CHAPTER III

RESEARCH METHOD

This chapter presents six topics dealing with the research method. Those are research design, population, sample and sampling, data collection method and research instrument, validity and reliability testing, normality and homogeneity testing and data analysis.

A. RESEARCH DESIGN

This research is aimed to examine the effectiveness of animation videos in students' higher order thinking skill (HOTS). Referring to the aim of the research, an experimental research is conducted. As Gay (1992) state that the experimental research is only of the research that can truly test hypotheses concerning cause and effect relationship. In an experimental research, the researcher can manipulate at least one independent variable, control over relevant variables and observes the effect on one or more dependent variable. In addition, Latief (2012) also said that experimental research is a powerful research method to establish cause and effect relationship with involving two or more variables, the variable that becomes the cause (independent) and the variable that becomes the effect (dependent). The independent variable in this research was animation videos, and the dependent variables were students' higher order thinking skill (HOTS) of narrative text.

Quasi-experimental research design was chosen because only assign randomly different treatments in two different classes, and cannot artificially

create groups for the experiment. Creswell (2012), state that quasi-experiments include assignment of participants to groups. The design was chosen for two reasons: First, this study conducted in the organized classroom setting in which the classes were not allowed to rearrange the students or subjects for research. Second, the school schedules which have been arranged by the school cannot disrupt. The variables of the research are shown in figure 2.1.

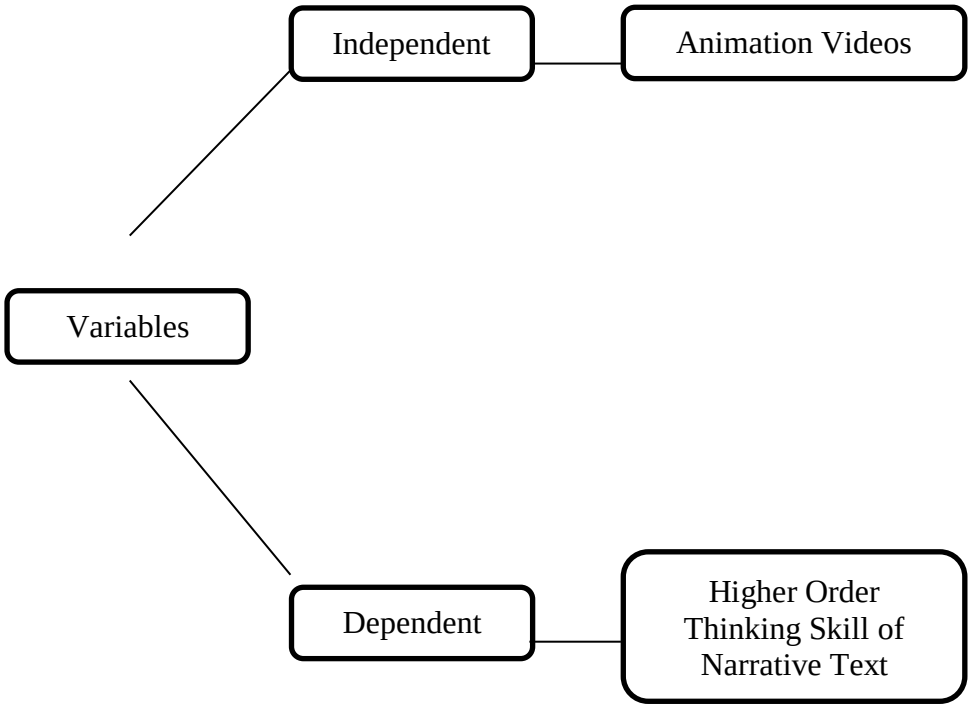

Figure 2.1 Variables of the Research

This design employed non-randomized control group pretest and posttest. Creswell (2012) state that the researcher assigns intact groups of the experimental

and control groups, administers a pre-test to both groups, conduct experimental treatment activities with only the experimental group, and then administers a post-test to assess the differences between the two groups. The pre-test in this research was used to check whether the two groups are equal or not before the treatment, while the post-test in this research was to find out the effectiveness of the strategy employed. The illustration of the non-randomized control group, pretest-posttest design in shown in the table 3.1

Table 3.1 The Nonrandomized Control Group, Pretest-Posttest Design

Group	Pretest	Independent Variable	Posttest
E	Y1	X	Y2
C	Y1	-	Y2

Note:

E: Experimental Group, the group that is taught by using Animation Videos strategy

C: Control Group, the group that is taught by using conventional strategy

Y1: Pretest is given to both groups to measure students' higher order thinking skill achievement prior to treatment

Y2: Posttest is given to both groups to know whether or not applying animation videos and conventional strategy has any effect to enhance students' higher order thinking skill

X: Treatment, applying animation video to the students experimental group

B. VARIABLES OF RESEARCH

The variables of this research consisted of two; the first variable was independent variable "cause" variable animation videos strategy while the dependent variable "effect" variable was Higher Order Thinking Skill (HOTS) of Narrative Text. The independent variable was the strategy of animation videos to the dependent variable students' higher order thinking skill to know the effectiveness of using this strategy can increase the students' higher order thinking skill of narrative text.

C. POPULATION SAMPLE AND SAMPLING

In this part, there were population, sample and sampling.

1. Population

The population of this research was the 8th grade students of Islamic Junior High School Ngunut Tulungagung at the first semester in 2019-2020 academic years. The 8th grade students were chosen because narrative text is the material that must be taught. There were 7 classes of the 8th grade students, which total number was 228 students. The male and female students were in the different class. Class A until D were male students and E until H were female. According to Gay (1992), state that population is a large number of groups to which is given a treatment by a researcher who's the result would be generalized.

2. Sample

The researcher choosed 2 classes as the sample. Therefore, this research used quasi experimental design, in which the researcher did not

assign the subject randomly. The researcher selects the sample by using non-probability sampling with purposive sampling form, because the population have no chance of selection or the sampling don't have any chance to be selected, so the researcher takes all the sample both two groups

According to Gay (1992), state that sample is the individual selected comprise. Selection of a sample is very important step in conducting a research study. The researcher determines the sample with the selection method on the knowledge, expertise and judgment. The classes are 8G and 8H with total number of two groups are 76 students selected to be a sample. After the researcher know about the real condition of the students. The researcher decided is 8H assigned as the experimental group and the 8G become control group.

3. Sampling

Gay (1992) states that sampling is the process of selecting a number of individuals for a research in such a way that the individuals represent the larger group from which they were selected. The researcher uses nonprobability sampling with the purposive sampling type. Ary (2010:149) stated that nonprobability sampling includes methods of selection in which elements are not chosen by chance procedures. It used when the application of probability sampling is feasible. The purposive sampling uses in this research because the researcher consciously selects subjects based on the consideration and purpose of the research herself.

D. DATA COLLECTION METHOD AND RESEARCH INSTRUMENT

In this part, the researcher states the data collection method and research instrument.

1. Data Collection Method

Tanzeh (2011) states that data collection method is a systematical and standard procedure used to collect data that is needed. In this research, the researcher collecting the data through administering test. It means that the researcher administers the test in the form of subjective test. In this test consist of six items. The items based on the category of Bloom's Taxonomy Revision. They are (1) Remembering, (2) Understanding, (3) Applying, (4) Analyzing, (5) Evaluating, and (6) Creating. The technique of collecting data is clarified as follows:

a) Pre test

The researcher conducts twice test both two groups. On the first test calls pre-test, it was followed by 76 students. The researcher allocated 90 minutes for conducts pre-test. The test is in the form of subjective test, because it is suitable for testing higher order thinking skill. The test contains 6 questions for HOTS test. It was done before treatment process by using animation videos and non-animation videos strategy. The test gives to them to know the prior knowledge of higher order thinking skill (HOTS).

b) Treatment

The treatment gives with different treatment both two groups. The experimental research gives treatment by using animation videos while the

control group gives treatment by using non-animation videos strategy. The researcher conducts the treatment in the classroom. The treatment conducted sixth meetings. The researcher wanted to measure the effectiveness of using animation video strategy really to help increasing the students' higher order thinking skill.

Before conducted the strategy, the researcher prepared the material and students' worksheet to support the treatment. The material was in the form of animation video of narrative especially of fable consisted about four videos with different title. The researcher took the material from YouTube.

In conducting the research, the researcher was helped by the English teacher of SMP Islam SGJ to handled control group, while the experiment group handled by the researcher herself. Both two groups needed six meeting to conduct treatment with different worksheet.

The first treatment for experimental group, the researcher applied the first part of animation videos strategy that was called *Pre-Task*. On this part, the researcher conducted brainstorming by asking the students question about what the topic that we wanted to discuss, the researcher introduced and explained the material about animation videos.

The second treatment, the researcher applied the second part of animation videos strategy that was called *Task-Cycle*. In this part were many activities that should be done by the students. The activities can be categorized into six. The *first* is remembering, the students are able to measure that the students can recall or remember the information. The *second* is understanding, the

students are able to measure that the students can explain the concept, principle, law or procedure. The third is Applying, the students are able to measure that the students can apply their understanding in new situation. The *fourth* is Analyzing, to measure that the students can classify the sections based on their difference and similarity. The *fifth* is evaluating, to measure that the students can state either good or bad towards a phenomenon or certain object. The last category is creating; the students are able to measure that the students can create a thing or opinion.

The third treatment, in this part was the last part of animation video strategy that was called *Language Focus*. The researcher encouraged the students to find their language problem that they encounter during task-cycle, while the students consulted their language problem that they encounter during the previous part. Next, the researcher guided the students to make reflection by giving feedback based on the lesson.

c) Post test

The researcher administer post-test after the students get treatment by using animation videos strategy. The researcher conducts post-test both of two groups, to know whether or not applying animation videos and non-animation videos strategy has any effect to enhance students' higher order thinking skill (HOTS). The result of posttests was compared to see whether the experimental group significantly out performed the control group.

In this research, the researcher used the data in the form of students' higher order thinking skill scores. The researcher got the data after administer

pretest and posttest both two groups. Administering the test result both two groups and scoring it with scoring rubrics. The researcher got the posttest from habitual activities in each meeting. The researcher gathered all scores and divided them based on the meetings. The researcher used type of achievement test, means that the test should be representative of structure and skill that will be tested then the test must be appropriate with the grade. Here, the test used to measure the students' higher order thinking skill, and the form of test was subjective test. Addition, the test is suitable for eight grades because the content of the test refers to syllabus for eight grades of Junior High School. The test was constructed by researcher herself. In other word, the test is called Researcher-made test, it means that the test was arranged by the researcher.

2. Research Instrument

According to Gay (1992) is a tool of to measure a knowledge skill, feeling, intelligence of an individual of group. The researcher used test to measure the students' higher order thinking skill. The researcher used type of achievement test, means that the test should be representative of structure and skill that will be tested then the test must be appropriate with the eighth grade.

Table 3.2 Instruments

No	Instrument	Variable to measure	function
1.	HOTS (Higher Order Thinking Skill) in Reading (pre-test)	Students' higher order thinking skill before treatment	To see the homogeneity
2.	HOTS (Higher Order Thinking Skill) in	Students' higher order thinking skill after	To test hypothesis

	Reading (post-test)	treatment	

Here, the test used to measure the students' higher order thinking skill, and the form of test is subjective test can measure skill and component that will be tested. Addition, the test is suitable for eight grades because the content of the test refers to syllabus for eight grades of Junior High School. The test is constructed by researcher herself. In other word, the test is called Researcher-made test, it means that the test was arranged by the researcher.

E. VALIDITY AND RELIABILITY TESTING

1. Validity

Latif (2010:236) stated that the objective of the validation was to analyze the test validity to enable the instrument to result in valid data. The researcher uses subjective test. It means that form both of two groups can measure the skill and component of students. To measure the test has a good validity, the researcher analyzed the test from *content and construct validity*.

a) Face Validity

Face validity if it looks as it measures what it is supposed measure. For example, a test which pretended to measure reading HOTS ability but, which did not require the test-takers to do reading HOTS might be through to lack face validity. This is true even if the test is constructing and criterion-related validity can be demonstrated. Face validity is hardly a scientific concept, yet it is very important. A test which does not have face validity may not be acceptable by test-takers, teachers, education

authorities, and employers. The researcher used face validity by consulting with the advisor and teacher.

In order to get face validity, prototype of the reading HOTS, test blueprint and expert validation form, is given to the expert to get judgment whether the test looks right to measure student's reading higher order thinking skill.

b) Content Validity

Latief (2016), state that content validity is concerned with the coverage of the materials will be measured or being tested. Besides, the content validity represents the test items in the test that cover and represent the material in the curriculum. In this research, validating the content validity conduct by analyzing the content of the test and the materials required in the English syllabus revision K13.The purpose of analyzing is whether the content of the test represents the reading materials in English syllabus.

The description of the test items used in reading higher order thinking skill can be clearly seen in **Table 3.3**

Objectives	**Category**	**Indicators**	**Number of Items**
To measure that the students can recall or remember the information	Remembering	Students are able to mention five points about the story	1
To measure that the students can explain the concept,	Understanding	Students are able to summarize the story at least two	1

principle, law or procedure		paragraphs	
To measure that the students can apply their understanding in new situation	Applying	The students can read what they summarized	1
To measure that the students can classify the sections based on their difference and similarity	Analyzing	The students can compare the characters of one player to another	1
To measure that the students can state either good or bad towards a phenomenon or certain object	Evaluating	The students can give comment to their friends' task	1
To measure that the students can create a thing or opinion	Creating	The students can retell the story by their own language	1

The reading higher order thinking skill was untilized in experimental and control groups. This procedure covers several steps, they are the purpose recognition, establishment of the test blueprint, devising the test items, expert review, and revision, try out test, and analysis and revision.

The specification on the test included the objective, general instruction, specific instructional, kind of test, test type, number of text, source text,

number of items, time allocation, equipment, and scoring. The components of blueprint include the macro skill, the objective, the category, the item types, the indicators, and the item number. In the devising the items is essay forms were constructed to the reading higher order thinking skill in this study. There were six questions with consist of remembering, understanding, applying, analyzing, evaluating, and creating category.

The next step was expert validation. The expert is one of lecturers in IAIN Tulungagung and other expert is one of English teacher of SMP Islam SGJ Ngunut who has specialized in reading higher order thinking skill. The expert checked whether the tests possesses evidence that meets the criteria of a good test.

Before the real test was given, the try-out of the test was done. The pilot testing was conducted on the particular subject since it has many chracteristics in common with the main subject of the resaerch. Further, the score was dissected to know the item facility/difficulty, items discrimination, reliability and the efficiency of the distractor. After trying out the test, the test items were anlyzed based on the students' score. The test items analysis covers anlysis of item reliability, analysis of items difficulty, analysis of item discrimination, anlysis of item validity.

c) Construct validity

Latief (2016), state that construct validity is the validity concerned with the theoretical construct will be measured. A test is considered construct validity if the items of test measure each of thinking aspect from a variable will

be measured through the test. In this research, validating the construct validity conduct by analyzing the objective of the test and the type in which the students ask to do the task.

Since the students' ability on reading higher order thinking skills are measured, the test must give in the reading used animation video. The construct validity evidence can be seen in **Table 3.4**.

Objective	**Type of test**	**Task**
Measuring the students' higher order thinking skill	Reading higher order thinking skill	Students are able to answer the reading higher order thinking skill

d) Criterion-related validity

Criterion related validity applied to know how far results on the test agree with those provided by some independent and highly dependable assessment of the candidate's ability. This independent assessment is thus the criterion measure against hich the test is validated. There are essentially two kinds of criterion-related validity: concurent validity and predictive validity. (Hughes, 2002)

The researchers use *predictive validity*. The predictive validity this concern the degree to which a test can predict candidates' future performance. An example would be how well a proficiency test could predict a student's ability to cope with a graduation at SMP Islam SGJ

Ngunut. The criterion measure here might be an assessment of the student's English as perceived by his or her teacher or researchers at SMP Islam SGJ Ngunut.

To apply this validity, the developer or the researcher might administer a certain test before the students begun the material about narrative text used animation videos explained by the researcher. After several times, the same group of students might take the same test and the scores, resulted from the first score until the last score are calculated for the correlation coefficient. The closer the correlation, to know the stronger the relationship among scores to predict the students' future. In this research, used *Pearson Product Correlation Coefficient (PPMC)* though SPSS 25.0 version to find the correlation coefficient among scores. The correlation among scores of reading higher order thinking skill can be seen from the table:

Table: 3.5 Result of Criterion-related validity (Predictive Validity)

Correlations

		Tryout1	Tryout2
Tryout1	Pearson Correlation	1	.996**
	Sig. (2-tailed)		.000
	N	34	34
Tryout2	Pearson Correlation	.996**	1
	Sig. (2-tailed)	.000	
	N	34	34

**. Correlation is significant at the 0.01 level (2-tailed).

The SPSS output suggests that the correlation coefficient is was 0,996. It means that there is a positive correlation between variables. It also suggest that the ρ-value is 0.000. Considering that 0.000 is smaller than 0.01, so the null hypothesis is rejected.

2. **Reliability**

Gay (1992) states that reliability is the degree to which a test consistently measures whatever it measures. Reliability is expressed numerically, usually as coefficient; a high coefficient indicates high reliability. Then is significant difference between the score of pre-test and the score of post-test both groups, so the result of research is reliable. To measure the reliability of the test, the researcher uses *Cronbachs' Alpha*, if the result of *Cronbachs' alpha* was higher than 0.05 (reliable index < 0.05), it means that the test was reliable.

To measure the reliability of the test, the researcher uses *Cronbachs' Alpha*, if the result of *Cronbachs' alpha* was higher than 0.05 (reliable index < 0.05), it means that the test was reliable.

Actually, the ideal test should be both reliable and valid. In this research, the researcher also used SPSS 25.0 for window to know the reliability of test instruments. The criteria of reliability instrument can be divided into 5 classes as follows (Ridwan, 2004), those are:

1. If the *alpha cronbach* score 0.00 – 0.20: less reliable
2. If the *alpha cronbach* score 0.21 – 0.40: rather reliable
3. If the *alpha cronbach* score 0.41 – 0.60: enough reliable

4. If the *alpha cronbach* score 0.61 - 0.80: reliable
5. If the *alpha cronbach* score 0.81 – 1.00: very reliable

The result of reliability testing of reading higher order thinking skill by using SPSS 25.0 can be seen from the table:

Table: 3. 6 Result of Reliability

Reliability Statistics

Cronbach's Alpha	N of Items
.997	2

To know the items is reliable or not it can be seen from Cronbach's Alpha column. The Cronbach's Alpha score = 0,997 means that it is reliable.

F. NORMALITY AND HOMOGENEITY TESTING

1. Normality

In this research works with statistic nonparametric to analyze the hypothesis. In statistic nonparametric the data that will be analyzed should in normal distribution. The technique that can be used to test normality by using *One- Sample Kolmogorov-Smirnove test* by the value of significance (α) = 0.05 rules as follow:

a. H_0 : If the value of significance > 0.05, means data is normal distribution
b. H_1 : If the value of significance < 0.05, means the distribution data is not normal distribution.

Table: 3.7 The Statistical Correlations of Tryout Score of Reading Higher Order Thinking Skill (HOTS)

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	GROUP	Statistic	Df	Sig.	Statistic	df	Sig.
READING HOTS	TEST 1	.172	34	.013	.926	34	.764
	TEST 2	.176	34	.009	.907	34	.892

a. Lilliefors Significance Correction

Based on the result of pretest and posttest in normality testing above, it is known that the significance of Test 1 is 0,764 and the significance value of Test 2 is 0,892. So, it can be conclude that test has normal distribution. To fulfill the provision of normal distribution is if the significance value or probability > 0,05.

2. Homogeneity

Homogeneity test intended to know whether the variance of data is homogeneous or not. In this research the researcher wants to know the variance score in class (group) sample with employs *Levene's statistic* by the value of significance (α) = 0.05. Before doing homogeneity testing, the researcher decides hypothesis in this homogeneity as follow:

a. H_0 : If the value of significance > 0.05, means data is homogeny

b. H_1 : If the value of significance < 0.05, means data is not homogeny

Table 3.8 The Statistical Correlations of Tryout Score of Reading Higher Order Thinking Skill (HOTS)

Test of Homogeneity of Variances

READING HOTS

Levene Statistic	df1	df2	Sig.
.238	1	66	.627

Based on the table above is known that the sig/p value is 0.627 higher than 0.05 means H0 can not rejected and Ha can rejected. So, it can be interpreted that the data is homogeny.

G. DATA ANALYSIS

In this research, to gain the difference between the experimental and the control groups *Independent Sample t-test* was used. Lodico (2010), states that independent sample t-test involves examination of the significant differences on one factor or dimension (dependent variable) between means of two experimental groups (control group vs. experimental group). *The Independent Sample t-test* was used to compare the means of one variable for two groups of cases, and it was used when the data were already normal and homogeneous. For practicality SPSS program 25.0 version was used.

There were one research hypotheses that should be tested. The hypothesis is the students' higher order thinking skill had better improved achievement significantly by using animation video those improved with using conventional strategy.

To test the hypothesis, the hypothesis was transformed into null hypothesis (symbolized $H_{0)}$. There is one null hypothesis. The Ho is the students' higher order thinking skill improved by conventional strategy.

To reject the null hypothesis the researcher stated the alternative hypothesis (symbolized Ha). There is one alternative hypothesis. The Ha is the students' higher order thinking skill improved by animation videos is better than that improved by conventional strategy.

The null hypotheses will be not rejected if the result of SPSS shows that the obtained significant level is higher than or equal to the level of significant .05 ($\rho>\alpha$). However, if the ρ-value is smaller than or equal to the level of significance .05 ($\rho<\alpha$) the null hypotheses will be rejected.