CHAPTER III

RESEARCH METHODOLOGY

3.1 Research Method

Based on the purpose of this research, a developmental research method was used in this research. This method was suitable for this research because according to Richey (1994), developmental research has been defined as the systematic study of designing, developing, and evaluating instructional programs, processes, and products that must meet the criteria of internal consistency and effectiveness.

3.2 Research Design

The method of this study consists of some stages that accordance with Richey and Klein (2005) which are literature review, analysis, product design, product development, product validation, product evaluation. According to Richey (1994), the developmental research also focuses on the impact of the product on the learner. In this study, one group pretest and posttest were conducted to analyze the impact of Android-based interactive multimedia towards students' critical thinking skills.

According to Fraenkel et al. (2012), one group pretest-posttest is a single group that not only measure or observe the effect of treatment but also before treatment. The pretest was conducted to investigate the student's prior knowledge about the matter. Then, the Android-based interactive multimedia was implemented in learning matter. And the last, it ended with a post-test to measure the enhancement of student's critical thinking skills. The questions of post-test were the same as the pre-test, the difference was of the test time. Table 3.1 is the design that used in this research.

Table 3. 1

One Group Pretest-Posttest Design

Pretest	Treatment	Posttest
O_1	X	O_2

(Source: Salkind, 2012)

O_1: Pre-test of student's critical thinking skills

X: Implementation of Android-based interactive multimedia

O_2: Post-test of student's critical thinking skills

3.3 Population and Sample

The subjects of this study were students and experts. There were five experts who validated that application. The experts consist of science expert, media expert, science teacher, and computer teacher.

The population of students in this research was the students of 7th grade Junior High School in Bandung Barat. Students who participated in this research came from "Private Junior High School X" in Bandung Barat. The sample consists of 13 females and 17 males. The average age of the sample is 12-13 years old and they have not learned about matter topic yet. The school uses Bahasa as the communication and delivery language in the teaching-learning process. The school applies National Curriculum which is Curriculum 2013. This school was selected because all of the sample have their own Android smartphone since the interactive multimedia used in this research based on Android.

There were 30 students of 7th grade Junior High School who participated in this research. The students were chosen through convenience sampling technique. Due to the condition when doing this research coincided with a pandemic, the research was conducted through an online system.

3.4 Operational Definition

In order to avoid misconceptions about this research, operational definitions are explained in this research. Those terminologies are explained as follow:

1) Android-based Interactive Multimedia

Android-based Interactive Multimedia is learning media that consist of text, image, audio, video, animation, and game. It was developed using Articulate Storyline 3. The output is an HTML file. Then, the file was converted into an Android application using web2apk builder. The quality of the application was assessed by experts' judgment rubric while the readability of the application was analyzed by students' questionnaire.

2) Expert's Judgement

Expert's judgment assesses the quality of the android-based interactive multimedia. In this research, the expert's judgment rubric was arranged based on the Learning Object Review Instrument (LORI) from the research of Leacock et al. (2017). This rubric consists of six indicators which are mechanical, information

structure, interface, motivation, learning content, and multimedia. The rubric also consists of blank spaces for the suggestions, opinions, or comments.

3) Student's Respond

The student's response analyzes the readability of the android-based interactive multimedia. Student's respond was analyzed using a questionnaire. The indicators of the questionnaire include mobile connectivity, materials, projects, quiz, mobile interface, and multimedia. This instrument also used the Likert scale and ratings. There are 4-rating scales consist of strongly disagree, disagree, agree, and strongly agree.

4) Critical Thinking

Critical thinking aspects included in this research based on Facione's critical thinking indicators consist of interpretation, analysis, evaluation, inference, and explanation. The students' critical thinking skill was analyzed using an objective test which consists of 20 questions of multiple-choice related to matter topic.

3.5 Assumption

The assumption as the foundation in this research as follows:

1) The use of Android-based interactive multimedia is effective to improve learning outcomes in learning chemistry (Hamimi & Sari, 2020).
2) The implementation of Android-based learning media can improve student's critical thinking skills (Rahmawati et al., 2020).

3.6 Hypothesis

The hypothesis that is tested in this research are as follows:

H_0: There is no significant enhancement of students' critical thinking after implementing Android-based interactive multimedia in learning matter.

H_1: There is a significant enhancement of students' critical thinking after implementing Android-based interactive multimedia in learning matter.

3.7 Research Instrument

The instruments are used to gain the data in this research. There are three types of instruments are used in this research which are experts' judgment rubric, questionnaire for students, and objective test. The instruments were used in this research are stated in Table 3.2.

Table 3. 2

Research Instruments

No.	Instrument	Data Obtained
1.	Experts' Judgement Rubric	The quality of interactive multimedia
2.	Students Questionnaire	Student' impression towards interactive multimedia
3.	Objective Test	Students' critical thinking skills

The description of the instruments is described as follows:

3.5.1 Expert's Judgement Rubric

The rubric that used in this research was the rating scale. This rubric was used to validate the quality of the application. This rubric was arranged based on the Learning Object Review Instrument (LORI) from the research of Leacock et al. (2017). LORI is a rule that is often used to measure all kinds of media used in learning. This rubric consists of six indicators which are mechanical, information structure, interface, motivation, learning content, and multimedia. The scale is 1 to 4 to determine the quality of each aspect. The rubric also consists of blank spaces for suggestions or comments.

The scoring result in this rubric is the overall average of all the scores given by the expert. The aspect of this rubric is presented in Table 3.3.

Table 3. 3

The Indicators and Aspects of Expert's Judgement

No.	**Indicator**	**Aspect**	**Score**			
			1	2	3	4
1.	Mechanical	Interaction usability Navigation				
		Grammar/ Visual communication Screen design				
2.	Multimedia	The use of additional media				
3.	Information Structure	Instructions to use the application				
		The sequence of information				
		Attraction of information				
4.	Motivation	Appeal of the application to motivate the students				

No.	Indicator	Aspect	Score			
			1	2	3	4
5.	Interface	Application interface				
		Permissions obtained resources				
		Originally				
6.	Learning Content	The linkage between curriculum learning with media				
		Objective learning				
		The use of scientific term				
		The depth and breadth of learning content				
		The linkage between learning content with critical thinking level				
		The clarity of learning material				
		User survey				

(Source: Leacock et al., 2017)

3.5.2 Student's Questionnaire

The questionnaire is a method of collecting data to obtain information from respondents (Sandu & Ali, 2015). In this research, the questionnaire is used as an instrument to analyze students' impressions after implementing android-based interactive multimedia. The indicators of the questionnaire include mobile connectivity, materials, projects. quiz, mobile interface, and multimedia. According to Hardani et al. (2020), the questionnaire often uses a checklist and rating scale. This tool helps simplify and measure student's attitudes. Usually, to evaluate attitudes and behavior Likert scale is used. There are 4-rating scales consist of strongly disagree, disagree, agree, and strongly agree. Each scale has different point. The scale for the students' questionnaire is shown in Table 3.4.

Table 3. 4

Likert Scale for Students' questionnaire

Scale	Criterion	Score
SD	Strongly Disagree	1
D	Disagree	2
A	Agree	3
SA	Strongly Agree	4

3.5.3 Objective Test

The objective test was used to measure the students' critical thinking skills before and after implementing interactive multimedia in learning matter. This test consists of 20 multiple-choice questions. The questions are related to the matter topic which is limited by core competence 3 and 4 and basic competence 3.3 in curriculum 2013. The questions consist of five types of critical thinking aspects based on Facione (2015) which are interpretation, analysis, inferring, evaluation and explanation. The blueprint of the objective test is presented in Table 3.5

Table 3.5

The Distribution of Test Items

No.	Critical Thinking Skills	Number	Total
1.	Interpretation	1, 2, 3, 4	4
2.	Inference	5, 6. 7. 8	4
3.	Analysis	9, 10, 11, 12	4
4.	Evaluation	13, 14, 15, 16	4
5.	Explanation	17, 18, 19, 20	4
	Total		20

In this research, the objective test was used for pretest and posttest. Pretest was conducted to know students' prior knowledge, while posttest was used to check the enhancement of students' critical thinking skills.

Before the instrument used for the pretest, the questions were judged by the experts and distributed to the students who have been learned about the matter to answer the questions. In order for the research instrument to be appropriate, the questions were checked to know the validity, reliability, difficulty level, distinguish power, and distraction. The data were analyzed by using ANATES. The techniques of question analysis are described below:

3.5.3.1 Validity

Validity is defined as the appropriateness, correctness, meaningfulness, of the evidence that supports any inferences from the scores Fraenkel et al. (2012). Validity shows the extent to which an instrument is able to measure what we need to measure. The instrument must be valid so that the research results can be trusted. Validity could be determined by calculating by the following formula:

$$\gamma pbi = \frac{\mathrm{Mp} - Mi}{\mathrm{St}} \sqrt{\frac{p}{q}}$$

(Kimberlin & Winterstein, 2008)

Where:

γ_{pbi} = Biserial correlation coefficient

M_p = Average of score of objects

M_i = Average of total scores

St = Deviation standard of total proportion score

P = students' proportion with the correct answer

Q = students' proportion with the wrong answer

The category of validity score is represented in the table 3.6.

Table 3. 6

Interpretation of Validity Score

Correlation coefficient	Validity category
0.80-1.00	Very high
0.60-0.80	High
0.40-0.60	Average
0.20-0.40	Low
0.00-0.20	Very Low

(Source: Kimberlin & Winterstein, 2008)

3.5.3.2 Reliability

Reliability refers to the consistency of scores from one administration of an instrument to another, and from one set of items to another Fraenkel et al. (2012). To The number of reliabilities, it is calculated by the formula as follows:

$$KR21 = \frac{k}{k-1}(1 - \frac{\hat{\mu}(k - \hat{\mu})}{k\sigma\hat{2}_x})$$

(Kimberlin & Winterstein, 2008)

Where:

$\hat{\mu}$ = Mean of total score

$\sigma^{\hat{2}}{}_x$ = Total score variance

k = The number of items on the test.

The interpretation of reliability is presented in Table 3.7.

Table 3. 7

Interpretation of Reliability Score

Correlation Coefficient	Reliability Category
0.80-1.00	Very high
0.60-0.79	High
0.40-0.59	Average

Correlation Coefficient	Reliability Category
0.20-0.39	Low
0.00-0.19	Very low

(Source: Kimberlin & Winterstein, 2008)

3.5.3.3 Difficulty Level

The difficulty level can be defined as the proportion of students who correctly answer a test item. The smaller the difficulty, the higher this percentage. This means that it is based on an inverse relationship. The more difficult the test items, the lower its index (Escudero et al., 2000). For measuring the difficulty level, it can be calculated by the following formula below:

$$Pi = \frac{Ai}{Ni}$$

(Escudero et al., 2000)

Where:

Pi = Difficulty index of item *i*

A_i = Number of correct answers to item *i*

N_i = Number of correct answers plus number of incorrect answers to item *i*

The interpretation of difficulty level can be seen in Table 3.8

Table 3. 8

Interpretation of Difficulty Level

Correlation Coefficient	Reliability Category
20 and below	Very difficult
21-40	Difficult
41-60	Average
61-80	Easy
81 and above	Very easy

(Source: Escudero et al., 2000)

3.5.3.4 Discriminating Power

The discriminating power of a test item refers to success or failure on an item indicating possession of the ability being measured by the item (Cohen et al., 2002). According to Escudero et al. (2000), the good test items should discriminate between those who get high test scores and those who get low test scores. There are usually two methods for evaluating test items' discriminating power, which are the

discrimination of coefficient and index. Discriminating can be measured by using the following formula:

$$Di = \frac{GA - GB}{N}$$

(Fraenkel et al., 2012)

Where:

Di = Discrimination index of item i

GA = Number of correct answers to item i among the 27% of those with highest test score

GB = Number of correct answers to item i among the 27% of those with lowest test score

N = Number of persons in the largest group (GA or GB)

The interpretation of discriminating power is presented in Table 3.8 as follows:

Table 3. 9

Interpretation of Discriminating Power

Discrimination Index	**Quality**	**Recommendations**
> 0.39	Excellent	Retain
0.30 – 0.39	good	Possibilities for improvement
0.20 - 0.29	Mediocre	Need to check/review
0.20 – 0.29	Poor	Discard or review in depth
< -0.01	Worst	Definely descard

(Source: Escudero et al., 2000)

The test item was analyzed using ANATES software to gain the data of validity, reliability, discriminating power, and difficulty level. The recapitulation of test item analysis is tabulated in Table 3.10 as follow.

Table 3. 10

Test Item Analysis Result

No.	Discriminating Power		Level of Difficulty		Validity		Note
	Value	Category	Value	Category	Value	Category	Acceptance
1	0.36	Good	0.48	Medium	0.315	Low	Revision
2	0.64	Excellent	0.58	Medium	0.460	Enough	Used
3	0.36	Good	0.64	Medium	0.301	Low	Revision
4	0.86	Excellent	0.54	Medium	0.662	High	Used
5	0.64	Excellent	0.26	Difficult	0.591	Enough	Revision
6	0.79	Excellent	0.62	Medium	0.577	Enough	Used
7	0.57	Excellent	0.66	Medium	0.479	Enough	Used

No.	Discriminating Power		Level of Difficulty		Validity		Note
	Value	Category	Value	Category	Value	Category	Acceptance
8	0.07	Poor	0.62	Medium	0.134	Very Low	Rejected
9	0.36	Good	0.26	Difficult	0.378	Low	Revision
10	0.43	Excellent	0.42	Medium	0.284	Low	Used
11	0.64	Excellent	0.32	Medium	0.499	Enough	Revision
12	0.29	Poor	0.36	Medium	0.335	Low	Revision
13	0.57	Excellent	0.72	Easy	0.591	Enough	Used
14	0.07	Poor	0.36	Medium	0.076	Very Low	Rejected
15	0.50	Excellent	0.60	Medium	0.378	Low	Used
16	0.64	Excellent	0.36	Medium	0.504	Enough	Used
17	0.79	Excellent	0.58	Medium	0.662	High	Used
18	0.57	Excellent	0.36	Medium	0.400	Enough	Used
19	0.50	Excellent	0.62	Medium	0.461	Enough	Used
20	0.71	Excellent	0.66	Medium	0.637	High	Used
21	0.50	Excellent	0.48	Medium	0.415	Enough	Used
22	0.21	Mediocre	0.92	Very easy	0.311	Low	Revision
23	0.36	Mediocre	0.38	Medium	0.328	Low	Revision
24	0.57	Excellent	0.68	Medium	0.530	Enough	Used
25	0.50	Excellent	0.64	Medium	0.431	Enough	Used
26	0.71	Excellent	0.64	Medium	0.651	High	Used
27	0.21	Poor	0.20	Difficult	0.232	Low	Used
28	0.86	Excellent	0.58	Medium	0.750	High	Used
29	0.79	Exccellent	0.44	Medium	0.539	Enough	Used
30	0.64	Excellent	0.54	Medium	0.575	High	Used

3.8 Data Processing Technique

The instrument has a different way to gain and process the data. The data processing techniques are explained as follow:

3.5.1 Objective Test

The objective test was used in pretest and posttest to analyze the effectiveness of the application towards critical thinking skills. The objective test consists of 20 questions of multiple choice. There are some tests to analyze the result of pretest and pretest which are homogeneity test, normality test, hypothesis test, and N-gain test. The tests were analyzed using SPSS version 25. The explanation is elaborated below.

3.5.1.1 Homogeneity Test

The homogeneity test is used to show that two groups of sample data come from populations that have the same variation. The level of significance used was 0.05. It means that if the signification value gain from SPSS is greater than 0.05, the data is homogeneous (Sugiyono, 2013).

3.5.1.2 Normality Test

The normality test is used to determine whether the population comes from a normally distrusted group or not (Fraenkel et al., 2012). The normality test is used to analyze the distribution of data. This test is the prerequisite for deciding the measures of statistical test for data analysis. If the sample is greater than 50, the method used is the Kolmogorov-Smirnov test and if the sample is less than 50, the Shapiro-Wilk test is used. The level of significance used is 0.05. The data is categorized as normally distributed if the significant value is greater than 0.05. If the data is normally distributed, then the data is analyzed using a Parametric test. Meanwhile, if the data is not normally distributed, the data is analyzed using a Non-Parametric test (Mishra et al., 2019).

3.5.1.3 N-Gain Test

N-Gain test is used to analyze the improvement from pretest to posttest. The N-Gain score can be measured by calculating the following formula.

$$< g >= \frac{\text{Postest} - \text{Prestest}}{Maximum\ score - Pretest}$$

The interpretation of N-Gain Score can be seen in this following table.

Table 3. 11

The Interpretation of N-Gain Score

Score interval	Categorize
0.7<g≥1	High
0.3<g>0.7	Medium
0.0<g>0.3	Low

(Source: Hake, 1999)

3.5.2 Likert Scale

Likert Scale is a rating scale which commonly used in questionnaires to measure a person's attitudes, respond and perception (Hardani et al., 2020). The

students' questionnaire was analyzed based on the percentage of the studenst who choose strongly agree, agree, disagree, and strongly disagree. And then, the expert's judgment rubric result was used to analyze the validity of the application based on (lewis. R. Aiken, 1985). The acceptable value of Aiken's V Index is based on an exact binomial test, which is represented by a table that lists the critical values of V based on the number of judges and ratings (lewis. R. Aiken, 1985). In this investigation, the number of judges and evaluations assigned a crucial value of 0.8. The Aiken's V index can be measured by calculating the following formula.

$$V = \frac{\sum S}{[n(c-l)]}$$

$$\sum S = r - l0$$

Where:

V = V index

$\sum S$ = The difference between the judge's validity ratings of the item (r) and the lowest validity category (l)

n = The number of validators

c = The highest validity category

r = The Score that given by the validator

$l0$ = The lowest validity category

3.9 Research Procedure

There are three main stages of procedure in this research which divided into preparation stage, implementation stage, and completion stage that elaborated as follows:

1) Preparation Stage
    a) Identification Problem
    b) Formulating the research objective
    c) Deciding research variables
    d) Reviewing the related literature review
    e) Constructing instruments
    f) Designing storyboard
    g) Designing mobile application
    h) Validating the research instrument and mobile application

i) Revising the research instrument based on experts' judgement
j) Check the readability of instrument

2) Implementation Stage
   a) Giving pre-test to the students
   b) Analyze the result of the pre-test
   c) Implementing interactive multimedia as learning media
   d) Giving questionnaire to students
   e) Analyze student's respond towards the application
   f) Assessing students' critical thinking skills by giving post-test
3) Completion Stage
   a) Collecting and analyzing the data
   b) Making a result and discussion
   c) Drawing the conclusion based on data analysis
   d) Reporting the result of the research

The arrangement of this research can be seen in this following picture.

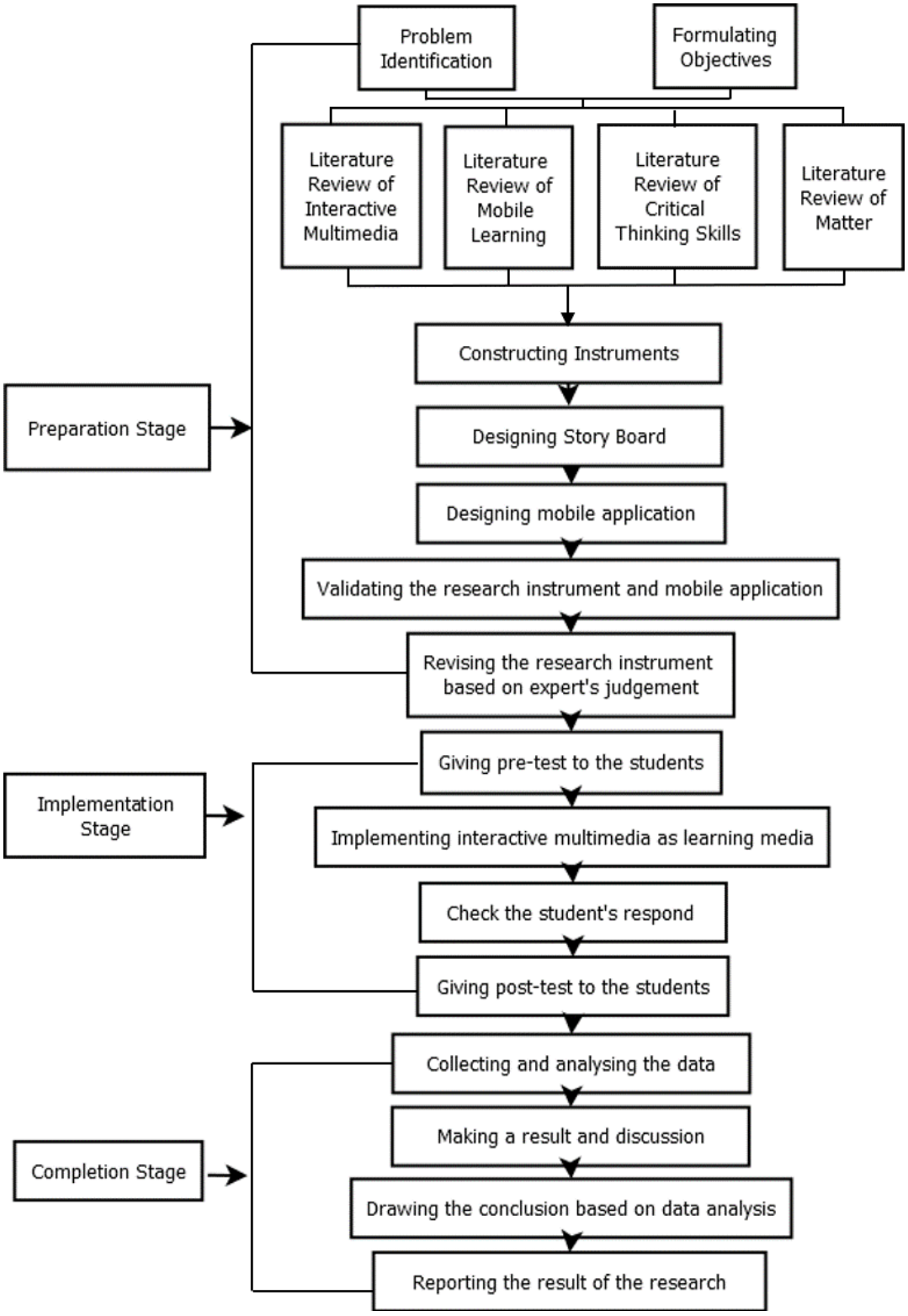

Figure 3. 1 Flowchart of interactive multimedia