

### CHAPTER III

#### RESEARCH METHOD

This chapter shows the procedure of conducting the research. It covers research design, population and sample of the research, setting of the research that contain time and place of the study, data and source of data, technique data collection and research instrument, also data analysis technique.

##### A. Research Design

The research design is the overall plan or structure used to answer the research question.<sup>1</sup> The design is used as a guide for a researcher to collect and to analyze the data. In conducting this research, it was used Quantitative design the numerical data needed to measure the English teacher's intra-rater reliability and analyzed it mathematically.

“Quantitative research is explaining phenomena by collecting numerical data that are analyzed using mathematically based methods (in particular statistics).”<sup>2</sup>

Statistics is a special mathematically method used to analyze data in this design. Statistical procedures were used in this study is Descriptive Statistics that

---

<sup>1</sup> Phyllis Tharenou, et.al., Ross Donohue, Brian Cooper. *Management Research Methods* (New York: Cambridge University Press, 2007), 16.

<sup>2</sup> Daniel Mujis. *Doing Quantitative Research in Education with SPSS* (London: Sage Publication, 2004), 1.

helps the researcher to organize, summarize and describe observations.<sup>3</sup> It means that after collecting the data of English teachers' pre- and post- essay test scoring in numerical, the researcher analyzed statistically then interpreted and concluded it descriptively. In addition, Inferential Statistic was also used in this study to generalize findings to the entire population from which the sample was drawn.<sup>4</sup> By this kind of statistic, this research wanted to show that whether the result of descriptive statistic is the real result or only incidental.

## **B. Subject of the Research**

### **1. Population**

A population is defined as all members of any well-defined class of people, events or objects which the generalization is made.<sup>5</sup> The population of this research was all English teachers in Al-Amin Islamic Boarding School Mojokerto. There are eight teachers who have the same quality of being English teachers in Al-Amin Islamic Boarding School. They have a college degree or national certificate in language.

---

<sup>3</sup> Donald Ary, et.al., Lucy Cheser Jacobs, Christin K. Sorensen. *Introduction to Research in Education Eight Edition* (USA: Wadsworth Cengage Learning, 2010), 101.

<sup>4</sup> Ibid., 148.

<sup>5</sup> Ibid., 148.

## 2. Sample

The sample is the smaller group or subset of the total population that the knowledge gained is representative of the population under study.<sup>6</sup> This study was used purposive sampling as the sampling method. Purposive sampling, or judgment sampling, is one of sampling method which the sample elements judged to be typical, or representative of the population.<sup>7</sup> The sample of the study was only six English teachers as the other two teachers have less experience in scoring essay test. So that this study only asked the six teachers to be the sample of this research as they have the same comprehension in scoring essay test for Al-Amin Islamic Boarding School. As the newest study about intra-rater reliability consistency, this study wanted to make the condition of collecting the data as natural as the English teachers do in usual scoring essay test.

## C. Setting of the Research

### 1. Place

The place of the study means the location where the researcher will do the research activity. This study took place on the grade eleven of Islamic Boarding Senior High School of Al-Amin Mojokerto which is placed on RA. Basuni Street no. 18, Sooko, Mojokerto, East Java 61361.

<sup>6</sup> Louis Cohen, et.al., Lawrence Manion, Keith Morrison. *Research Methods in Education Sixth Edition* (London: Routledge, 2007), 100.

<sup>7</sup> Donald Ary, et.al., Lucy Cheser Jacobs, Christin K. Sorensen. *Introduction to Research...* 156.

## 2. Time

The time of the research means when the researcher will start to do the study until finish especially for collecting data. This research started from February, 20<sup>th</sup> 2016 up to April, 23<sup>rd</sup> 2016. All teachers took part in two scheduled data collection sessions:

- 1) One in February, 22<sup>nd</sup> 2016 (pre-scoring)
- 2) One in April, 18<sup>th</sup> 2016 (post-scoring)

### D. Data and Source of Data

The data needed to answer the research question was the students' essay test score in pre- and post-scoring. This research only focused on essay test which was tested in grade eleven of Al-Amin Islamic Boarding Senior High School Mojokerto as the essay material is focused on this level. The question guide of essay test was made in discussion by the English teacher of grade eleven and the researcher. So that it has adjusted with their habit in organizing essay test and made it as usual as they do their daily test. See Appendix 1.

Whereas the source of data was the English teachers' mark of their students' essay test. There are for about six English language teachers; three males and three females, who teach six classes; three classes for girl class and the others for boy class. This study asked all of them to score the essay test. Even there were more than two graders who will score the essay test, the researcher examined only the intra-rater reliability of each rater, not inter-rater reliability.

## E. Technique of Data Collection and Research Instrument

Collecting data means identifying and selecting individual for a study, obtaining their permission to study them, and gathering information by asking people several questions or observing their behaviors which is formed as a collection of numbers (test scores, frequency of behaviors) or words (responses, opinions, quotes).<sup>8</sup> This study uses documentation technique that usually involves quantitative data in the form of archival records.<sup>9</sup> It was in the form of English teachers' pre- and post-essay test scoring. Actually the teachers had used some specific scores needed to be achieved by students but it was not detail. Therefore the teachers were given a rubric to help them to be more specific in their assessment. The rubric was adapted from journal by Viphavee Vongpumivitch entitled *Classroom Writing Teacher's Intra- and Inter-rater Reliability: Does It Matter*<sup>10</sup> from National Tsing Hua University as the analytical rating grid that was adjusted with the teachers' scoring scale before. See Appendix 2.

These data collection sessions are described below.

- a. PRE: From 40 essays in grade eleven, Raters were asked to score the same 10 essays for three male teachers and another same 10 essays for three female teachers. It means 20 of 40 essays were chosen randomly.

---

<sup>8</sup> John W. Creswell, *Educational Research*....10.

<sup>9</sup> Phyllis Tharenou, et.al., Ross Donohue, Brian Cooper. *Management Research Methods* ..... 124.

<sup>10</sup> Viphavee Vongpumivitch. *Classroom Writing Teacher's Intra- and Inter-rater Reliability: Does It Matter*. Journal of International Conference on English Instruction and Assessment National Tsing Hua University, 2006.

- b. POST: Two months later, all teachers as raters were given the other 10 essays again. The 10 essays consisted of 5 same essays (in pre-scoring) and 5 different essays. Like in pre-scoring, they were asked to score those essays again. Some teachers did not remember having read and scored the same essay before and others remembered having read it but they forgot the specific score given.

Then a research instrument is a tool for measuring, observing or documenting the data of the research.<sup>11</sup> The instrument was students' answers of essay test. For the pre-score, the English teachers graded the students' essays a day after the test. Whereas for the post-score, the English teachers gave the score on the copied of the same essays; like in the pre-score, in the next two-month.

#### **F. Data Analysis Technique**

According to Creswell, analyzing data involves drawing conclusions about it; representing it in tables, figures and pictures to summarize it, and explaining the conclusions in words to provide answers to the research question.<sup>12</sup> As quantitative approach used, the researcher would analyze the data using statistic descriptive. The students of XI classes would be conducted the essay test. After doing the test in handwriting, the English teachers gave the essay to the researcher to be copied. Then, the researcher gave the first copy to the English

---

<sup>11</sup> John W. Creswell, *Educational Research* .... 14.

<sup>12</sup> Ibid., Page 10.

teachers to be given the first score (pre-score). After scoring, the teachers gave the first essay copy to the researcher to make the recapitulation.

The second score (post-score) would be given in the next two months. Like the first score, the researcher gave the second copied to the teachers and they graded for the post score. Then the researcher made the second recapitulation to be analyzed and compared with the post-score.

After this step, the data was analyzed by the easiest rater reliability calculation pattern called Cronbach Alpha Coefficient ( $\alpha$ ) as an alternative procedure for calculating the split-half method reliability. It was one of analysis in descriptive statistics to summarize the intra-rater reliability consistency of English teachers at Al Amin Islamic Boarding School Mojokerto in scoring essay test.

The data was analyzed by using reliability analysis in SPSS 23 for each rater in each category of assessment. The steps were:

1. Opening SPSS 23 software;
2. Changing the name of both variables to be “PRE” and “POST”;
3. Changing the number in “Decimal” column to be 1 and choose “Scale” in “Measure” column;
4. Putting the data in the “Variable View”;
5. Clicking “Analyze”, choose “Scale” then “Reliability”;
6. In the reliability box, moving all of rater variables to the right box for analysis;

7. Choosing “Alpha” as this research used Cronbach alpha coefficient for the reliability analysis;
8. Clicking “Statistics” and check “Intraclass Correlation Coefficient”;
9. Choosing “Two-Way Mixed” in category as this study had a population of raters;
10. Choosing “Average Measures” in values as this research needed the mean for the calculation;
11. Choosing “Consistency” in reliability as this study wanted to use the subsequent values for other analyses;
12. Clicking “Continue” and “OK”.

The result of this analysis will be in the form of decimal (ex: 0.333). This measurement had to be over 0.7 before it can be concluded that the test was internally consistent.<sup>13</sup> In addition, there were some levels of reliability that can be interpreted from the coefficient. These levels were used to identify the consistency of each rater in each category. The interpretation is described below.<sup>14</sup>

### 3.1 Table of Reliability Interpretation

<sup>13</sup> Daniel Muji. *Doing Quantitative Research*.... 73.

<sup>14</sup> Abd. Rozak - Wiwin Sri Hidayati, *Pengolahan Data dengan SPSS*. (Jombang: STKIP PGRI Jombang, 2013), 143.





as this research was held only once. As the variables of this study were equal or the same subject, it was appropriate to use t test, especially for dependent sample; called paired t-test; to analyze it.

The paired t test was calculated by using SPSS 23. The data was imported to each cell in “Variable View” and changed the number in “Decimal” to be 1 and chose Scale in Measure like in analyzing the reliability before, PRE and POST. It was calculated by clicking the “Analyze” menu, choosing “Compare Means” then selecting “Paired-Samples T Test”. This application would show the new box and both variables were selected and moved to “Paired Variables” box then click “OK”. Afterwards it would present the output of the data.

The steps in analyzing the inferential statistic use paired t test were:

1. Deciding the level of significance. The most commonly used level of significance in the behavioral science is the 0.05 and the 0.01 levels.<sup>16</sup> If the result was not significant in both levels, it would be tried in other levels such as 0.10; 0.20 or 0.50.
2. Calculating the paired  $t$  test in SPSS 23.
3. Checking the t-test with t-table and p-value “Sig. (2-tailed)” with the level of significance whether the null hypothesis was rejected or not.

### **G. Reliability and Validity**

This research used students’ works formed essays as the instrument that was examined in pre- and post-scoring. The instrument of the research must

---

<sup>16</sup> Ibid., 166.

complete the rules of reliability and validity. The reliability of measuring instrument is the degree of consistency with which it measures whatever it is measuring.<sup>17</sup> It means that the instrument has to produce the same result in any kinds of measurement. As the instrument formed in essays, the instrument's reliability could not be measured mathematically. Nonetheless, this study tried to make the instrument reliable by keeping the data authenticity and avoiding the possibility of subjective value. There were some ways committed to keep the reliability of the data, such as:

- No changing any data (students' work) which was scored by the English teachers in pre- and post-scoring (the same essays).
- Copying the essay answers by photocopy machine twice, one is for pre-scoring and another is for post-scoring which will be given in the next two-week after the first scoring, so that the examiner scored the students' original handwriting.
- Omitting the students' name in each answer sheets of essay test and change it by their numbers of attendance list so that the rater did not understand whose writing it is.

Not only reliability but also validity of the research instrument is very important. Validity means the extent to which an instrument measured what it claimed to measure.<sup>18</sup> As the daily essay examination, the question guide of essay

---

<sup>17</sup> Ibid., 236.

<sup>18</sup> Ibid., 225.

