclusion and recommendation based on the result and discussion.
    d. After receiving guidance, the research data and analysis were approved by the supervisors to undergo the "thesis defense". The approval form supervisor can be seen in Appendix B.3.
    e. Submitting the article to Journal Sinta 3 with the journal's name as "INSECTA: Integrative Science Education and Teaching Activity Journal". Submission acknowledgement email from Journal INSECTA can be seen in Appendix B.4. The research procedure of this research can be seen in Figure 3.1.

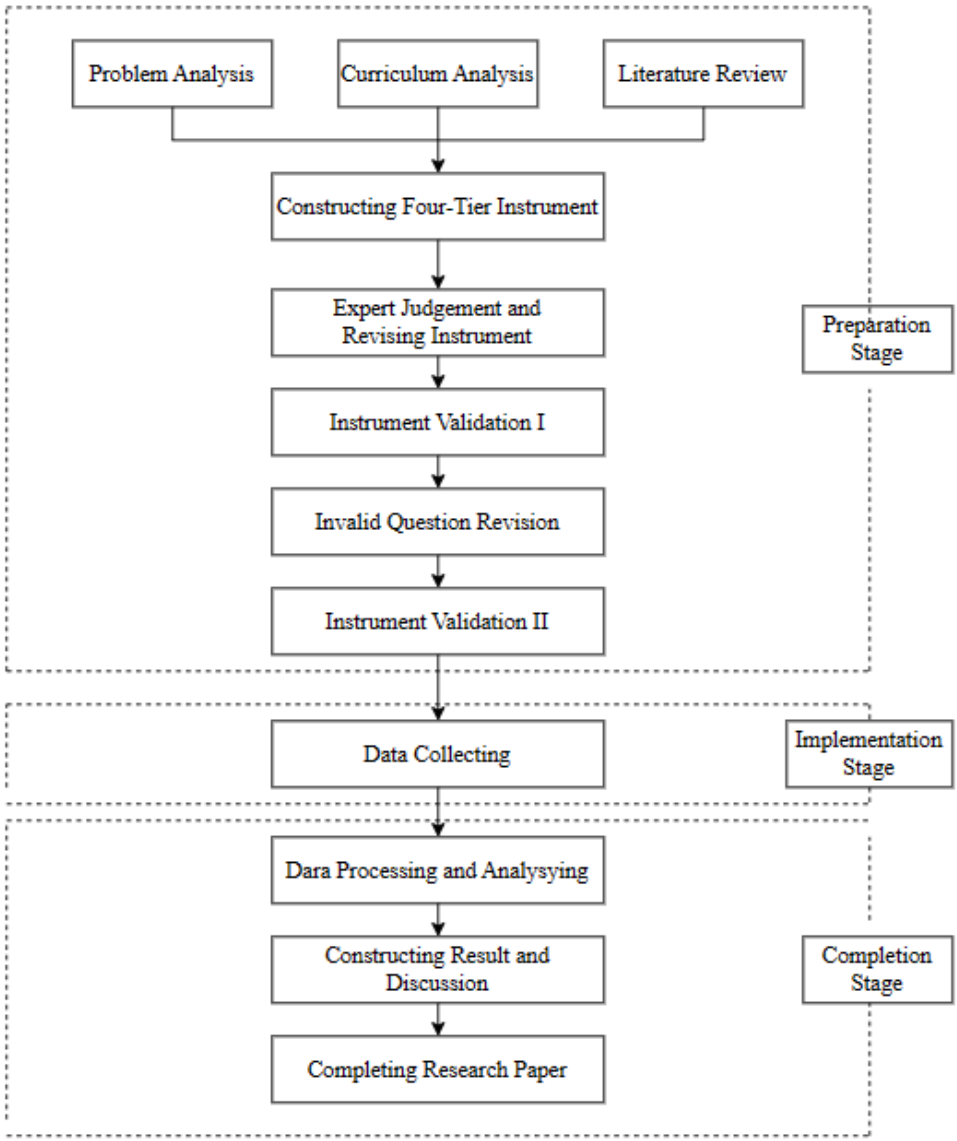

Figure 3.1 Research Procedure