

CHAPTER III

RESEARCH METHOD

In this chapter, several points regarding the methodology of the research are elaborated. It covers research problem, research design, research site, participants, instruments, variables and hypothesis, stages in collecting data, and method of data analysis.

3.1 Research Problem

This study was intended to identify the effects of teaching reading using Islamic texts toward students reading comprehension. The participants were the nine graders of Islamic middle school in Bandung. Moreover, it was aimed to find out the effects of teaching reading using Islamic texts toward the students reading comprehension and to find out the students attitude toward the Islamic reading texts given to them.

3.2 Research Design

The study is aimed to find the effect of teaching reading using Islamic text on the students reading comprehension. It was done by examining the students reading comprehension before and after the treatment to see a significant difference of reading comprehension between students who are taught by using general texts and Islamic texts. It was conducted based on the mix method using the concurrent embedded design (figure 3.1). It is the use of one data collection phase, during which both quantitative and qualitative data are collected simultaneously. Cresswell (2009; 2012) explained that this approach has a primary method which guides the research data collection and a secondary database that provide a supporting role in the procedures. Without any priority, the secondary method (quantitative or qualitative) is embedded, enclosed, within the prominent method

(qualitative or quantitative). The mixing of the data from the two methods is to integrate the information and compare one data source with the other, and to be accomplished on a discussion session of a study or to be reside side by side as two different pictures that provide an overall composite assessment of the problem.

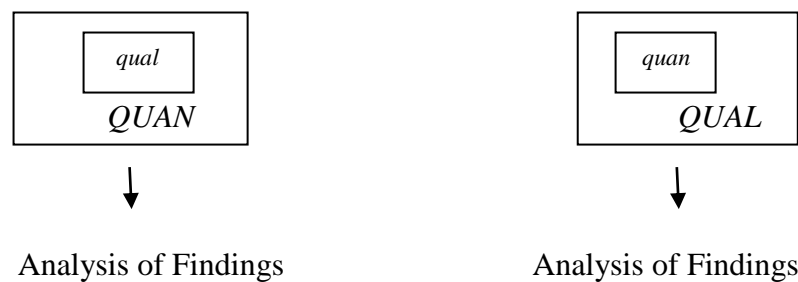


Figure 3.1 Concurrent Embedded Design, Source: Cresswell (2009)

In this study, a descriptive statistic and the quasi experimental design which applied pre-test post-test control and experiment group were used to answer the first research question on exploring the effect of using Islamic texts on the students' reading comprehension. Meanwhile, qualitative method was also applied to give supporting data to answer the research question. Subsequently, questionnaire was used and analyzed quantitatively to answer the second research question. It was given to the experiment group to see their attitude toward Islamic reading texts given to them.

To answer the first research question, two groups of ninth grade of Islamic Middle School consisting of 30 students each were randomly assigned. One class was the control group and one was the experiment group (Hatch and Lazaraton, 1991, cited in Fitriyani, 2008). The experiment group was given the Islamic texts, while the control group was given the general ones. Before and after the treatments, both groups were tested using the same reading comprehension test. The design of the study is as follows:

G1 (random) T1XT2

G2 (random) T1_T2

G1= Experiment

G2= Control

T1= pretest

T2=posttest

X=treatment

In this design, the pre-test (T1) was administered in the two groups to collect data based on their reading comprehension scores and to select homogeneous subjects. At the next stage, the treatment was applied for the experimental group, while the control group followed their usual class routine with general narrative texts. Then, the two groups took the post-test. The aim of this stage was to compare the performance of the two groups in this phase and to check null hypothesis. The two groups were also being interviewed to have secondary data to support the first finding. The data was analyzed and compared with the student post-test result to see if there is any correlation on why the Islamic texts have such effect toward the students' reading comprehension.

Meanwhile, to answer the second research question, the students in the experiment group were given questionnaire to find out their attitudes toward Islamic texts, covering three aspects: cognitive, affective, and behavioral. The cognitive aspect is sought through six items of the questionnaire regarding the students' choice on type(s) of reading texts which considered ideal to be used by Islamic middle school students. The affective aspect, through six items of the questionnaire, is intended to find out whether they feel that Islamic reading texts can help them improving their comprehension ability. Finally, the behavioral aspect is studied through four items to know whether their interest in reading more English texts and in learning English in general are increased by the use of texts (Fitriyani, 2008).

3.3 Research Site

The study was conducted in a private Islamic middle school in Bandung, West Java. It consisted of 12 classes with about more or less thirty students in each class. There were four classes of VII grade, four classes of VIII grade, and four classes of IX grade.

Similar to other Islamic middle school, this school gave more Islamic subjects to the students, such as Qur'an, Hadith, Fiqh, Aqidah, Akhlaq and Arabic, than it was in general middle schools. However, the English subject was given with the same amount of credits as given in general middle schools. In related to academic competencies, especially of English competencies, in general, the students were said to be in middle level competence, to be similar to other common middle schools' students in Indonesia. However, a few classes there were designed to be special classes which consist of the best students in terms of their achievement in most of the subjects.

3.4 Research Participants

Based on the focus of this study, the participants of this research were the students of a private Islamic middle school in IX grade. The Islamic middle school chosen was the one in which the researcher has an access to do the research. To choose the participants, all the four classes of the IX grade were given a pre-test, and two classes which the pre-test scores of reading comprehension were normal and homogenous were chosen

3.5 Research Instruments

In conducting this research, some instruments are used to collect the data. The instruments for the first research questions are pre-test and post-test of reading comprehension, covering literal and inferential comprehension as required by KTSP 2006. The reading comprehension test for pre-test has similar level of

validity, reliability, discriminating power and level of difficulty with the reading comprehension test for posttest.

Both pre-test and post-test consist of only general reading text groups in order to avoid bias, i.e. in case that the experiment group gets higher score than the control group in posttest, it can be because of familiarity of the experiment group with the Islamic narrative texts prior the posttest, and not merely because of the treatment. (Fitriyani, 2008)

Both the pre-test and post-test were consisted of 50 items. Before they were used, the reading comprehension test were tried out to 40 students which had similar characteristics (Appendix 1 and 7). However, not all items can be used because some items were either not good in discriminating power, too easy or too difficult, thus, the number of items were dropped into 30 for each tests. The scoring of those test ranges from 0 to 10, which means each correct item is scored 0.25. Thus, when the students can answer all the items correctly, she or he gets 10. The result of the post-test were then analyzed and compared with the data from interview. The interview was conducted after the posttest, and was employed to both groups.

To answer the second research question, a set of attitude questionnaire was used, it was consisted of 16 items (Appendix 10). The set was given to the students involved in the experimental group. The set contains of questions asking the students' responses toward Islamic narrative texts within the three aspects mentioned earlier.

The questionnaire was adopted from the questionnaires used in Fitriyani's (2008) study, which she developed by following the guidelines from Oppenheim (1982). The questionnaire was chosen due to the similarity of the research objectives of her previous study with this study. The scale had been administered on her study and its validity was ensured. The types of the questionnaires is Likert-Scale with five options: Strongly disagree, Disagree, Uncertain, Agree, and Strongly Agree. It is stated in Oppenheim (1982) that this type of scale is less laborious and often achieved a reliability coefficient of .85.

Oppenheim (1982) regarding designing attitude statements further explains that the odd items should be positive statements while the even items are negative statements. It means item number one is the opposite statement of item number two. It means there are 8 pairs of items in the questionnaire. This composition is aimed to make sure the students' consistency on their responses to the questionnaire. The scoring for positive items is 1 for Strongly disagree, 2 for Disagree, 3 for Uncertain, 4 for Agree, and 5 for Strongly Agree. While for the negative statements the scoring is reversed to 5 for Strongly disagree, 4 for Disagree, 3 for Uncertain, 2 for Agree, and 1 for Strongly Agree. The higher the total score results the more positive the students' attitude toward the issue being addressed.

Fitriyani (2008), see also Abidin et al. (2012) explained that attitude questionnaires covered three aspects of attitudes which are cognitive, affective, and behavioral. The first aspect which consists of 6 items (number 1-6) were intended to find out students' belief and insight on which type of reading text is suitable for them the most. The second aspect (number 7-12), was aimed to find out the students' affective on whether a particular type of reading texts interest them better and whether it increases their reading comprehension. The last aspect (number 13-16) were intended to find whether a particular type of reading texts helps increasing their interests in reading English texts and in learning English in general.

3.5.1 Pre-test, Post-test, Narrative Texts and Questionnaire

3.5.1.1 The Selection of the Items in Pre-test, post-test and Narrative texts.

For the pre-test and post-test instruments, the multiple choice form is chosen for several reasons. It is chosen as it is considered to be practical. Brown and Abeywicakrama (2010) describes practical test as stays within budgetary limit, can be completed by the test-taker within appropriate time constraints, has clear directions for administration, appropriately utilizes available human

resources, does not exceed available material resources, and consider the time and effort involved for both design and scoring. Moreover, multiple choice form can test knowledge quickly within large groups and be used to provide quick feedback. It can also be analyzed with regard to difficulty and discrimination. (Epstein et al, 2002; Higgins and Tatham, 2003; Kuechler and Simkin, 2003).

The items of both pre-test and post-test and the general narrative texts to be given to the control group were carefully selected from some standardized textbooks. While the Islamic general and narrative texts to be presented in experiment group were also carefully selected and modified from those available in the internet. The general texts were retrieved from <http://englishstories12.blogspot.co.id>, and the Islamic texts were retrieved from <https://storiesformuslimkids.wordpress.com>. Since the texts are authentic and they aren't written for the purpose of teaching English, there are several things that must be considered in choosing the authentic materials for reading texts. Teacher must consider the suitability, exploitability, readability, and the presentation of the content. Suitability refers to whether the texts interest the students, relevant to their need, and represent the type of material that the students will use outside the classroom. Exploitability refers to whether the texts can be exploited for teaching purposes, what purpose should the text be exploited, and that skills/strategies can be developed by exploiting the text. Readability refers to whether the texts are too easy/difficult for the student, structurally too demanding/complex, and contain new and relevant vocabulary. Meanwhile presentation refers to whether the texts "look" authentic, "attractive", grab the student's attention and make him want to read more (Berardo, 2006). Thus, both the general and Islamic reading texts used in this study had been carefully chosen in regarding to their suitability, exploitability, readability, and presentation.

3.5.1.2 Validity of the Pre-test, Post-test, Narrative texts and Questionnaires

According to Brown and Abeywicakrama (2010), a test is considered effective, appropriate, and useful or so called "a good test" when they have the five criteria which are practicality, reliability, validity, authenticity, and washback. In terms of

validity, Fitriyani (2008) conveys that according to Heaton (1988), there are several types of validity. Some of them which are related with her study meaning also related to this study, are face validity, content validity, and external validity. Mousavi (2009, cited in Brown & Abeywicakrama, 2010) stated that face validity refers to the degree to which a test *looks* right, and *appears* to measure the knowledge or abilities it claims to measure, based on the subjective judgment of the examinees who take it, the administrative personnel who decide on its use, and other psychometrically unsophisticated observers”. Bachman (1990) in Brown and Abeywicakrama (2010, p. 35) said that fair tests can be made by using:

- a well-constructed, expected format with familiar tasks
- tasks that can be accomplished within an allotted time limit
- items that are clear and uncomplicated
- directions that are crystal clear
- tasks that have been rehearsed in their previous course work
- tasks that relate to their course work (content validity)
- a difficulty level that presents a reasonable challenge

However, the pre-test and post-test used in this study have been designed in accordance to those items above. Thus, the face validity has been fulfilled.

The content validity or content-related validity is usually observational and can usually be identified if the test maker can clearly define the achievement that is about to be measured. In other words, content validity means the test should measure what is supposed to be measure. It has to be related to their course work and asses real course objectives (Brown & Abeywicakrama, 2010). In this study, the content of the pre-test and post-test is relevant and is constructed based on the curriculum and the course objectives which is to measure the students' comprehension on narrative texts. Thus, it can be said that the pre-test and post-test have content validity.

The external validity of this study is reached by correlating the Try Out results with the Try Out participants' English score they have gained in the first

mid semester. Here the Pearson Product-moment Correlation formula is used (Hatch & Lazaraton, 1991).

The Spearman Brown Formula is:

$$r_{xy} = \frac{N\sum xy - (\sum x)(\sum y)}{\sqrt{\{N\sum X^2 - (\sum x)^2\} \{N\sum Y^2 - (\sum y)^2\}}}$$

The computation shows the result of .77 for pre-test and .82 for posttest, which means that the tests are valid since r table with α .01 and df 28 is .4026 (Appendix 3 and 9). The Islamic narrative texts and general reading texts are also validated. They have been consulted to an expert of material development in Postgraduate School of Indonesia University of Education and the texts have been stated to be valid.

Concerning the questionnaire for the experiment group, it was adopted from Fitriyani (2008) and is considered valid as she stated that it has been tried out to similar population where the results showed that all of the 16 items are significant and have tobs higher than table with df =8, α =.05 is 2.306 .

3.5.1.2 Reliability of the Pretest, Posttest, and Questionnaires

The reliability of the pre-test and post-test items was confirmed with the Spearman Brown Formula. The measurement used reliability index using split-half technique. The tried-out scores were first calculated and the fifty items were divided into two parts. Then, each of the students score were analyzed. Finally, the total score of first half and second half, together with the total of squared first and second score were put in the Pearson Product Moment formula. The result was then put into the Spearman Brown Formula to know the overall reliability (Hatch and Lazaraton, 1991).

The Spearman –Brown Formula is:

$$r_{\text{total test}} = \frac{2r_{\text{split half}}}{1 + r_{\text{split half}}}$$

From the calculation, it is shown that the reliability index is .84 for pre-test and .84 for posttest, while r_{table} with $df = 38$ and $\alpha = .05$ is 0,3120 (Appendix 2 and 8). It means that the r of the test is greater than the table. Thus, the items are reliable.

The questionnaire for experiment group that was adopted from Fitriyani (2008) was also considered reliable as she stated that its internal consistency had been analyzed. It is reliable as the r index of the questionnaire for group is .822, which means that the questionnaire has very high reliability index.

3.5.1.3 pre-test and post-test Items Difficulty

The items difficulty indexes were also calculated using a formula of Gronlund, (1982):

$$P = \frac{R}{T} \times 100$$

where P = the percentage of who answered the item correctly; R = the number who answered the item correctly, and T = the total number who tried the item.

Meanwhile the criteria used are:

$P = 0.00$ = too difficult
 $0.00 < P \leq 0.30$ = difficult
 $0.30 < P \leq 0.70$ = medium
 $0.70 < P < 1.00$ = easy
 $P = 1.00$ = too easy

After being tried out, the analysis showed that all the 50 items pre-test questions' could actually being used, but it was not final yet since the result of the items discriminating power should be considered. The items difficulty of the pre-test consisted of 8 difficult items, 24 medium items, and 18 easy items.

Meanwhile, the items difficulty of the post-test consisted of 4 difficult items, 29 medium items, and 17 easy items. (Appendix 4 and 10)

3.5.1.4 Discriminating power

The indexes of items discriminating power were calculated using a formula from Gronlund (1982) too. The formula used for analyzing discriminating power is:

$$D = \frac{RU - RL}{\frac{1}{2} T}$$

where D = the index of discriminating power; RU = the number in the upper group who answer the item correctly; RL = the number in the lower group who answer the item correctly; $\frac{1}{2} T$ = a half of the total number of students included in the item analysis.

Meanwhile the criteria used are:

- D = 0.00 = very bad
- 0.00 < D ≤ 0.20 = bad
- 0.20 < D ≤ 0.40 = medium
- 0.40 < D ≤ 0.70 = easy
- 0.70 < D ≤ 1.00 too easy

The analysis shows that from 50 items for pretest, 1 item has very good discriminating power, 15 items are good, 14 items are medium, 10 items are bad, and 10 items are very bad. So, of all there are only 30 items can be chosen. Meanwhile from 50 items of the posttest, 5 items are good, 25 items are bad, 9 items are bad, and eleven items are very bad (appendix 6 and 12). So, of all, there are also only 30 items can be chosen.

Having analyzed the try out results of both pre-test and posttest, 40 items were selected for each test. The items selected are those which have difficulty level range from easy, medium, and difficult and level of discriminating power range from very good, good, and medium, with the same composition of this range on both pre-test and posttest.

3.5.2 Interview

In collecting qualitative data interviews, it is important to consider what form of interviewing will best help the researcher to understand the central phenomenon and answer the questions in the study. There are a number of approaches to interviewing which include using open-ended questions on questionnaires (Cresswell, 2012). Thus, it is important to consider the interview approach to use based on the accessibility of individuals, the cost, and the amount of time available. As to the limitation of the time on the study and also the availability of the participants, the use of focus group interview was conducted.

There are two aims for conducting focus group interview as conveyed by Creswell (2012). First is to collect shared understanding and to get views from several individuals. It is advantageous when the interaction among interviewees are similar to and cooperative with each other and is useful when the time to collect information is limited and individuals are hesitant to provide information because some individuals may reluctant to provide information in any type of interview. Second is, to select several students to be individually interviewed. It is relevant with the use of focus group interview as described by Hatch (2002), that they are often used to supplement other qualitative data besides to be the basic data collection strategy of a qualitative study.

3.6 Variables and Hypothesis

The characteristics of this study including the variables and hypothesis are shown in table 3.1

Null Hypothesis (H0)	There is no significance different between reading post-test scores of experiment and control group.
Research Hypothesis (H1)	There is a significant difference between reading post-test scores of experiment and control group.
Significance Level	.05; two -tailed
Dependent Variable	Reading Comprehension Scores
Design	Pretest and post-test control design

Measurement	Scores (interval); Independent measure
Other feature	Random selection
Statistical procedures	paired sample tests and independent t-test

Table 3.1 Diagram showing the characteristics of the study

3.7 Stages in Collecting the Data

There were several steps done in collecting the data. The pre-test and post-test materials, the questionnaire items, the interview items, and the general and Islamic reading materials were first consulted with the experts, then prior to this research, the pre-test and post-test materials were tried out to subjects from similar population to make sure the validity and reliability of the instruments.

In deciding the treatment and control group, four groups were selected and given the same reading test which consisted of 30 items and of only general narrative texts. After that the result were analyzed to see the homogeneity of the groups, and whether the data were normally distributed. Finally, two groups which are homogeneous and have their data distributed normally were chosen

After that a series of treatment was given to the experiment group, where the Islamic reading texts were given for four meetings. Meanwhile, the control groups were given the usual, general narrative texts. The procedure, strategy, steps of instructions, and students' activities in both groups were similar. The amounts of the meetings were in accordance to the term program's time allocation that had been slotted by the school. The text type used was narrative text as it was the material which supposed to be given in the first term of grade IX. Students were given a narrative text with different focus of discussion on its every meeting. The syllabus is on the following time table.

Table 3.1 Time table

Meetings	Activity
1	Introduction (pre-test)
2	Pre-Test: 4 classes
3	Basic purpose and types of narrative texts

	Structure of narrative texts
4	Features of narratives
5	Specific language features of narratives
6	Specific language features of narratives continued
7	Control and Experiment group's Post-test
8	Control and Experiment group's Interview
9	Experiment group's Questionnaire

After the treatments have been finished, both groups were given the same reading post-test which is a different one from the pretest. However, the level of difficulty, discriminating power, reliability and validity are more or less similar.

As the treatment was completed, an interview was conducted to both control and experiment groups to gain supporting data to answer research question 1. The item questioned were related to students' preference on certain type of texts and whether those types of the texts increase their motivation in reading English texts and improve their reading comprehension. The interview was conducted using Bahasa Indonesia in order to get better understanding for the students in comprehending the questions, and in order to have them express their answer better without any limitation.

Prior to that, a set of questionnaires was also given to the experiment group to find out their attitude toward the Islamic reading texts. The questionnaire was intended to see which of their preference. Whether it is general text or Islamic text they prefer better, and whether those specific types of texts improve their comprehension ability and their interest to read more English texts.

3.8 Data analysis

Having collected the data, several steps were conducted to analyze the data obtained. These steps conducted for the data analysis were divided into two parts which were analyzing the students' tests and analyzing students' questionnaire.

3.8.1 Tests

The data were collected through pretests and posttests. The paired sample t-test and independent t-test were used to answer the first research question. The second research question was answered by analyzing the questionnaire given to the experiment group. The post-test results of the experiment group was also compared with the questionnaire result to answer the third research question. The t-test computation was calculated using SPSS 17 for windows. Several steps were conducted for the t-test computations. They were stating the null hypothesis, setting the alpha level at 0.05, computing the t-value which was done by SPSS 17, and comparing the result of the tests.

Prior to the above steps, there were two other tests which are needed to be considered preceding the t-test computation. These tests were normality test and homogeneity test. These tests are elaborated in the following paragraphs.

The normality was used to measure whether or not the data from students score were normally distributed. The hypothesis for the normality test are as follow:

H0: the data are not distributed normally

H1: the data are distributed normally

The criteria for the hypothesis is H1 is accepted if the $\text{sign} > \alpha$, with the level significance 0.05.

The result of normality testing for the pre-test of control and experiment group can be seen in Appendix 14 A. Based on the table, it can be seen that the significant values of the pre-test control group were .200. The significant value of control and group was higher than .05, therefore it can be said that the pre-test scores of both groups were normally distributed. Hence, independent t-test using parametric test can be used (Hatch and Farhady, 1982). Meanwhile, the post-test scores from both groups were also analyze whether or not they are normally distributed. Based on the table in Appendix 14 A, the significant value of those

groups was higher than .05, it can be concluded that the distribution of post-test of control and experiment group were normally distributed.

To find out the homogeneity of both groups, the data were tested by one-way anova test. The criteria for homogeneity test are:

H0: there is a significant difference in the level of ability (not equal)

H1: there is no significant difference in the level of ability (equal)

The criterion for the hypothesis is: H1 is accepted if the result of homogeneity test is higher than .05 ($Sign > \alpha$)

The result of homogeneity for pre-test of control and experiment groups can be seen in Appendix 14 B. From the table, it can be seen that the significant value was .386 which was higher than .05. It indicates that the null hypothesis was rejected and therefore the pre-test score of control and experiment groups were equal.

3.8.2 Questionnaire

Meanwhile, questionnaire was used to find out the answer for the second research question. The data on the questionnaire sheets which were in the form of checklist were coded and scored. The data were then analyzed, interpreted, and explained in relation to the theories discussed in chapter I

The answers of the questionnaire were then categorized and related to the research question, i.e., the students' choice on the type of reading text they choose, whether or not the type of the text improve their reading comprehension ability, and whether or not the use of the text increase their interest in reading more English texts and in learning English in general. The findings from the questionnaire were then also discussed related with the students' post-test scores to answer the second and third research questions on their attitude toward the use of Islamic text in teaching reading and the effect caused by the use of the Islamic text on students' comprehension.

3.8.3 Interview

The data from the interview session using open ended questionnaire were first coded and categorized. The category was based on the mentioned description of the interview guidelines which covers students' identity, students' preference on certain type of texts and whether those types of the texts increase their motivation in reading English texts and improve their reading comprehension. As to emphasize this Creswell (2012) and Mackey and Gass (2005) explained that transcribing and coding techniques conducted to analyze the interview data were intended to provide information and other findings related to the focus of the research.