

## **CHAPTER III**

### **RESEARCH METHODOLOGY**

This third chapter discusses about (1) Research Design, (2) Place and Time of The Research, (3) Population and Sample, (4) Subject and Object of The research, (5) Research Variable, (6) Research Procedure, (7) Technique and instrument of Collecting Data, (9) Technique of Data Analysis.

#### **3.1 Research Design**

In this research, the researcher used the mixed method to get the data. This method had the designs were included both qualitative and quantitative structures in the design, data collection and analysis (Mertens, 2010:293). Mixed method was used of the both open-ended and closed-ended questions and statistical and text analysis to analyze the data by the researcher from collected the data (Creswell, 2014:45). A mixed methodology would be used for the experimental data collection, using numerical and verbal data, in order to collect rounded, reliable data. And the type of mixed method that used was the exploratory sequential design. It of the design with the first analyzes of the qualitative aspect and followed the quantitative aspect to get the result of the study (Creswell, 2012:543). The quantitative data was build or explain the result of qualitative analyze.

The qualitative data of this research is for analyze of the items test based on the syllabus and a good assessments of multiple-choices items with the instruments analysis of curriculum and criteria of a good assessments check-list based on Brown & Heatons' book . Moreover, the quantitative data of instruments' test and statistic to analyze the level of difficulty, discrimination power, validity and reliability of the distracters from students' score and students' paper sheet.

### **3.2 Place and Time of the Research**

This research was conducted in the south district Gresik consists of three sub-districts. And were selected three schools in the south district Gresik at an area KKM. Those are MTs Sunnan Ampel in Kedamean, MTs Tarbiatul Aulad in Cerme and MTs Irsyadul Athfal in Benjeng. The implementation of this research was on April, 11<sup>th</sup> and 29<sup>th</sup> 2017.

### **3.3 Population and Sample**

Population is a generalization region consist of subject and object that they has specific quantity and characteristics which was determined by writer to studied and got a conclusion for the studied. Those populations are any items, individuals, not only people or just number but also the objects/subjects and anything that has specification about identifying feature of objects or objects (Sugiyono, 2015:80). The population of this study are seventh grade students in South district Gresik at academic year 2016/2017. In this population there are 5 sub-districts; Driyorejo, Wringinanom, Kedamean, Cerme, and Benjeng. Then, from 5 sub-districts there has 43 – 50 schools (Mts). This final test is get from KKM (Kelompok Kerja Madrasah) in MTsN Metatu in Gresik.

A sample is a part of total and characteristics from the population. The sample is defined as any group of individuals that is selected to represent a population (Sugiyono, 2015:81). The samples of them were English final test of the first semester at seventh grade students in one school of three sub-districts Gresik. There areas of the schools are MTs Sunan Ampel in Kedamean, MTs Tarbiatul Aulad in Cerme, and MTs Irsyadul Adfal in Benjeng. The sample of the study determined from sampling technique. The sampling technique is technique for taken sampling from the research. And in this research, the writer had chosen the sampling technique of sample random sampling technique, because not all the schools want to be the sample in the research. The sample

random sampling technique was called simple taken technique because the technique did not look at the strata in population region of the taken the sample (Sugiyono, 2015: 82). It is simple randomizing, that a simple was opportunity of self-regulating and an equal of being included in the sample (Pandey, 2015:47).

### **3.4 Object of the Study**

The object of this study focused on the test instrument and the result. The researcher taken from the schools' document in 2016/2017 of academic year in South District Gresik and those documents were collected in order the data, then these data were analyzed to give a breakthrough how the quality of the test.

### **3.5 Research Instrument**

In this research, instrument as tools for collecting data to analyzed and draw the conclusion of the result in this study. Thus, the conception of research instrument carefully and in an appropriate manner is greatly fundamental. With regard to different desire data, research instrument may different from one study and another. Research instrument in the forms of English final test was measure the quality of the items test that applied in South District Gresik. For the details function all of the tools was work as follows:

#### **3.5.1 Curriculum Checklist**

This instrument was used to know the appropriateness and test-items within curriculum and the context of the items test-pack of in the syllabus to analyzing how much the material can measure the students' ability and the levels of the content (Cunningsworth, 2008:60). It the ways to know how far the test has fulfilled the instruction materials that almost related to analyzed from selection of content and grading of the paper test question, which made by

MGMP in South District Gresik and used on 43 – 50 schools in Gresik. The curriculum-checklist related of syllabus on national curriculum now.

In the curriculum checklist, the researcher was analyzing of each numbers in the paper test questions. Which the number of the items tests that includes in the basic competence point 3.1 until 4.4. And how many numbers in each basic competence, then how many percentages at the each basic competence. It was to answer the first and second question problem.

### **3.5.2 Characteristics and Criteria of Good Assessment Checklist**

The checklist used for identify some characteristics and criteria of English Final test paper based from Heaton and Brown's book. The researchers look at test-items and analyze them whether they had fulfilled the good test or not. From this instrument and technique, the writer got the analyzed of the second of the question problem and knew of the qualitative data based on a good assessment checklist of multiple-choice items of the test paper that applied on South District Gresik schools.

### **3.5.3 Ms. Excel of Analysis's Application**

The Ms. Excel of analysis's application was the application for calculate the data that included in quantitative aspect. A computation to calculate the data to items difficulty levels and the discrimination power. Then the researcher was manage the result from the data analyze accordance with the aspects which fulfill the requirements of the organize items test.

#### **3.5.4 SPSS Version 16.0**

SPSS was statistic software to analyze the research data. SPSS was statistical package for the social sciences, it the statistical product and service solutions. The writer used the SPSS version 16.0 to managed and analyze the data that got from students' paper sheet of the random sampling in South District Gresik.

The SPSS program to find the validity and reliability data in this research. From the fifty questions of ninety students that writer gets from the three schools of the random sampling data. The valid and reliable data to know the items test would be able and be agreeable to measurement the quality in the English items test.

#### **3.6 Technique of Collecting Data**

The data resources for this research were the schools' documents and students' scores in the South District Gresik as random sampling data. The schools documents were test and blueprint made by the teachers to develop the test. The test consists of fifty items in the form of multiple-choice. The form of multiple choices included the matter in syllabus, and for the example were greeting, family, introduction, time, etc.

The students' score were acquired form the re-administrations of the test in three different schools because to draw the data of students' score in an area MGMP of South district Gresik at seventh grade students. Meanwhile, the re-administrations data was aimed to see whether the test were reliable, valid, check the items difficulty, discrimination levels, the appropriateness the test to students in the 43-50 schools, and also to see what students can achieved the goals of the test or not in South district Gresik in academic years 2016/2017. Then, the details explanation of each instrument to get the data as follows:

### **3.6.1 Observation**

The observation was the first technique for gets the samples and the data as through survey of this research. According to Khotari (2004:17) that technique collecting data by observation is not suitable in inquiries where large samples are concerned. The observation is providing to get the purpose for this research (Khotari, 2004:96). The researcher exerted observation to getting the schools as the sample and getting the data, like the paper test question, students answer sheets, students' score in South District Gresik in this research. From the observation researcher can find out which schools can be referable as the sample. And the desired data can be obtained in accordance with the aspects under the study.

### **3.6.2 Paper Test Question**

The paper question test was comprised of fifty questions' multiple-choice. The test pack was taking from English Final Test used by teachers in MGMP South District Gresik that used in first semester in academic year 2016/2017. The test pack was applied in seventh grade students and the sampling is school in three sub-districts Gresik.

The test-pack was matched with curriculum national or instructional material to saw the quality of the test items in case of the qualitative data. After the writer analyzed the quality of the test items in paper test question, the writer would like use the paper test question to analyze with the curriculum and good assessments checklist for answer the second question problem.

### **3.6.3 Students Answer Sheet**

The students answers' sheet is the data that writer found from the schools that applied from the three schools from the random sampling. Used by the writer to analyze the data from the quantitative aspect to know the difficulty levels and discrimination power from this research and to answer the third of the problems.

### **3.6.4 Students' Score**

The students' score was the result from the English final test in the first semester. And it was use to know students' answer distribution. It was to find out the validity and reliability data that writer would like to analyze the quality of the data, and then it was to answer the second of the problems.

### **3.6.5 Data and Source of Data**

The researcher get the data from the items of English final test, the key answers, the students' answer sheets and syllabus of English for seven grade of Junior High School in South District Gresik academic year 2016/2017.

The data of this research were components of items test, the test items and students' score based on the students' answer sheets. In addition, the resources of the data were the questions test and the students' paper sheets.

## **3.7 Research Procedure**

The procedure is the steps of implemented to gets the goals of this analyze. And it describes the steps taken by the researcher in conducting this research. The steps as follows this table below:

**Table 1 - Table 3.7 the Research Procedure**

No.	Time	Activities	Information
1.	January, 6 <sup>th</sup> 2017	Problem and potency analysis	The writer begun has an idea of how teachers made a test to measure their students. Then, to know what and how a good test is?
2.	February, 4 <sup>th</sup> 2017	Collecting theory	The researcher used some theories and references to adopt and adapt the important things to create this research.
3.	March 1 <sup>st</sup> – April 29 <sup>th</sup> , 2017	Collecting data	To collect the data, the researcher used some form and documents to analyzing this study. Those are: copying the English final test from of seventh grade students, copying the students' score, collecting the items test and analyzing the result of the items test to know the validity, reliability, items difficulty, and discrimination level of the test.
4.	May 1 <sup>st</sup> – June 30 <sup>th</sup> , 2017	Analyzing the data	The data analyzed in this research was taken from the students' score of seventh grade students in South district Gresik in the academic year 2016/2017. The students' score used to analyze the quality of items difficulty, reliability, discrimination level, item validity that according to the characteristic and criteria of good test in EFL / ESL in the quantitative analyzed.
5.	July 10 <sup>th</sup> , 2017 until August, 2017	Take the Result and the Conclusion	After was analyze all of the data with the aspect qualitative and quantitative, the researcher was take the result what she get and find out the data of this items test in English final test at South District Gresik in 2016/2017 academic years. Then, make the conclusion and the suggestions of the items test. And the last said that it the



			items test was a great instrument of the assessment or not or it was enough instrument to measure the students' abilities as students of ESL.
--	--	--	---

### **3.8 Technique of Data Analysis**

#### **3.8.1 Qualitative Data Analysis**

It was distributed through analysis and studied on no-statistical structures on the test-items. It had two aspects on this sub-chapter that the researcher was going to analyzed. Those are analysis of appropriateness to curriculum and analysis of characteristics of a good test item according based on syllabus and Heaton and Brown. Then, the analysis of instructional tools, the characteristics and criteria of good test counted in analysis of language use in it.

##### ***3.8.1.1 Analysis of the Appropriateness to Curriculum***

The test-items were reviewed whether or not they had matched with Standard and Basic Competent especially in junior high Schools as MTs in an area of South District Gresik at seventh grade students. An analysis of instructional materials dealt through appropriateness of the content on the paper test question that they usage with instructional materials of teaching and

learning process based in syllabus in 2013 curriculum based on revision. The data would be analyzed based on the curriculum checklist. Then, to make it easily and clearly the writer shown the illustration based on sub-chapter 3.5.1 for this step as follow in the instruction based on table below:

**Table 2 – Table 3.8.1.1 The Curriculum Checklist**

<b>Core Competence</b>	<b>Basic Competence</b>	<b>Themes / Materials</b>	<b>Items Test that Appropriate with the Basic Competence</b>	<b>Total number of items test (<math>\Sigma</math>)</b>	<b>Percentage of Total Numbers of Particular Items Represent the Elated Basic Competence</b>
3. Memahami pengetahuan (faktual, konseptual, dan prosedural) berdasarkan rasa ingin tahunya tentang ilmu pengetahuan, teknologi, seni, budaya terkait fenomena dan kejadian tampak mata.	3.1 Mengidentifikasi fungsi sosial, struktur teks, dan unsur kebahasaan teks interaksi interpersonal lisan dan tulis yang melibatkan tindakan menyapa, berpamitan, mengucapkan terimakasih, dan meminta maaf, serta menanggapi, sesuai dengan konteks penggunaannya	Interaksi antara peserta didik di dalam di luar kelas yang melibatkan tindakan menyapa, berpamitan, berterimakasih, meminta maaf yang dapat menumbuhkan perilaku yang termuat di KI			

*(Syllabus on 2013 curriculum based on revision)*

In the table above the writer used the six columns that it applied from syllabus. The first column was description about core competence, second column was about basic competence, third column filled of themes / materials, the next column of items test that appropriateness with basic competence, and the fifth column was filled of total number of items that included with the themes and basic competence, and the last column was contain of the percentage from each basic competence that included from the test items. It was seen on the appendix of the appropriateness regarding to the curriculum and syllabus, that was explain about the analyzed the curriculum check-list by the writer usage from get the valid data to known the answered of question problem number one of the one about the kind of the test that included in the first semester of the basic competence in syllabus from the first theme until the last theme.

### ***3.8.1.2 Analysis of the Characteristics of a Good Test***

One way to known what is a good questions test or not to assigned of mastered the matter that teacher gift for their students and to know how far the test could applied in schools as students EFL/ESL, the researcher used a checklist which there was a measurement of a good test according to Brown & Heaton's theories used in the students' assessment. It was a good assessment checklist had fifteen points that should be in form of multiple-choices to questions test. And it was shown from the illustration according to sub-chapter 3.5.2 in this table below as follows:

**Table 3 - Table 3.8.1.2 Good Assessment Checklist**

<b>No.</b>	<b>Characteristics' of Good Multiple Choice</b>	<b>Checklist (√)</b>
1.	Presented out of simple context	
2.	The Stem; the choices from which the students select their answers are referred to as option/responses/alternative	
3.	The option in multiple choice are: a. One option is correct option or key b. Other option is distractors	
4.	Etc.	

*(Heaton. 1990:25 & Brown. 2004:55)*

In the table there was three columns, those were number, characteristics'' of a good multiple-choices and space for checklist ( $\sqrt{\phantom{x}}$ ), which contains some requirements to thought as a good test or not. Then, the researcher would like to fifteen analyzed data usage this checklist. Then, made a description based on theories and seen the appendix of the test appropriateness toward students levels to know the application and find to this check-list. If there was an item that did not included in the terms of a good test and which one of the item did not included one or more requirements of a good test, it was analyzed on its error and the writer submitted for its improvement.

### **3.8.2 Quantitative Data Analysis**

The Quantitative analysis was determined of the collected data from based questioned about the validity, reliability, difficulty levels and power discrimination. The data analyze would be managed by usage the SPSS and the formulated of statistical theory which of the data from students' score and paper sheets that writer got.

#### ***3.8.2.1 Items Difficulty Analysis***

The items difficulty analysis was used to known the quality in each items test. According for Sudjana (2011:137) that the analysis of items difficulty looked from students' perspective in their ability and understanding to answered the questions test. It was evaluate and determined the proportion and criteria of each question that included of easy, medium and hard. Moreover, to measure the item difficulty was found the difficulty index or level of difficulty, the researcher was found to as formula below:

$$I = B/N$$

**Notes:**

I = the difficulty index for each items test

B = Number of test-takers who answered correctly in each items

N = Number of test-takers' responding in each items

The criteria used were the smaller index obtained that the test was more difficult and conversely, the bigger index was easier it would be. The criteria index of difficulty level as follows:

**Table 4 - Table 3.8.2.1 Index of Difficulty Levels**

No.	Index of Difficulty Level	The Categories
1.	0 – 0.30	Hard
2.	0.31 – 0.70	Medium
3.	0.71 – 1	Easy

*(Sudjana, 2011:137)*

The table above to found the report and conclusion after the researcher analyze the data from the data of 3 schools each sub-district in South District Gresik that applied in the 93 test-takers from fifty questions of multiple choice items.

### **3.8.2.2 Discrimination Power Analysis**

Analysis of items discrimination has a purpose to known the ability of the questions in students distinguished who has categorized in high achievement between students who has classified as less or weak achievement. Items discrimination as discrimination power is a group which an item differentiates

between high and low ability test-takers (Brown, 2004:59). And before calculating the discrimination power, the writer found of the rank from the students to known the low- and high-group. It was the rank formula as follows:

$$27\% \times N \text{ (Total of test-takers)}$$

Another ways and table from Ross & Stanley to analyze the items difficulty level as likes:

**Table 5 - Table 3.8.2.2 Rose & Stanley of Items**

Percentage	Option				Categorizes
	2	3	4	5	
<b>16</b>	0.16n	0.213n	0.24n	0.256n	Easy
<b>50</b>	0.50n	0.667n	0.75n	0.80n	Medium
<b>84</b>	0.84n	0.20n	1.26n	1.344n	Hard

**Notes:**

2 option for true-false question

3 – 5 option for multiple-choices question

N is 27% from total number of the test-takers

The method to discriminate power analyzed by using the following formula:

$$SR - ST$$

Where:

SR = Test-takers in lower-group who answered incorrectly

ST = Test-takers in high-group who answered incorrectly

The criteria of discrimination power could be seen in the table below which was proposed by Ross & Stanley:

**Table 6 - Table 3.8.2.2 the Items Test Discrimination Power of the Index (Ross & Stanley)**

Number of Test (N)	N (27% N)	Option			
		2	3	4	5
28 – 31	8	4	5	5	5
32 – 35	9	5	5	5	5
36 – 38	10	5	5	5	5
<b>Etc. =&gt; Seen on the table of Ross &amp; Stanley</b>					

*(Sudjana. 2011:142)*

Then, the steps before the writer analyze using of discrimination power's formula were correction the entire students paper sheet, made a list of the students' score based on the rank that students get. Therefore was determined the samples until 27% from numbers of test-takers for students in high-group and low-group. Do items analysis was calculated the wrongly answered from students in high- and low-group. Henceforth, counted the differences of test-takers who students that answer incorrectly of the high-group and low-group. Compare the score taken from the value at the r table Ross & Stanley. The last, made a conclusion whether the test items has "discrimination power or not" in each question with criteria "had discrimination power" if the value of the differences the test-takers who answered wrongly between low-group and high-group has equal or not than table' value.

### 3.8.2.3 Analysis of Validity

The validity test could found the item-tests of the valid or invalid of the items test. It was discussed a test to measured which enough the requirements of the validity test to measure the students' ability. The writer used with the formula as follows:

$$r = \frac{n(\sum XY) - \sum X \sum Y}{\sqrt{[n\sum X^2 - (\sum X)^2][n\sum Y^2 - (\sum Y)^2]}}$$

(Priyanto, 2008:18)

#### Where:

- rx<sub>y</sub> = correlation coefficient items-total (Bivariate person)
- N = number of test-takers
- ΣX = number of test items
- ΣY = total score of test items
- ΣXY = multiplication of items score and total score
- ΣX<sup>2</sup> = quadrate of number of test items
- ΣY<sup>2</sup> = quadrate of total score of test items

The other formula of validity item-test as follow:

$$r_{i(x-i)} = \frac{n \sum i(x-i) - (\sum i)(\sum (x-i))}{\sqrt{[n \sum i^2 - (\sum i)^2][n \sum (x-i)^2 - (\sum (x-i))^2]}}$$

(Priyanto,2008:22)



Where:

- i** = Item score (correlation)
- (x-i)** = Total score item minus item score
- n** = Total of subject

If the writer used formula above, the variable (x-i) or total score item minus number of score item. The tested used two-sided tests with the significant of alpha 0.05. The criteria of test as follows:

- a. If  $r \text{ count} \geq r \text{ table}$  ( two-sided tests with 0.05 sig), then instrument or correlation items with score total is valid
- b. If  $r \text{ count} < r \text{ table}$  (two-sided tested with 0.05 sig), then instrument or correlation items with score total is not valid.

#### ***3.8.2.4 Analysis of Reliability***

This analyzes to known what the test should usage to measure more than two times or not. And the writer measure reliability of multiple-choices test items used the Alpha formula below:

$$r = \left[ \frac{k}{k-1} \right] \left[ 1 - \frac{\sum \sigma_b^2}{\sigma_t^2} \right]$$

(Priyanto. 2008:25)

**Where:**

$r_{11}$  = Reliability instrument

$k$  = Number of test-items

$\sum \sigma_b^2$  = Total of test-items

$V_t^2$  = Total of variants

The significant test was performed at alpha 0.05 would be significant. Criteria of a research instrument was reliably, if the alpha's value was greater than the product in r's table, or used a limited of less than r's table. The test was performed using SPSS 16 software.