

CHAPTER I

INTRODUCTION

1.1 The Background of the Study

In education quality, one of the important components is evaluation. The evaluation is defined as the process in making decision and solution for education process based on the result of tests, other assessments, and others reports. Based on Brown & Priyanvada (2010: 9). For English language education, the evaluation is conducted in many aspects of education such as curriculum, strategies of teaching, references and test items. Thus, the good evaluation is believed will have good benefits in national education because by evaluating the system of education, the government can improve the quality of education system.

Testing refers to an effort to measure the result of students learning in teaching learning process. Consequently, the teachers should have an ability to arrange and good test and analyze of a test. Therefore, the accuracy and the carefulness of teachers may have a big impact on the increase the quality of teaching particularly in giving the judgement of students ability.

Teacher are professional educators who have the main duties to educate, teach, supervise, train, assesses and evaluate students, according to *Undang-undang Guru dan Dosen No.14, 2005*. Based on this rule, in term of evaluation, the teacher must have well competence to arrange students evaluation. The purpose of evaluation to measure students ability. In this research, the researcher will measure the ability of students reading comprehension on eleventh grade of SMA Negeri 8 Medan. In addition to that, the outcomes of learning reading comprehension can be measured true evaluation. A good evaluation is important instrument to know whether the students need a help or not. A patient from the teacher (as a help) is needed for the lesser students ability.

Test have been widely use to demonstrate level of proficiency of the students, and at the same time function as policy instruments to implement educational standards stated by Phakiti & Roever (2011 : 29). This study is a part of test development process which aims to analyse the reading comprehension test items. A teacher who is competent in terms of pedagogics must be able to make quality questions. The quality of the questions is viewed from the aspects of difficulty items, item discrimination power and level of distractors.

Based on the researcher experience during teacher training program at SMA N8 Medan. The English teacher of SMA N8 Medan distribute 15 multiple choices to the students. In term of item difficulties the researcher found that 60% categorized average, 26,6% categorized difficult, 6,6% categorized easy and 6,6% categorized too easy. Based on this fact the test above (15 multiple test) categorized is lack. In term of the discrimination power showed only 1% which means it is very low, and the result for reliability showed -1,71 categorized low and level of distractor showed 18,6% categorized poor and need to be analyze. Based on the fact above above it can be concluded that the teacher are not professional to make the test.

Based on the observation above which indicate the weakness of English teacher who teach English for grade eleventh at SMA N8 Medan, the researcher has a great desire to investigate final reading tests of grade eleventh of SMA N8 Medan by using ANATES program (software). The researcher applied this program since it is new program, effective and able to measure the quality of item test and short time. The researcher prefer to choose ANATES because it free to download and easy to use. The quality of item test is viewed from four aspectsnamely: item difficulty, item discrimination power, the distractors efficiency and reliability. As a result, the research will explain whether the final exam test is good or not.

1.2 The Problem of the Study

This research about analysis of test items in final examination attempts to answer following questions :

1. What are the qualities of test items reading in the final examination result at the eleventh grade of SMA Negeri 8 Medan based on reliability, item difficulty, item discrimination and effectiveness of distractor?
2. What is the result that indicate by ANATES to show the effectiveness of test item to evaluating to students?

1.3 The Objectives of the Study

Based on the problem statement, this study is intend to find out the qualities of test item in final examination at eleventh grade of SMA Negeri 8 Medan based on:

1. To describe the reliability, item difficulty, item discrimination power and effectiveness of distractor at eleventh grade of SMA Negeri 8 Medan.
2. To describe the effectiveness of test item to evaluating to students.

1.4 The Scope of the Study

To limit the scope of the study, the writer only discuss such as follows :

1. This study will be doing at eleventh grade students of Senior High School in the academic year 2018/2019 of SMA Negeri 8 Medan.

This study will talk about item analysis which consist of multiple choice test to analyze the quality of test items based on level of difficulty, discriminating power and effectiveness of distractors

1.5 Significances of Study

1. Theoretically

- a. Give the information about quality of English Final Examination Test for the second semester of the eleventh grade students of SMA Negeri 1 Medan in 2018/2019 academic year.
- b. Give the information about quality of English Final Examination test for the second semester of the eleventh grade students of SMA Negeri 1 Medan 2018/2019 academic based on level of difficulty, discriminating power and effectiveness of distractors.

2. Practically

- a. English Teacher, this research hope can use to improve knowledge of analyzing the test items especially in English material.
- b. Students, can help the students to know deeply about analyzing the test items theory.
- c. Other researcher, to provide reference in conducting similar research study about item analysis program.

CHAPTER II

REVIEW OF LITERATURE

2.1 Theoretical Framework

In order to answer the research questions of this study, English items test analysis of multiple choice of seventh grade at SMA Negeri 8 Medan is review use terms of the relevance

of test items to the school based curriculum for level of Senior High School, difficulty level, discrimination power and the effectiveness of distractor.

2.2 Reading

According to writer, reading is the process of looking at a series of written symbols and getting meaning from them. When we read, we use our eyes to receive written symbols (letters, punctuation marks and spaces) and we use our brain to convert them into words, sentences and paragraph that communicate something to us.

David Nunan (1989 :33) stated that reading is not an invariant skill, that there are differencet types of reading skills that correspond to the many different purposes we have for reading. So, in the classroom, in students reading activities, the writer is sure that they have many purposes, among others are to graduate from their school and to provide themselves with the knowledge to continue their studies whatever their purposes are. In order to achieve the goal, the comprehension ability in reading is needed.

According to Day and Bamford (1998 : 12) reading is the construction of meaning from a printed or written message. It means the construction of meaning involves the reader connecting information from the written message with previous knowledge to arrive at meaning and understanding.

2.2.1 Kinds of Reading

2.2.1.1 Skimming

Skimming is high speed reading that can save time and help the readers get through lots of material quickly. It is different from other kinds of high-speed reading, such as previewing, and scanning. By skimming, the readers can get general sense of passage or book, not spesific details. The goal in skimming is not to read the text throughly, but it is

used to find out the kind information that the readers seek, and if it does determine what to read more closely.

2.2.1.2 Scanning

Scanning is very high-speed reading that the reader do when they are looking for spesific piece of information. When the readers scan, the readers have questions in mind. In scanning, we do not read every word but only keywords that will answer our questions. Scanning is quickly reading to find the spesific information. Brown (2000 : 308) stated that, scanning is quickly searching for some particular piece or pieces of information in a text. By scanning, a readers means to glance rapidly through a text either to search a spesific piece information (e.g name, date) or to get an initial impression of whether the text is suitable for a given purpose.

2.2.1.3 Previewing

Previewing is something that we already do in our daily life. For example, when we receive a letter, we usually look first at the return address or the stamp to find out where it came from and who sent it. Then, we will make some guesses about what it will be about. By previewing for just a few second, we can pick up a great deal of information about the text we are going to read.

2.2.1.4 Close Reading

According to Lher F and Osborn J (2001 : 45) close reading requires carefull attention to all the words or sentences in a selection to understand it full meaning. After we have skimmed a source and decided to read all or part it closely, we read carefully to comprehend ideas and record information. While these to purposes can undoubtedly overlap, awareness of them as separate activities will help to focus.

2.2.1.5 Guessing from Context

Guessing from context is the best strategy to do when we come to a word that we don't know when we are reading. Using prior knowledge of the subject and the ideas in the text as clues to the meanings of unknown words, instead of stopping to look them up.

2.2.1.6 Paraphrasing

Richard and Schmidt (2002 : 384) state that paraphrase using other word or phrases, often in an attempt to make the meaning easier to understand. Paraphrasing stopping at the end of a section to check comprehension by restating the information and ideas in the text.

2.3 The Definition of Test

A test is supposed to be able measure learning outcome which distinguish the every single student's ability between students already mastered and not yet the learning material. Therefore, testing is one of the powerful tools to measure students abilities as well as enhance their attitudes towards learning based on Hughes (2003 : 6). According to Brown (2004 : 3) test is a method of measuring a person's ability knowledge, or performance in given domain. Test have become the tools use to complete the requirements in the process admission. Test become a crucial thing in the journey to individual's success.

The importance of test have become a crucial milestone in the journey to individual's success. Brown (2001:128)) argues that tests represent a social technology deeply embedded in education, government and business as such it provides the mechanism for enforcing power and control which can determine the future of individuals. It is clear that test becomes an important part of individual's progress.

Talking about test it can not separate from assessment. We may be attempt to think that "testing" and "assessing" are synonymous, but they are not. According to Linn and Gronlund

(2000: 31-32) assessment is a broad term which covers the full range of procedures used to obtain information about student's learning through observation. Ratings of performances or projects, and paper and pencil test. It is also covers the value judgments formation focusing the learning progress. While, a test is a spesific type of assessment which typically consists of a set of questions conduct during a fix period of time under reasonably comparable conditions for all students.

Based on the definitions above, it can be concluded that a test is a particular type of assessment which is use to measure student's behaviour or performance by posing a set of questions conduct during a fix period of time under reasonably comparable conditions for all students.

2.4 Kinds of Test

There are many types of test to measure students achievement. According to Nurgiyanto (2001 : 60-64) kinds of test divided into teacher made test and standardize test.

2.4.1 Teacher Made Test

Teacher made test is a test which has arranged by teacher this test is intended to measure students achievement after teaching process conducted by teacher. Test items arrangement as intended to measure students achievement generally done by a teacher in certain object study. Test items arrangement basically related on purpose and description of the material which has taught. It is possible have differences between one teacher and another teacher although they have some object study. Teacher made test has a limited scope and limited students score achievement.

Analyzing and revising of item test is rarely done by teacher that is why the reliability of teacher made test is low. This weakness can be finished if teacher want to learn

and apply the technique of arranging and processing of assessment result. It purposed that this test intended to determine 1) purpose achievement level, 2) students comprehension level 3) to give score to students' outcomes report.

2.4.2 Standardized Test

Standardized test can be divided into two test, they are aptitude test and achievement test. The procedures used to gain the content of achievement test is quite different with aptitude test. Achievement test used to gain the knowledge and skill which have learned in educational level and the items used for material measurement. Standard in the test meant that do by all of students with same limitation. It arranged by a team. After arrangement finished, it has tryout to several students. Tryout results will be analyzed by using item analysis to know the difficulty level and discriminating power. Easy item or difficult item will be revised or discarded.

Standardized test used by all schools in same region. It usually has been completed with directions about test implementation and scoring system. This test used to : 1) then information about students outcomes 2) compare students achievement in every school. Other kinds of test based on the function in learning activity could be classified into four kinds, they are formative test, summative test, diagnostic test, and placement test (Purwanto, 2011 : 67-70).

2.4.2.1 Formative Test

Formative test is intended as test which used to determine students comprehension after learning activity. Every learning program or learning material must comprehend the student's in line with the learning purpose which has arranged. Formative test is examined to determine the effectivity of teaching learning process which has comprehend by students.

Formative test in learning practice called as *ulangan harian* or daily test. In teaching planning, learning process is a sub material is planned in a lesson plan. Lesson plan covers learning purpose, material, method, learning strategy, media and evaluation. Evaluation which is planned in a lesson plan is evaluation that is done based on formative test. Formative test given in the end of learning program, so it could be called as post test.

2.4.2.2 Summative Test

Summative test is intended as a test used to determine student's comprehension toward all of materials which have given in a certain period time such as mid semester and semester. In teaching practice, summative test known mid-term test and final test semester.

2.4.2.3 Diagnostic Test

Learning outcomes evaluation has diagnostic function. Test which has used to do diagnostic evaluation is diagnostic test. Diagnostic test used to identify student's difficulty and investigate kinds of difficulty which is faced. Teacher could solve the students difficulty with appropriate solution. Example of diagnostic test is to teach descriptive test, teacher should be sure if students comprehends about simple present tense. Before starting about descriptive test, teacher gives diagnostic test to determine student's comprehension about present tense.

2.4.2.4 Placement Test

Placement test is intended to collect the data about learning outcomes test which is needed to grouping students with their interest and their talent appropriately. This grouping is done in order to give appropriate materials with their interest and talent. In learning practice, placement test used for example in Senior High School to grouping students in

IPA, IPS, Bahasa. Placement test serves data to put students in a class appropriate with their interest and talent.

2.5 Reading Assessment

Assessment is a systematic approach to collecting information and making inferences about the ability of a student or the quality or success of a teaching course on the basis of various sources of evidence. To collect the information of the students' improvement in reading comprehension, some assessment of reading are essential to be carried out. Unlike speaking and writing, the reading process and product cannot be seen and observed specifically. For this reason, all assessment of reading must be carried out by inference. Some considerations are needed in designing assessment of reading, such as the types of genres of written text, the components of reading ability and specific tasks.

Furthermore, the types of reading performance will influence the assessment tasks as well. Brown lists a number of possible tasks for assessing perceptive, selective, interactive and extensive reading. Related to the specifications of reading in this study that is assessing students' interactive reading performance so there are three types of possible assessment tasks applied as follows :

1. Multiple-Choice

Multiple choice in this study provides not only the vocabulary and grammatical items but also the context to assess the students' understanding of information in the text. The context is presented by putting a pair or part of a text followed by questions in which the students have to respond correctly.

2. Impromptu Reading Plus Comprehension

This type of assessment involves impromptu reading and responding to questions. It is commonly used in proficiency test. In this test, students are provided a reading passage followed by questions and have to respond to the items. The set of questions in impromptu reading covers the comprehension of some features of reading : (1) main idea (2) expressions/idioms/phrases in context (3) inference (4) grammatical vocabulary in context. These specifications and the questions are in line with strategies of effective reading :skimming for main idea, scanning for details, guessing word from context, inferencing using discourse makers etc which are assessed in this study.

3. Short Answer Task

In this type of assessment, reading passage is presented, and the students read questions that must be answered in a sentence or two. The question might cover the same specifications similar to the impromptu reading. Those three types of assessments task have the combination of form-focused and meaning-focused objective. They cover the objective of reading assessment especially in the comprehension issues and embody the evidences of students reading comprehension.

2.6 Item Analysis

Item analysis is a process which examines students responses to individual test items (questions) in order to assess the quality of those items and the test as a whole. It is a valuable, powerful technique available to teaching professionals and instructors to increase their test construction skills, identify specific areas of course content which need greater emphasis or clarity, and improve other classroom practices. According to Thompson and Levitov (1985 : 163) item analysis investigates the performance of items considered

individually either in relation to some external criterion or in relation to the remaining items on the test.

Students may find various types of test in the class in periodically assessment may be found the test with instruction such as answer briefly, fill in the brackets, choose the best answer, or match the word. Students may be seen another instruction like discuss with your friends, say the words with correct pronunciation, or practice the dialogue with your pair. These instruction reflect that there are several types of test. Based on the instruction, there is subjective test and objective test.

2.6.1 Subjective Test

Subjective test commonly called as essay (*uraian*). Essay test is a type of test needs to explain the answer in words. Students allows to select, construct and present the answer with their own words. The characteristic of this test is started by : explain, why, how, compare, conclude, etc.

Test with essay forms needs 90 until 120 minutes in 5 until 10 questions. This test demands student's comprehension to organizing, interpreting and connecting the meanings and statements. It is useful for students to remind, recall and have high creativity.

1) The advantages of subjective test :

- a. Easy to prepared and arranged
- b. It is not give too much opportunity to speculate or guessing the answer
- c. Push the students to be brave in proposing their opinion and construct it in good sentences.
- d. Give the opportunity to students to present their statement with their own words and their own way

- e. Know how far student's comprehending with the problem which is tested
- 2) The disadvantages of subjective test
- a. Validity and reliability level is low because difficult to determine which aspect of student's ability which has comprehend
 - b. Less of representative in explaining all of materials scope because the question tested is limited.
 - c. The way to asses is influenced by subjectivity aspects.
 - d. Assesment is more difficult because need individuals judgment from assesor
 - e. Need more time in correcting and it can not replace by other assesor
- 3) The guideline in arranging
- a. Test item could cover main idea from the material which is tested.
 - b. Test item shoud not take from a book directly. It needs to be served with other words.
 - c. When arranging, test item have been completed with answer key and assesment direction
 - d. More various question like " explain", "how", "why", in order to know students comprehension with material.
 - e. Question formulation should be easy to be understand
 - f. The question should be spesific.

Example :

Explain about the story in this book

This question is less spesific. It is better if add statement become:

Explaining the story in this book on page 34 entitled " Cinderella", please explain :

- a. Setting

- b. Plot
- c. Characters
- d. Point of view
- e. Weather

2.6.2 Objective Test

Objective test is a test which is with objectively checking. It is intended to solve the disadvantages of essay. Objective test uses more questions than essay. Objective test needs 60 minutes to answer 30 until 40 items

1) The advantages of objective test

- a. Contain more positive aspect, like more representative materials, more objective and avoid subjective elements.
- b. Easier and quick to correct the answer.
- c. Assessment can be replace by other assesor.

2) The disadvantages of objective test

- a. Preparation in arranging the test is more difficult than subjective test because the item is much.
- b. The item only used to recall the memory.
- c. Give opportunity to guess answer.
- d. Easier to cheat during doing the test.

3) Kind of subjective test

- a. True – false

True false test serves statement. The statement could be correct and could be incorrect. Student has to put a circle in “ T “letter if statement is correct and put and put a circle “F” letter if the statement is incorrect.

Example :

T – F. Objective test gives opportunity to students in guessing the answer

True- false form divides into two kinds, as follow :

1. With correction, students should give the correct answer if they choose “ false”
2. Without correction, students only put a circle in “T” or “F” letter without give the correct answer.

The advantages of true false test :

- a. Cover wide material and serve in short statement.
- b. Easy to arrange.
- c. Can be used frequently.
- d. Can be seen quickly and objectively.
- e. Easy to understand.

The disadvantages of true – false test :

- a. Confusing
- b. Easy to guess
- c. Only intens on recalling the memory

2.6.3 Multiple Choice Test

Multiple choice test consist of an information about incomplete definition. To complete the definition, students should chose one from some possible answers. Multiple choice consist of information (stem) and the possible answer (options). The possible answer (option) covers one correct answer and two until four distractors.

The advantages of multiple choice :

- a. It easier to construct high-quality test items.
- b. Despite its superiority in many respects to other simple form objective test.
- c. It requires selection of the correct answer.

The disadvantages of multiple choice :

- a. The difficulty of finding a sufficient number of incorrect but plausible distractors.
- b. Its difficult in arrangement.
- c. Items only measure memory aspect.
- d. The effect of students habit toward multiple choice item (testwise) in students' outcomes (Linn and Gronlund, 2009 : 203-204)

2.6.4 Matching Test

Matching test consist of a set of question and a set of answer. Each queastion has an answer included in a set of answer. Student finds and puts the answer in certain question. The guideline in arranging matching test :

- a. A set of question should not more than ten questions because its confusing students and decreasing the homogeneity of question.

- b. The number of answer should have more number than the number of question. It push students to thinking more.

Example :

Match the cities and the name of province :

- | | |
|-----------------|---------------------------------|
| 1. Cirebon | a. North Sumatera |
| 2. Demak | b. West Nusa Tenggara |
| 3. Pasuruhan | c. East Kalimantan |
| 4. Lubuklinggau | d. West Kalimantan |
| 5. Depok | e. West Java |
| 6. Singaraja | f. North Sulawesi |
| 7. Balikpapan | g. Central Java |
| 8. Martapura | h. East Nusa Tenggara |
| 9. Gorontalo | i. Central Sulawesi |
| 10. Ende | j. South Kalimantan. Yogyakarta |
| | l. East Java |
| | m. Bengkulu |
| | n. Jakarta |

2.7 Multiple Choice Item

Multiple choice test consists of a problem and a list of suggested solutions. The problem or a notice about an incomplete notion (Linn and Gronlund, 2009:194). The multiple choice format for test items is characterized as follows:

- a. The problem may be stated as a direct question or incomplete statement (the stem of the item).

b. The list of suggested solutions may include words, numbers, symbols, or phrases are called alternatives (also called choices or options). The correct alternative in each item is called the answer and the remaining alternatives are called distractors (also called decoys or foils) stated by Linn and Gronlund (2009: 194)

a. Avoid repetition in options.

Example:

When is paprika added to the omelet?

- A. It is added before heating the skillet.
- B. It is added before cutting the meat.
- C. It is added while blending the ingredients.
- D. It is added after cooking the omelet.

Should be

- A. Before heating the skillet
- B. Before cutting the meat
- C. While blending the ingredients
- D. After cooking the omelet

b. Omit the articles in options

Example:

What part of Grendel did Beowulf remove?

- A. a toe
- B. the nose
- C. the head
- D. an arm

should be

- A. toe
- B. nose
- C. head
- D. arm

c. Handling measurement units

Example:

What is the length of the line?

- A. 2 inches
- B. 3 inches
- C. 4 inches
- D. 5 inches

Should be

- A. 2
- B. 3
- C. 4
- D. 5

Unless the objective to be tested requires the understanding of standard abbreviations for units of measure, the units should not be abbreviated in the stem or options:

1 mm

Should be

2 millimeters

d. Use third person

Example:

According to American Epileptic Society standards, what first aid should you perform when you witness someone having an epileptic seizure?

- A. You should keep the person on his back.
- B. You should turn the patient on his side.
- C. You should give the person something to drink.
- D. You should put something between the person's teeth.

Should be

According to American Epileptic Society standards, what first aid should be performed when someone has an epileptic seizure?

- A. Keep the person on his back
- B. Turn the patient on his side
- C. Give the person something to drink
- D. Put something between the person's teeth
- e. Italics and boldface

Italics are employed in accordance with conventional usage for foreign terms and for the names of books, plays, long poems, periodicals, paintings, and films. Example:

Which is the best source for finding the 1980 population of Istanbul, Turkey?

- A. The World Almanac
- B. Compton's Picture Encyclopedia
- C. The Guinness Book of World Records
- D. Webster's School and Office Dictionary

Contrary to conventional usage in writing, however boldface rather than italic type is used to designate words as words, as for example:

Read the paragraph below and answer the question that follows.

After several weeks at sea, the ship ran into bad weather. The crew had to work hard and their tempers grew mean. When the captain discovered the first mate sleeping on the job, he was furious. Then he learned that the mate had passed out from overwork. Everyone on board was eager to see fair skies again.

What does **overwork** mean in this paragraph?

- A. Hard work
- B. Lack of work
- C. Work above
- D. Too much work

Boldface type should also be applied if the words appear in the options. Example:

Read the paragraph below and answer the question that follows.

Ms. Theodore is trying to weasel out of paying a legitimate bill by saying we did not do the work on her car properly. I cannot believe that our respected legal system would let someone who has an honest debt evade paying it.

Which is a loaded word in this paragraph?

- A. weasel because it makes Ms. Theodore seem sneaky
- B. properly because it makes Ms. Theodore seem picky
- C. bill because paying bills is unpleasant
- D. legitimate because it is a formal word

f. Ordering options

Example:

$$35 + 50 = \dots$$

A. 85

B. 86

C. 87

D. 88

Heaton (1990: 27) shows the positive effect of using multiple choice in measuring learning outcome. It can measure various learning outcomes, from simple to complex. Some advantages of multiple choice item are good for testing knowledge of grammar, vocabulary, rather than the ability, measuring students' ability to recognize grammatical form. Not only advantages, multiple choice item has negative criticism too. It is constructing negative characteristics especially reduce students knowledge and creativity.

Linn and Gronlund (2009: 195) give a brief explanation dealing with the selection one of form of stem. The selection of direct question and incomplete statements as the form of stem in each item depends on some factors. The use of direct-question is easier to write, more natural for younger students, and more likely to present a clearly formulated problem. On the other hand, the incomplete statement is more concise. From the statements above, the multiple-choice item consists of two parts: (1) the stem, which presents the problem in the form of direct-question or incomplete statement, (2) options, which covers list of possible answers, in which one of the option is the correct answer and the others are distractors.

2.8 ANATES Software

Analysis data refers to find and set systematically the data which have collected from the collecting data process to be easily to understand in this study. The analysis quantitatively is doing to order investigate the test items from students answer or response and the key answers. Anates is a software for statistic calculation. Anates V4 develop by Drs.Kanoto M.Pd and the programmer is Yudi Wibisono ST. This software very effective to calculate about reliability, item difficulty, item discrimination power, and level of distractor. The advantages in using this program is this program can be use to analyze the test item of multiple choice and it really helps for investigate the data.

How to use anates software :

1. Open the application and the application will ask how many question and how many students then fill it.
2. Input the answer key of each number of test
3. Input the student's answer one by one
4. Then data will process and give the result based on item difficulty, item discrimination and level of distaractor that you choose before.

2.9 Previous Research

This research is conducting based on the previous study. The first one is taken from Sari (2016) with entitled “ Item Analysis of English Mid-Term for the Second of the Seventh Grade Students of SMP Negeri 2 Wonosari in 2015/2016 Academic Year”. She discussed about the quality of multiple choice item in mid-tem of seventh grade students and how she gets the result from multiple choice test. The similarity between Sari's research and the writer's research is use some test, that is multiple choice item, using same skill that is reading . The differences of this research is the writer will doing this research in Senior High School but Sari done her research

in Junior High School, research setting (sample, place and time) is different, Anis use 35 of multiple choice test while the writer use 50 items test and the way of analyzing data. Sari's research use ITEMAN Software while the writer use ANATES software and formula to analyze the data.the reference from this research are theory based on some experts and the formula (using the same formula), the source of book and the example of test.The result are mostly 68% the quality of test item is good and other is poor. While the references from this thesis are some theory based some experts, formula of the technique of analyzing data.

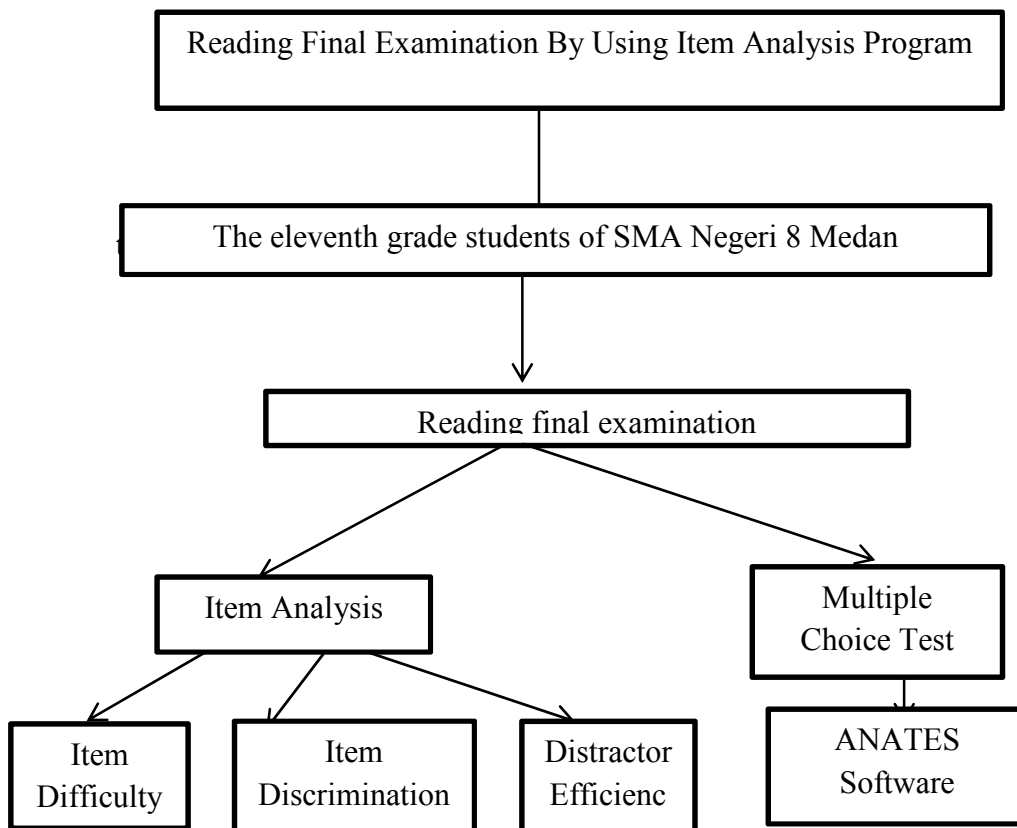
The second one taken from Danuwijaya (2018) with entitled “ Item Analysis of Reading Comprehension Test for Post-Graduate Students”. He discussed about reading comprehension to improve item quality and investigate item quality of the test. The similarity between Danuwijaya's research and the writer are the kinds of test is multiple choice, using formula. The differences are Danuwijaya has done his research in post graduate, using software called Lertap and Classical Test Theory, research setting (time, sample and place) at Universitas Indonesia, while the writer will doing this research in SMA Negeri 8 Medan at June 2019, using ANATES software and the writer use 50 multiple choice but Danuwijaya use 100 test of multiple choice. The references is the method of collecting data, theory, explaining about item analysis. The result showed that item difficulty level was most average.

2.10 Conceptual Framework

Test have been widely use to demonstrate level of proficiency of students. The quality of test is determined with the test item through analysis process. A test can be accept as a good test if it has the good item difficulty, test development etc. Item analysis is a crucial part in a test development process as it functions to provide information about items that should be improved

in terms of quality for later tests or even be eliminated due to misleading. An item analysis program use

to describe the quality of items. This research doing to find out the analysis result.



CHAPTER III

RESEARCH METHODOLOGY

3.1 Research Design

The research design of this study will be using Qualitative-Quantitative Research. According to Jhon (2009: 1) qualitative inquiry employed different philosophical assumptions, strategies of inquiry, and methods of data collections, analysis and interpretation. The goal of

qualitative is collecting the data, arranging the data and interpreting the data. The analysis of the data will take from the students reading final examination (multiple choice item). Quantitative research is a process to find the data in form a number as a tool to find the information related to the objective of the research. Both of this used to obtain the quality of multiple choice item which is used in the English final examination test.

Qualitative approach in this research use to show which one is good item, average item and poor item of multiple choice item. Quantitative approach use for identifying the level of difficulty, the discriminating power and the effectiveness of distractor of eleventh grade of SMA Negeri 8 Medan.

3.2 Population and Sample

The research will be doing at SMA Negeri 8 Medan and the sample of this research is students who are in the eleventh grade of SMA Negeri 8 Medan on June 2019.

1. Population

In this research the population are answer sheets of students who take English final examination which test make by the English teacher that consist of 30 items of multiple choice. The total of students in SMA Negeri 8 Medan 1000 students. Because the students who participate in test are more than 100, in this research, the writer will take only some students answer sheet as the sample.

2. Sample

Sample is a part of characteristic which have by the population. The technique of sampling in this research is sample proportional random. The writer choose the sample randomly from the test result for each class.

3.3 Instrument of Data Collection

Instrument of the problem in this research using instrument reading multiple choice item tests. Students will ask to answer multiple choice test. The test are administrate to get the data of student. To help their test, they are allowing to use dictionary and grammar book to help them found out the answer of the test. Then the answer will input in ANATES software to determine the result of level of difficulty, item discrimination and effectiveness of distractors.

3.4 The Technique of Collecting Data

The data will be collecting from the students multiple choice item sheet that has divide before. The test will classified by difficult classification(item difficulty, item discrimination and effeiciency of distractor). The writer will give the instruction to ask the students answer the test and tell how long it to finish the test.

The data will be collect by using these following step :

1. Answering the multiple choice with the correct answer.
2. Collecting all the answer paper and question sheet after finished.
3. Arrange all the scores from the highest to the lowest score.

3.5 The Technique of Analyzing Data

There are some formula to analyzing data in multiple choice item :

3.5.1 Computing Level of Difficulty

To get the difficulty level of the test item, the formula is considered appropriate in the analysis.

$$P = \frac{B}{JS}$$

Where :

P = difficulty index

B = the number of students who got the item right

JS = the total number of students who tried to answer the item

Difficulty level of an items is judged through the following range

Table 3.1 Item Difficulty Index

Difficulty Index (P)	Interpretation
1.00 - 0.30	Difficult
0.30 - 0.70	Satisfactory
0.70 - 1.00	Easy

3.5.2 Computing Item Discrimination Power

This is the formula to get the item discrimination power on test item :

$$\frac{BA - BB}{JS}$$

$$D = \frac{JA - JB}{PA - PB}$$

Where :

D = Discrimination Power

JB = the number of students in lower group who tried the item

BA = the number of students in upper group who answer correctly

BB = the number of students in lower group who answer correctly

PA = index of difficulty in upper group

PB = index of difficulty in lower group

The maximum size of the index is +1.00 and the minimum is -1.00. The negative value means that the test item discriminates in the wrong direction and is not satisfactory. Positive value shows that the test item discriminates in the desired direction and it is satisfactory. Discriminating power of an item is judged through the following range :

Tabel 3. 2 Item Discriminating Power

Discrimination Power (D)	Interpretation
0.00 – 0.20	Poor
0.20 – 0.40	Average
0.40 - 0.70	Good

0.70 – 1.00	Excellent
-------------	-----------

3.5.3 The Effectiveness of Distractors

The distractors can function well if at least chosen by 5% of all learners who participate in test. Criteria for assessing the use of distractors can adapt based on Arikunto (2005 :220) are :

- a. Acceptable, because it has been chosen at least 5% the total number of students who try to answer. A good distractor can distract more students from lower group than students from upper group.
- b. Refused, because it does not exist by the students to choose the distractor.
- c. Revision, because it attracts more students from upper group than students in lower group. It is revision distractor if choose by less of 5% of the total number of students who try to answer. A distractor can be called revision only on the sentence structure. It can be re-write the effective revision.

3.6 Data Analysis Procedure

The procedure of analysis data :

1. Distributing multiple choice test
2. Collecting the answer sheet after they finish to answer the question.
3. Analyze the score based on item difficulty, item discrimination power and effectiveness of distractor based on the result that showed by the ANATES program.
4. After get the result, analyze the test item examination that which the test consist of item difficulty, item discrimination power and effectiveness of distractor and which the test item that need to use and not use (need revise).

5. Make a conclusion.