

### **CHAPTER III**

#### **RESEARCH METHOD**

This chapter presents and discusses research method which consists of research design, population and sample, research instrument, validity and reliability testing, normality and homogeneity testing, data collecting method, treatment, data analysis, and hypothesis testing.

##### **A. Research Design**

The researcher used an experimental research design using quantitative approach, Quasi-experimental research design with nonrandomized control group, pretest-posttest design. Quasi-experimental research designs are similar to randomized experimental designs in that they involve manipulation of an independent variable but differ in that subjects are not randomly assigned (Ary, 2010: 316). In quasi-experimental there are two types of groups, they are experimental and control group. Those groups are given pretest and posttest, what makes two groups different is that the experimental group is given treatment after pretest.

**Table 3.1 Design of The Study**

Group	Pretest	Treatment	Posttest
Experimental	✓	✓	✓
Control	✓	-	✓

The procedure of Quasi-experimental research design was, first, selected two samples (two classes) that at least appeared to be similar (ability or value). Then the experimental group and control group were selected from two classes. After that, all of the groups were given pretest but before doing the test, all the students in a class were explained about what they would write. The next was, giving treatment to the experimental group by conducting a recount text learning by using peer review for assessing in the steps of making product of writing. The treatment classes was conducted by giving the material of recount text first, after that they were asked to make the example of the text and then all the students practiced peer review activity. First, they assessed the outline of their friends' text, and the second, they conducted peer review for their friends' final text by using the peer review guidelines (see treatment) and gave score based on themselves. Besides that, they were given time to give another feedback to their friends using their own perception and language. The researcher directed and monitored all the activities that should be done. While the control class was not given peer review in learning recount text. The last was giving posttest and comparing the result of two groups.

## **B. Population and Sample**

Ary et al. (2010) states that the small group that is observed is called a sample, and the larger group about which the generalization is made is called a population. A population is also defined as all members of any well-defined class of people, events, or objects. The population of the research was students of first grade of MAN 1 Tulungagung.

The sampling technique that was used by researcher was purposive sampling where the sample of the research was chosen because it fits with the criteria that were needed. In this research the samples were the A.19 MIPA 4 and A.19 MIPA 5. Both of those class has the same material, it is recount text and they are in the same level of study.

Beside that, in verifying that two classes were equal, the researcher calculated pre-test score of both classes. The researcher used Mann-Whitney U Test in SPSS 16.0 version. The researcher used Mann-Whitney U Test because one of the data of the classes was not normally distributed based on the result of normality testing (see normality testing result). The result showed that the significant level/ *p value* was 0.834. The *p value*: 0,834 is higher than the  $\alpha = 0.05$ , so it indicated that there is no difference in variance between both of data or in the other words those classes were equal. The result is presented as in table 3.2 below:

**Table 3.2 Mann-Whitney U test for Pre-Test**

Test Statistics <sup>a</sup>	
	Score Pre-test
Mann-Whitney U	300.000
Wilcoxon W	576.000
Z	-.210
Asymp. Sig. (2-tailed)	.834

a. Grouping Variable: Group

### C. Research Instrument

In a research, instrument is needed to collect the data. Wilkinson & Birmingham (2003:3) defines research instruments as devices to obtain relevant

information for the research project. In this study the researcher used test as the instrument to gather the data. The types of the test were pretest and posttest.

The pretest was given as the first step of collecting the data. Both of the groups, experimental and control were given pretest. The pretest was used to identify the students' preliminary knowledge of recount text and achievement in writing the text. Beside that, pre-test was also used to know that both of classes are equal. The test itself asked them to make a recount text about historical event. The researcher gives 60 minutes (25 minutes for getting any information and making an outline + 35 minutes for making the whole text) to make a recount text based on the topic that was selected by them.

The posttest was given after the control group finished the Recount Text material and the experimental group had been given the treatment by conducting peer review. The peer review itself not only by giving a comment in spoken and written but also having a guideline (see treatment) so that in all steps of making a text had peer review activity twice.

In giving the score, both in pretest and posttest, the researcher used scoring rubric the analytical scoring rubric, the scoring rubric is presented as follows ;

**Table 3.3 Scoring Rubric Adapted From Cohen (1994:328-329)**

Aspect	Criteria	Score
Content	- Main ideas stated clearly and accurately, change of opinion very clear	5
	- Main ideas stated fairly clearly and accurately, change of opinion relatively clear	4
	- Main ideas somewhat unclear and inaccurate, change of opinion somewhat weak	3
	- Main ideas not clear or accurate, change of opinion weak	2
	- Main ideas not all clear or accurate, change of opinion very weak	1

Organization	- The recount text is well organized and perfectly coherent.	5
	- The recount text is almost organized well and coherent, but there are some incomplete sentences.	4
	- The recount text is disorganized but main ideas still clear, logical but there is incomplete part of recount text.	3
	- The recount text ideas is disconnected, the main idea is not too clear on each paragraph but still can be understood, lacks logical sequencing, and incomplete part of recount text.	2
	- The recount text is unorganized and incoherent.	1
Vocabulary	- The choice of words idioms, preposition and word forms are right.	5
	- The choice of words and idioms are right but there are some mistakes in word forms and preposition.	4
	- The choice of words is sufficient but there are some inappropriate of idioms, preposition, and word forms.	3
	- Many mistakes in choosing words, idioms, preposition, and problems in word forms that could change the meaning.	2
	- Very poor knowledge of words, idioms, preposition, and word forms.	1
Grammar	- No errors, full control of complex structure	5
	- Almost no errors, good control of structure	4
	- Some errors, fair control of structure	3
	- Many errors, poor control of structure	2
	- Dominated by errors, no control of structure	1
Mechanics	- Mastery of spelling, capitalization, and punctuation	5
	- Few errors in spelling, capitalization, and punctuation	4
	- Fair number of spelling, capitalization, and punctuation errors	3
	- Frequent errors in spelling, capitalization, and punctuation	2
	- No control over spelling and punctuation	1
<b>Total score</b>	<b>25 x 4</b>	<b>100</b>

## D. Validity And Reliability Testing

### 1. Validity Testing

Validity testing is the extent to which inferences made from assessment results are appropriate, meaningful, and useful in terms of the purpose of the

assessment. The process of gathering evidence to support (or fail to support) a particular interpretation of test scores is referred to as validation (Ary et al., 2010: 226). There are four types of validity, they are content validity, construct validity, face validity, and criterion-related validity. In this study, the instrument is tested by using content validity, construct validity, and face validity, it is those are relevant with this research.

a. Content validity

Content validity is related to ability of instrument in measuring content or concept of what is desire to measure in the way that the instrument is representative (Brown, 2003). Before making each test the research made a blueprint of the test item. The blueprint itself consisted of Basic Competence, material, indicator of the test and type of the test. The basic competences that are used 4.7, 4.7.1 and 4.7.2 about the material of recount text. The indicator is the students is given some topics of Indonesia historical events and they had to make a recount text from the selected topic. The type of the test itself is essay. Based on the explanation above, it is showed that the test is valid based on content validity.

b. Construct Validity

Construct is any theory, hypothesis or model that attempts to explain observed phenomena in our universe perception (Brown, 2003:25). In other word, construct validity can be fulfill if a test can measure what we are supposed to measure. In here the researcher want to assess the writing skill of the students. The researcher has conducted the writing test and its

activity is in accordance with Salma (2016) that writing is an action, a process of discovering and organizing our ideas, putting them on the paper, reshaping and revising them. Moreover, according to Brown (2003:220), the categorization of writing performance belongs to responsive. It is because in responsive, the writing performance is connecting sentences into a paragraph and creating a logically connected sequence of two or three paragraph. Genres of writing include brief narratives, descriptions, short reports, lab reports, and summary. So the pretest and posttest is valid based on construct validity.

#### c. Face Validity

Face validity itself relates to a test's content. Face validity refers to the extent to which examinees believe the instrument is measuring what it is supposed to measure. Face validity is important to ensure acceptance of the test and cooperation on the part of the examinees (Ary et al, 2010:228). Oluwatayo (2012) defines face validity to researchers' subjective assessments of the presentation and relevance of the measuring instrument as to whether the items in the instrument appear to be relevant, reasonable, and unambiguous and clear. He explains more about the criteria that should be gained to make the test valid based on face validity such as; appropriateness of grammar, the clarity and unambiguity of items, the correct spelling of words, the correct structuring the sentences, appropriateness of font size, the structure of the instrument in terms of construction and well- thought out format. This research is done to know

the effectiveness of using peer review as method to improve students' writing achievement in writing recount text, so the test should in the form of writing test. Related to this research, the researcher also asked the students to write a recount text. In doing a face validity, the researcher give the test instrument to the lecturer and the English teacher of A.19 MIPA 4 and A.19 MIPA 5. After checking the instrument, they confirmed that the test is valid based on face validity.

## 2. Reliability Testing

Brown (2003: 20) states that a reliable test is consistent and dependable, so if we give the same test even on different occasions, the test result should be similar. In addition, the test is reliable if the score is steady over time. Reliability is also seen as the degree to which a test is free from measurement errors, since the more measurement errors occur the less reliable the test (Fraenkel & Wallen, 2003; McMillan & Schumacher, 2001, 2006; Moss, 1994; Neuman, 2003). There are some factors that may contribute to the unreliability of a test. Mousavi (804: 2002) states there are some possibilities that may influence such as: fluctuations in the student, in scoring, in test administration, and in the test itself. In anticipating those problems in the reliability of the test, the researcher gives clear instructions on the test sheet and the existence of test specification (time allocation, class, and semester).

Not only that, the topic of recount text that they would be written is also included. The topic itself is telling the historical event, students are given the choices that relate to the topic, for the example in pre-test, they are; *Peristiwa*



*Rengasdengklok, Kongres Sumpah Pemuda I, Berdirinya Budi Utomo, and Peristiwa Bandung Lautan Api.* The test sheet itself is typed so that it is readable. In assessing the students work, the scoring rubric (see research instrument) was also used as the guidance in giving the score and also to decrease the subjectivity of the corrector.

Another way to check the reliability of the test was done by using the procedure of inter-rater reliability. Inter-rater reliability itself can be done by two or more observers produce similar quantitative results when observing the same individual during the same time period (Ary et al, 2010:228). Whether Livingston (2018:15) defines inter-rater reliability as the consistency of the scoring process, on a test for which the scoring involves judgments by raters. It is the agreement between the scores produced by different raters scoring the same responses.

After having the scores the researcher will use SPSS 16.0 version to analyze the reliability of the test. The procedures were, inputting the data, then click 'analyze' and choose 'scale'. After that click the 'reliability analyses'. There will be a table after clicking it, then move the data on the left table to the right table. After that click 'statistics' and choose item, ;scale, and scale if item deleted on "descriptives for" part and in inter-item part choose "correlations". The result is shown after that and the 'cronbach's alpha' that will determine how reliable the test is will be existed. Here are the classifications of the cronbach's alpha:

1. cronbach's alpha > .9 – Excellent,

2. cronbach's alpha  $> .8$  – Good,
3. cronbach's alpha  $> .7$  – Acceptable,
4. cronbach's alpha  $> .6$  – Questionable,
5. cronbach's alpha  $> .5$  – Poor,
6. and cronbach's alpha  $< .5$  – Unacceptable. (George and Mallery, 2003 : 231).

Before giving the test sheet to the sample of the research, the researcher tried it on the students who had similar background (MIPA class) to the sample of the research. The class that was chosen A.19 MIPA 3. After that the researcher analyzed the reliability of the test instrument by using SPSS. Here are the result of reliability testing of pre-test and post-test;

**Table 3.4 Reliability Testing of Pre-Test**

Reliability Statistics		
Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.731	.799	2

**Table 3.5 Reliability Testing of Post-Test**

Reliability Statistics	
Cronbach's Alpha	N of Items
.901	2

The cronbach's alpha of pre-test is 0,731, from the classifications of the cronbach's alpha it belongs to acceptable, that was why the pre-test was reliable.

While the cronbach's alpha of post-test is 0, 901 which belongs to excellent, and confirmed that the post-test was also reliable.

## **E. Normality And Homogeneity Testing**

### **1. Normality testing**

Normality testing is used to examine whether a set of data is belong to normal distribution or not. After doing the normality testing then the researcher can determine whether the statistical test uses parametric test or non-parametric test. In this research the normality testing was done toward students score in pretest and posttest.

The main tests for the assessment of normality are Kolmogorov-Smirnov (K-S) test , Lilliefors corrected K-S test , Shapiro-Wilk test, Anderson-Darling test, Cramer-von Mises test, D'Agostino skewness test, Anscombe-Glynn kurtosis test, D'Agostino-Pearson omnibus test, and the Jarque-Bera test. Considering the sample in this research that only 50, among those normality tests, the researcher used Shapiro-Wilk tests that can be conducted in the SPSS 16.0 version by clicking on Explore procedure (Analyze → Descriptive Statistics → Explore → Plots → Normality plots with tests) (Elliot and Woodward, 2007). Shapiro-Wilk itself is one of normality testing that is used for the sample 10 – 70 (Oktaviani & Notobroto, 2014). Razali and Wah (2011) also stated that Shapiro-Wilk shows the best result of distribution, followed by Lilliefors

and Kolmogorov-Smirnov (K-S) test. In this research the researcher used Shapiro-Wilk result to know the normality of the test.

## 2. Homogeneity testing

This test is used to know the variables are equal in variance or not. The researcher used SPSS 16.0 version to test the homogeneity. In deciding the variable is homogenous or not, it was compared with  $\alpha = 0,05$ , if the result shows that it is higher than  $\alpha$ , then it is homogenous, but if the result shows that it is lower than  $\alpha$ , then it is heterogeneous.

## **F. Data Collecting Method**

Collecting data is the process of gathering and measuring information on variables of interest, in an established systematic fashion that enables one to answer stated research questions, test hypotheses, and evaluate outcomes (Kabir, 2016: 201). In other words, collecting data is a systematic and standardized procedure to obtain the data. In this research the researcher used Administery test. Administery test is a set of stimuli presented to individual in order to elicit responses on the basis of which a numerical score can be assigned (Ary et al, 2010:216). The administery test is classified in two part, they are pre-test and post-test. The explanation is presented as follows.

### 1. Pre-test

This test was done in the first meeting of control and experimental classes. This test is aimed to know the students preliminary knowledge of recount text, their

achievement in writing and also the equality of those classes. The students are given the test with the instruction which asks them to make a recount text.

## 2. Post-test

This test was done in the end of the meeting or after the treatment. Both of the class were given the same test. This test is aimed to know the improvement of the students in writing recount text after giving peer review activity for experimental class and without using peer review activity for control class. Just like in pre-test, the students were asked to make a recount text but having different topic from the pre-test. After it, the researcher got the students' scores and compared them to know how effective peer review activity in improving students writing achievement.

The research was done after having a coordination with the principal of the school (MAN 1 Tulungagung) and the English teacher of the classes that were used. The research was done in four meetings of each class. On A.19 MIPA 4, the research could be done all in offline learning. Unfortunately, on experimental class, for the last meeting had to done in online learning trough WhatsApp group. It was done as the policy of the President in preventing corona virus. All the lesson time was done around 90 minutes, but for the experimental class, the researcher had to divided it into two meetings (around 20 minutes for the treatment and the post-test spent around 60 minutes). The schedule of the meetings on two class are presented in the following tables:

**Table 3.6 The Schedule of The Research On Control Class**

No	Date	activity
1	February 5 <sup>th</sup> , 2020	Giving recount text material and pre-test
2	February 12 <sup>th</sup> , 2020	Giving recount text material and exercise
3	February 19 <sup>th</sup> , 2020	Making a recount text
4	February 26 <sup>th</sup> , 2020	Post-test

**Table 3.7 The Schedule of The Research On Experimental Class**

No	Date	Activity
1	February 12 <sup>th</sup> , 2020	Giving recount text material and pre-test
2	February 19 <sup>th</sup> , 2020	Giving recount text material, exercise and trying <i>peer review</i> activity on group work
3	February 26 <sup>th</sup> , 2020	Making a recount text and giving treatment I (practicing peer review on the outline that have been made). Revising and continuing on drafting.
4	April 13 <sup>th</sup> , 2020	Giving treatment II (online)
5	April 15 <sup>th</sup> , 2020	Post-test

## G. Treatment

The treatment classes was conducted by giving the material of recount text first, after that they were asked to make the example of the text and then all the students were divided into groups consisted of two students. According to O'Muirheartaigh (1990) peer review itself can be employed in three forms; pre-peer review, while-peer review and post-peer review;

### a. Pre-peer review

In the pre training teacher clarifying the benefits of peer review and then grouping the students. Teacher shows how to give feedback/review by doing an example. It can help students to structure their papers more clearly.

### b. While-peer review

In this section students begin their peer review and teacher is a monitor.

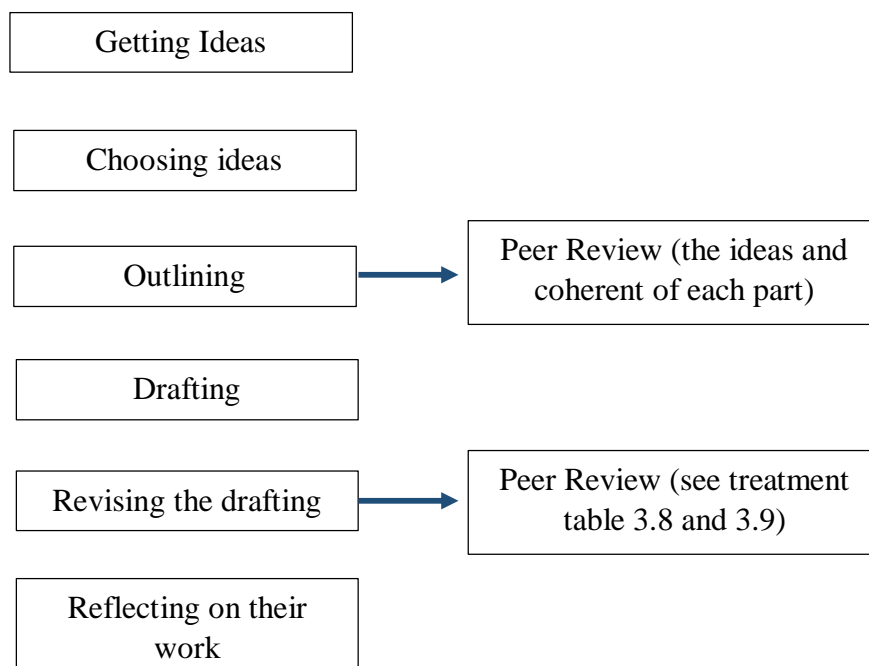
c. Post–peer review

Students reflect on what they did during peer review and they write about their experience.

In practicing peer review, the researcher did some modifications. After students were asked to make a recount text, the researcher added explanation about how to do peer review to assess their friends' work by using guideline and peer review sheet that have been provided. Unfortunately, the researcher had to modify the form of their work to make the samples understood how to do peer review easily in online meeting. Not only in the end of the students writing activity (the final text) but they did peer review twice in writing steps.

The writing steps itself modified from the book “Writing from Start to Finish” by Grenville (2001:11). There were six steps in writing as follows:

**Figure 3.1 The treatment steps in experimental group**



The figure is defined as follows:

1. Getting ideas

This part determines how the whole story would be. There are four ways to generate an idea such as: making a list (brainstorming), making a cluster diagram, researching or independent investigation, freewriting. In here the samples were able to find any information related to the topic that had been chosen. They made a list of sentences that was relevant to the topic. By doing this step they also reduced a chance to get stuck in writing the whole text.

2. Choosing ideas

In this part they looked at all and assess their ideas. This was where the writer/sample started to discriminate between the ideas that definitely can't use, and ones that had some potential. The writer should know what the purpose of the text is. In writing recount text the text is purposed to persuade or inform or both, so the text would be: can this idea be used as part of an argument, or as information about the topic? The answer is yes if the idea would give the reader facts about the subject, a general concept about it, or an opinion about it, or if the idea could be used as supporting material or evidence.

3. Outlining (putting these ideas into the best order—making a plan).

The samples had to put their selected ideas into an order as the researcher presented below:

- a. **Beginning**—an introduction, telling the reader where they are and what kind of thing they're about to read.
- b. **Middle**—the main bit, where you say what you're there to say.



- c. **End**—some kind of winding-up part that lets the reader know that this is actually the end of the piece (rather than that someone lost the last page).

After finished the third steps, all the experiment students did peer review. In this part, each students gave written or spoken comment to their friends' work in pair. It focused on the outline that has been made and the coherence of each part (beginning, middle, end), but it was still possible for the students to give some comments on ideas that might be not relevant. When this peer review activity was applied, students spent longer time than the prediction. It was because they were tired (the teaching was at 11.45 – 13.15, they also just finished the Chemistry class for three hours of lesson time (it is 135 minutes)). The researcher had to create a fun atmosphere so that they could concentrate more.

4. Drafting (doing a first draft from beginning to end, without going back)

In this part the students wrote the full version of their text by developing their ideas in the beginning, middle, and end. They had to develop the sentences (by giving supporting details) that they thought relate to the idea and did not need to think about the correct grammar, verb, comma, etc. They just had to complete the text.

5. Revising the drafting (cutting, correcting, adding or moving parts of this draft where necessary).

After completing to make a full text, they reread their work and revised anything necessary. The samples deleted unimportant word or sentences, add sentences (as long as having the correlation), moving some

sentences that still had wrong position. Even they could revise the wrong spelling and punctuation. Not only by the sample itself, but his/her friend helped them to do this by giving advices.

The second treatment was implemented again in this step. After the texts had been collected at the previous meeting, the procedure was, the researcher asked them one by one to give review consists of comment, suggestions, and score on their friends' texts. The comment and suggestions had to be appropriate with the guideline that had been provided. The guideline just like presented below:

**Table 3.8 Correction Guideline for Experimental Group**

<b>Correction Guideline</b>	
<b>You should check :</b>	Grammar and verb
	Spelling
	Preposition
	Capitalization
	Punctuation and indentation

Beside that, this research was also used another guideline in giving peer review, it was;

**Table 3.9 Peer-Review Sheet Adapted From C. Waller (1994). ESL Middle School Teacher, Fairfax County Public Schools, Virginia.**

<b>Reader's name :</b> <b>Author's name :</b> <b>Title of piece :</b>	<b>Date :</b> <b>Class :</b> <b>Score :</b>
<b>Peer Evaluation</b>	
The text is: _____	
_____	
_____	
_____	
The best part of the text is : _____	
_____	
_____	
The text can be improved by : _____	
_____	
_____	
_____	
_____	

The treatment that used this correction guideline and peer review sheet above were done in online situation because of the President policy about Work From Home (WFH) on preventing corona virus. The researcher directly gave example on how to give comments on their friend's text in a new form, because it was too difficult for students to fill the guideline. So they were explained and done the review on their friends' text using Indonesian in a new form, but still what they should review had to follow the guideline. About the score consideration, it was depend on the students itself, there was no scoring rubric in giving the score on their friends' text.

After they sent back comment, suggestions, and score to the researcher, then the researcher made a WhatsApp group. All the reviews from the experiment class were sent there, and the researcher asked them to pay attention on the review of their own texts. After sending them, the researcher concluding all mistakes that often made by them on their texts.

6. Reflecting (checking for grammar, spelling and paragraphs).

They checked how the reviewer gave comments, suggestions, and also score on their work. Not only the students but the researcher gave comment or conclusion from reviews that have been resent in WhatsApp group on students' texts. This step was also done through *WhatsApp* group.

The researcher directed and monitored all the activities that should be done in every steps.

## **H. Data Analysis**

In this research, the researcher used quantitative data analysis. Quantitative data analysis is also called statistical analysis. This technique was used to find out the significant difference in writing achievement of the first grade students of MIPA 5 and MIPA 4 in writing recount text that used peer review and those who are not used peer review.

There are two inferential statistical procedures, they are parametric and non-parametric tests. *t* Test is a type of parametric method; they can be used when the distribution of data is normally distributed, equal variance, and independence (Kim, 2015). It is used to compare the means of two groups.

In non-parametric test, there is Mann-Whitney U Test which is a type of non-parametric test is most likely to be recommended when the normality assumption of the  $t$  Test for two independent samples is not gained (Sheskin, 2000 : 291). Moreover, Conover (1980, 1999), Daniel (1990), and Marascuilo and McSweeney (1977) stated that the consideration of having Mann Whitney test is the distributions from which the samples are derived are identical in shape. The shapes of the underlying population distributions, however, do not have to be normal. The assumption of identically shaped distributions implies equal dispersion of data within each distribution. Because of this, they note that like the  $t$  Test for two independent samples, the Mann–Whitney U Test also assumes homogeneity of variance with respect to the underlying population distributions.

In this researcher, the researcher used the Mann-Whitney U Test to know whether the hypothesis is rejected or not. In this case the researcher calculates it by using SPSS 16.0 version