An Item Analysis of English Summative Test on Difficulty Level, Discriminating Power, and Distractor Effectiveness

(A Study of the Seventh Grade Students of SMP N 3 Ungaran   in the Academic Year 2018/2019)

a final project

submitted in partial fulfillment of the requirement

for the degree of *Sarjana Pendidikan* in English

by

Niken Nuraini Munifa

2201412010

ENGLISH DEPARTMENT

FACULTY OF LANGUAGES AND ARTS

UNIVERSITAS NEGERI SEMARANG

2019

APPROVAL

This final project entitled An Item Analysis of English Summative Test on Difficulty Level, Discriminating Power, and Distractor Effectiveness (A Study of the Seventh Grade Students of SMP N 3 Ungaran in the Academic Year 2018/2019) has been approved by a board of examiners and officially verified by the dean of the faculty of Language and Arts on 22 August 2019.

1. **Chairman**
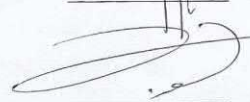   Dr. Hendi Pratama,S.Pd., M.A.
   NIP 198505282010121006

2. **Secretary**
   Bambang Purwanto, S.S., M.Hum.
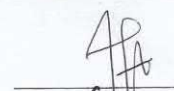   NIP 197807282008121001

3. **First Examiner**
   Arif Suryo Priyatmojo, S.Pd., M.Pd.
   NIP 198306102010121002

4. **Second Examiner**
   Bambang Purwanto, S.S., M.Hum.
   NIP 197807282008121001

5. **Third Examiner/ Advisor**
   Galuh Kirana Dwi Areni, S.S., M.Pd.
   NIP 197411042006042001

Approved by
The Dean of the Language and Arts Faculty

Dra. Sri Rejeki Urip, M.Hum.
NIP 196202211989012001

ii

PERNYATAAN

Dengan ini saya,

Nama   : Niken Nuraini Munifa

NIM   : 2201412010

Prodi/ Jurusan : Pendidikan Bahasa Inggris S-1/ Bahasa dan Sastra Inggris

Fakultas Bahasa dan Seni Universitas Negeri Semarang, menyatakan dengan sesungguhnya bahwa skripsi/ tugas akhir/ final project yang berjudul:

An Item Analysis of English Summative Test on Difficulty Level, Discriminating Power, and Distractor Effectiveness (A Study of the Seventh Grade Students of SMP N 3 Ungaran in the Academic Year of 2018/2019)

yang saya tulis dalam rangka memenuhi salah satu syarat untuk memperoleh gelar sarjana ini benar-benar merupakan karya saya sendiri, yang saya hasilkan setelah melakukan penelitian, pembimbingan dan diskusi, dan pemaparan atau ujian. Semua kutipan, baik yang langsung maupun tidak langsung, baik yang diperoleh dari sumber kepustakaan, wahana elektronik, maupun sumber lainnya, telah disertai keterangan mengenai identitas sumbernya dengan cara sebagaimana yang lazim dalam penulisan. Skripsi ini membubuhkan tanda tangan sebagai keabsahannya, seluruh isi karya ilmiah ini tetap menjadi tanggungjawab saya sendiri. Jika kemudian ditemukan pelanggaran terhadap konvensi tata tulis yang berlaku, saya bersedia menerima akibatnya.

Demikian, surat pernyataan ini dapat digunakan seperlunya.

Semarang, 23 Agustus 2019

Niken Nuraini Munifa

NIM. 2201412010

MOTTO AND DEDICATION

You can't skip chapters.

That's not how life works.

You have to read every line.

Meet every character.

You won't enjoy all of it.

Some chapters will make you cry for weeks.

You will read things you don't want to read.

You will have moments when you don't want the pages to end.

But you have to keep going.

Stories keep the world revolving.

Live yours, don't miss out.

-Courtney Pappernill

For

My beloved family,

Bapak, Ibun and my brother.

ACKNOWLEDGEMENTS

First of all, I would like to express my greatest gratitude to Allah Subhanahu Wata'Ala, the Almighty, for all the blessing given to me during the accomplishment of my final project. Then, shalawat and salaam are only given to Prophet Muhammad Shalallahu 'Alaihi Wasallam, the best teacher in this world.

I present my sincere gratitude and appreciation to my kind-hearted advisor, Galuh Kirana Dwi Areni, S.S., M.Pd., who give ideas, times, suggestions and guidances to me so patiently during this final project writing. I would also like to extend my gratefulness to all lectures of the English Department and the whole staffs of the Faculty of Language and Arts, Universitas Negeri Semarang who taught, guided and help me during the study so that I can overcome a lot of difficulties and learn many things.

Most importantly, I would like to give my deepest gratitude to my parents for their love, prayers, support and anything they have done for me. And also to My beloved brother for supporting me.

Lastly, my sincere thanks is given to all of my friends in English Department 2012 especially Izzun, Riyan, Shofy, Roti, Yusti, Rosi, Pas, Wahyu who always share their laughter and support me in past years. In addition I would like to thanks to everyone who has helped me in compleating my final project.

ABSTRACT

Munifa, Niken Nuraini. 2019. An Item Analysis of English Summative Test on Difficulty Level, Discriminating Power, and Distractor Effectiveness (A Study of the Seventh Grade Students of SMP N 3 Ungaran in the Academic Year 2018/2019). A Final Project, English Department, Faculty of Languages and Arts, Universitas Negeri Semarang. Advisor : Galuh Kirana Dwi Areni, S.S, M.Pd.

Keywords: *summative test, multiple choice item, difficulty level, discriminating power, distractor efficiency*.

The aim of this study is to find empirical evidence of whether or not the summative test items have good characteristic of test in terms of difficulty level, discriminating power, and distractor efficiency. This study was held in the seventh grade of SMP Negeri 3 Ungaran. The study design used in this study was Item Analysis Research. The respondent of this study are 100 students. The writer chose the top 27% for the upper group and top 27% for the lower group. Then, the data analysis used a formula from Arikunto's theory. After conducting the research, the writer found that from 40 items, there are 15 (37.5%) acceptable items to meet the criteria of index difficulty. Besides there is 1 (2.5%) too difficult item and there are 24 (60%) unacceptable items due to easiness of the items. For the discriminating index, the writer found there are 17 (42.5%) poor items, 1 (2.5%) items has worst result, which has to be discarded, 11 (27.5%) acceptable items, and 11 (27.5%) mediocre items test. Moreover for the distractor effectiveness, the writer found there are 50 (47.67%) distracters with functioning distractor and 70 (58.33%) distracters with non-functioning distractor. In conclusion, English summative test for seventh grade of SMP N 3 Ungaran did not meet the criteria of effective and acceptable test.

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION

This chapter presents background of the study, reasons for choosing the topic, limitation of the study, purpose of the study, research question, significance of the study, and outline of the report.

1.1 Background of the Study

Evaluation is an important part of teaching learning process. It is conducted to make sure whether the teaching learning process has been running well over the term or not. In education, evaluation becames the important role because it shows the result of learning process. The objective of the evaluation itself is to help the teacher ascertain the degree to which educational objectives have been achieved, to review the effectiveness of teaching method and to help the teacher knows their pupils as individual.

According to Russel & Airasian (2008), teacher can make an evaluation through assessment and measurement. Once assessment information is collect, teacher use it to make a decision, reflection, or judgment about pupils, instruction, or classroom climate. Evaluation is one of the steps that cannot be separated in teaching learning process. In another words evaluation can be defined as a system, a decision, or a judgment about the process in determining the value of student's performance based on information that had been collect, synthesized and reflected by the teacher.

There are several methods in getting the data for the evaluation purpose. One of them is by using a test. This test could be a teacher-made test or standardized test. In the teacher-made test, the teacher who makes the test should know and master the principles and the steps that must be done in making the test. By this knowledge the teacher will get a clear figure about the general systematic framework of evaluation.The test items are supposed to be well constructed so it can be used efficiently. To be an effective test, it has to fulfill the criteria of a good test. They are validity, reliability and practicality. Brown (2003) stated that the test is recognized as valid if it can measures what supposed to be measured. It is reliable if the result of the test is similar even the test administered with the same standard for several times. A test is practical if it easy to do, and administer.

It is not an easy way to know the quality of the test item. To acknowledge whether the test has accomplished the standards of a good test, the teacher should evaluate the test item in many steps. The investigation that teacher did in order to know the quality of each item test is called item analysis. Item analysis is helpful for improving teachers' skills in the test construction and recognizing specific areas of course content that need a greater emphasis or clarity. The characteristics that determine an item analysis test are item difficulty, item discriminator, and item distracter. The item difficulty means the rank of difficulty for each item test for students. The item discriminator tells how well each item test differentiates the comprehension ability among the upper and the lower students. Lastly, item distracter indicates how effective each alternative or option for an item on multiple choice questions.

Based on the preliminary study conducted by the writer on the seventh grade students of SMP Negeri 3 Ungaran, the writer found some facts about construction and content of the evaluation. The writer found that some of the seventh grade students test questions in English final summative test academic year 2018/1019 were confusing due to ambiguous options. Some of the options are the same and the instruction of the question was not clear. Furthermore, the summative test was a teacher-made test which was designed by the teacher herself. The teacher designs the item test based on item bank by using search engine from the internet that may not fit with the local content. If the test item is not suitable to the local content, the difficulty level on the test item is not fairly distributed.

1.2 Reasons for Choosing the Topic

By giving a good test, the students will have an opportunity to get a quality result in learning process fairly. Concerning on the problems that the writer found on the preliminary study, the writer believes the presence of study that focuses on discussion about item analysis is necessary. The research will not only find out the quality of the test item which used in SMP Negeri 3 Ungaran, but also to know the weakness in each item test. The writer decided to focus on item analysis as the main topic of this study entitled "An Item Analysis of English Summative Test on Difficulty level, Discriminating Power, and Distractor Effectiveness" because it is an important key to knowing whether or not each item in English summative test has classified into a good item test.

1.3 Limitation of the Study

The discussion of the study will focus on the item analysis of English final summative test in the second semester of seventh grade students of SMP Negeri 3 Ungaran. The item analysis refers to the process of collecting, summarizing, and using information about individual test items. The items that will be analyzed are:

1. Difficulty level which helps us to decide if the test items are at the right level for the target group.

2. Discriminating Power which allows us to see how well the students know the material that has been tested.

3. Distractor efficiency which gives information whether or not the item test has an efficient distractor.

1.4 Purpose of the Study

The objective of the presents study is to find empirical evidence of whether or not the summative test items have good characteristic of test in terms of difficulty level, discriminating power, and distracter effectiveness.

1.5 Research Question

Based on the limitation of the problem, the writer conducts a research to find out the percentages of item analysis in multiple choice item. So the writer formulates the problem as follow "Do the test items of English final summative test used for seventh grade students of SMP Negeri 3 Ungaran fulfil the criteria of a good test based on difficulty level, discriminating power, and distracter effectiveness?"

1.6 Significance of the Study

The result of this study is expected to enhance the will of writer generally, and society at large to identify poor items in evaluation test through item analysis, therefore the test will give students score fairly based on their competence.

1.7 Outline of the Report

This final project report has five chapters as the following:

Chapter I is an introduction, It presents background of the study, reasons for choosing the topic, limitation of the study, purpose of the study, research question, significance of the study, and outline of the report.

Chapter II is review of related literature. It consists of previous studies and theoretical review.

Chapter III presents research methodology. It presents place and time of the study, population and sample, research instrument, method of collecting data, and method of analysing data.

Chapter IV is an analysis of data and discussion. It presents the research results and discussion.

Chapter V is conclusions and suggestions.

CHAPTER II

REVIEW OF RELATED LITERATURE

There are two subchapters in this chapter. They are review of previous studies and review of theoretical study.

2.1 Review of the Previous Studies

There are some similar researches that have been conducted by other researchers. Related to this study, the writer choose some references about previous studies which are close to the previous studies. Those previous studies discuss about item test analysis.

The first reference is according to Amelia (2010) by her research entitled "An Analysis of the English Summative Test Items in terms of Difficulty Level for the Second Year Students of MTs. Darul Ma'arif Jakarta". The purpose of this research is to measure the difficulty level of the English Summative Test items by calculation the student's correct response from the upper and lower group with J.B. Heaton's formula referred from his book "Writing English Language Test".

Research question of this research is "Does the English Summative Test items for the second year students of MTs. Darul Ma'arif Surakarta have a good quality in terms of difficulty level?". The result of this research interpreted by the Suharsimi Arikunto's criteria of items referred from his book "Dasar-dasar Evaluasi Pendidikan". There are 20 items regarded as difficult level, 20 items regarded as moderate level, and 9 items regarded as easy level.

All of the items have been counted by dividing the total of difficulty level of the items with the total number of students. The research showed that the result is 0.45. In the end of this research, the researcher has been concluded that the English summative test items for the second year students of MTs. Darul Ma'arif qualified as a good test. It can be seen from the difficulty level of all item which is in the moderate level, because it ranges from 0.30 up to 0.70.

The next research came from Prayoga (2011) who did an analysis about difficulty level of English summative test for the second grade of Junior High School in the odd semester 2010/2011 at SMP N 13 South Tangerang. This study categorized as quantitative research because the researcher used some numerical data which were analyzed statistically. Also, this study categorized as a descriptive analysis because it was intended to describe the objective condition about the difficulty level of the English summative test.

In this study the researcher took 93 students as a sample. The findings of the study were moderate items had the highest percentages with 66.7%. Followed by difficult items with 20 %, and easy items with 13.3%. Overall, the difficulty level of the test was moderate level with 0.50 index of difficulty. It means that the test had a good difficulty level.

The third research conducted by Lestari (2011). It was an item analysis about the discriminating power of English summative test at the second year of SMP N 87 Pondok Pinang. This study categorized as a descriptive analysis because it was intended to describe the objective condition about the discriminating power

by analyzing the quality of English summative test items in discriminating student's achievements. This study is considering as a quantitative research because the researcher use some numerical data which was analyzed statistically.

The researcher only took 60 students as an ordinal sampling on her study. The findings of the study showed that the English summative test which tested to the second grade students of SMPN 87 Pondok Pinang had a good discriminating power. There were 35 items ranging from 0.25 until 0.75 (70%), which means the test items full filled the criteria of a positive discriminating power.

Based on the previous related findings there are some differences between the researches. The researchers above analyzed difficulty level and discriminating power. There is one analysis of the test item that did not conduct by the researchers. Besides the difficulty level and discriminating power, the effectiveness of distractor in a multiple choice item should be concerned by the researchers.

Therefore, the writer interested to conducting a study in item analysis to overcome the weakness occurs in those previous studies. This research will prove whether multiple choice items of English summative test categorized as a good test or not. This study will hopefully, turn out quite different from those studies above, because the writer is not only analyze the difficulty level and discrimination power but also include the distractor efficiency analysis in each test items.

2.2 Review of Theoretical Study

2.2.1   The Understanding of a Test

Marsh (2010) stated that assessment is an activity done by teachers to obtain information about the knowledge, skills and attitude of students. This action can involve the collection of formal assessment data as an objective test or informal assessment data as observation checklist. In doing the assessment, there is a long process in collecting, synthesizing and interpreting the information in which testing, measurement and evaluation take parts. Brown (2003) described a test as a media to conducting the assessment. It is a method to measure the ability of a person in knowledge or performance in a given program. A test also created as a device to monitor the development of the program, to diagnose the difficulties in the program and to measure the performance of the test taker in and at the end of the program based on Sudaryono (2012) theory.

The purpose of a test are as a method to monitor the development of the program in teaching-learning process, diagnose the difficulties during and at the end of the program by measuring the ability of test taker intelligence both in knowledge and or performance. Based on the definitions above, test is very important either for the teachers nor the students. The importance for the students through a test, they will know their achievement in learning the material. While for the teachers, through a test, they will know a students who have understood the material so that the teachers can give more attention to the students who have not understood yet.

2.2.2   Types of Test

2.2.2.1 Proficiency Test

According to Heaton (1988), the proficiency test is concerned simply with measuring a student's control of the language in the light of what he or she will be expected to do with it in the future performance of a particular task. While Brown (1996) stated that a proficiency test assess the general knowledge or skill commonly required or prerequisite to entry into (or exemption from) a group of similar institution.

Arthur Hughes stated that proficiency tests are designed to measure test taker's ability in language regardless of any training they may have had in that language. In contrast to achievement tests, content of proficiency tests are not based on the syllabus or instructional objectives of language courses. Rather, those are based on a specification of what candidates or test takers have to be able to do in the language in order to be considered proficient.

Proficiency tests are kinds of tests designed to measure people's ability in a language, regardless of any training they may have had in that language. The content of a proficiency test, therefore, is not based on the content or objectives of language courses that people taking the test may have followed. Rather, it is based on a specification of what candidates have to be able to do in the language in order to be considered proficient. Proficiency tests are often used for placement or selection, and their relative merit lies in their ability to spread students out according to ability on a proficiency range within the desired area of learning.

2.2.2.2 Achievement Test

As its name reflected, the purpose of achievement test is to establish how successful individual students, groups of students, or the courses themselves have been in achieving objectives. Brown (2001) stated that an achievement test is related directly to classroom lessons, units, or even a total curriculum.

Mehres and Lehmann stated on the book A guide to language testing: *Development, Evaluation and Research,* achievement test may be used for program evaluation as well as for certification of learned competence. It follows that such tests normally come after a program of instruction and that the components or items of the tests are drawn from the content of instruction directly Henning (2001). According to Bill R. Gearheart (1974), the achievement test attempts to measure the extent to which pupil has achieved in various subject area.15 The measurement based on those opinions is usually done at the end of learning process or program.

Thus it can be inferred that achievement tests are used to measure the extent of learning in a prescribed content domain, often in accordance with explicitly stated objectives of a learning program. Achievement tests are also used by teacher to motivate students to study. If students know they are going to face a quiz at the end of the week, or an end of semester achievement test, the effect is often an increase in study time near the time of the test. The primary goal of the achievement tests is to measure past learning, that is, the accumulated knowledge and skills of an individual in a particular field.

According to Hughes (2013), there are two kinds of Achievement test:

1) Summative Tests (Final achievement tests)

Summative assessments, in contrast, are efforts to use information about students or programs after a set of instructional segments has occurred. Their purpose is to summarize how well a particular student, group of students, or teacher performed on a set of learning standards or objectives. Information obtained from summative assessments is used by teachers to determine grades and to explain reports sent to students and their parents stated in (Arends (2012). Moreover, Cotton (2004) states that, summative test assessment methods are made to determine what a student has accomplished at the beginning or the end of language course, the teacher gives a final mark to the students.

According to Suwandi (2009) there are four types of summative test that be used in classroom include: a) Performance task; b) Written product; c) Oral product; d) Test. In summative testing Fusher (2012), it is expected that test scores to carry generalizable meaning; that is, the score can be interpreted to mean something beyond the context in which the learner is tested.

It is concluded that, summative test is administered at the end of a course of study. They may be written and administered by ministries of education, official examining boards, or by member of teaching institutions. This test is designed to know how successful students have mastered the previous materials of a long period of course.

2) Formative Test (Progress achievement tests)

This is a way of measuring progress would be repeatedly to administer final achievement tests, they are hope to increase scores indicating the progress made. Airasian (2102) stated that formative tests take place while interacting with students and focused on making quick and specific decisions about what to do next in order to help students learn. They all rely on information collected through either structured formal activities or informal observations made during the process of instruction. Formative tests are typically designed to measure the extent to which students have mastered the learning outcomes of a rather limited segment of instruction, such as a unit or a textbook chapter. These tests are similar to the quizzes and unit tests that teachers have traditionally used, but they place greater emphasis on (1) measuring all of the intended outcomes of the unit of instruction, and (2) using the results to improve learning (rather than to assign grades) based on Huges (2013) theory.

The result of formative test gives the information about how well students have mastered a particular material. The purpose is to identify the students' learning successes and failures so that adjustments in instruction and learning can be made. The formative test also determines whether a student has not been mastered the learning tasks being taught, it can be prescribed how to remedy the learning failures. Formative test is intended to monitor learning progress during the instruction and to provide continuous feedback to both pupil and teacher concerning learning successes and failures.

2.2.2.3 Diagnostic Test

Brown (1996) stated that a diagnostic test is designed to determine the degree to which the specific instructional objectives of the course have been accomplished. Heaton (1988).also stated that diagnostic test is widely used; few tests are constructed solely as diagnostic tests. Note that diagnostic testing is frequently carried out of groups of students rather for individuals. In summary, diagnostic tests are designed to diagnose a particular aspect of a language and can be used to check the students in learning a particular element of the course. For example: it can be used at the end of a chapter in the course book or after finished one particular on lesson.

2.2.2.4 Placement Test

The placement test provides an invaluable aid for placing each student at the most beneficial position in the instructional sequence Gronlund (1977). The purpose of placement test according to Brown (2001) is to place a student into an appropriate level or section of a language curriculum or school. A placement test typically includes a sampling of material to be covered in the curriculum (that is, it has content validity), and it thereby provides an indication of the point at which the student will find a level or class to be neither too easy nor too difficult, but appropriately challenging. In summary, placement tests are intended to provide information that will help to place students at the stage or in the part of the teaching learning program that most appropriate with their abilities. So the performance students will be seen after they have been mastered.

2.2.3 Categories of a Good Test

Test as an instrument of obtaining information should have a good quality. The quality of a test will influence the result of the test itself. Once the test has a good quality, the right information will be gained and used to make accurate decision to the student's achievement. Brown (2001) stated that a well-constructed test should have five main characteristics which involve validity, reliability, practicality, authenticity and washback.

Validity is the degree to which the test actually measures what is intended to measure. Reliability is consistent and dependable. A practicality is means of financial limitations, time constraints, ease of administration, and scoring and interpretation. Then, authenticity is defined as a concept that is a little slippery to define, especially within the art and the science of evaluating and designing tests. Meanwhile, washback is the effect of testing in teaching and learning.

Moreover, a good test should be good at its item analysis. Brown (2004) also stated that, "there are three main components of item analysis, they are: difficulty level, discriminating power and the effectiveness of the distractor". Meanwhile, according to Purwanto (2009), a good test item should have three criteria; moderate difficulty level, high discriminating power and distractor analysis which work effectively. An effective and good test should have the items that belong to moderate level. The item that is too easy or difficult potentially weaken the quality of the test and the valid data of information about students's achievement will not be acquired.

2.2.4 Item Analysis

2.2.4.1 Difficulty Index.

The difficulty index is the percentage or proportion of students who answered the item correctly. Ryan and Ory (1993) said that the Higher the percentage of the students who answer correctly, the easier the item is. To obtain specific data, the writer differs students into two groups; upper and lower group. The writer here uses top 27% of high group and the 27% bottom of the lower group method. The 27 percentage is shown by Truman (1971) that it was small enough to clearly identify high and low students group, yet large enough to provide a sufficient number of score as a base for item statistic. The difficulty index is found by dividing students who get the correct items with the total number of students whom taking the test, after that multiply with 100% based on Arikunto (1999). The formulation of difficulty index is shown below:

$$ID = \frac{U_R + L_R}{N} \; \chi \; 100\%$$

Where:

ID: Index difficulty

U_R: The number of pupils who answered the item correctly in the upper group

L_R: The number of pupils who answered the item correctly in the lower group

N: The number of pupils who took the test.

Table 2.2.4

The classification of Index Difficulty (Arikunto, 1999)

Difficulty Level	
0.00 – 0.30	Difficult
0.30 – 0.70	Medium
0.70 – 1.00	Easy

The test is appropriate for being tested if each item in the examination passed by half of the students. Karmel (1978) stated that difficulty Index is relevant for determining whether students have learned the idea of being tested. It also plays an important role to determine the ability to discriminate the students who know the concept of materials being tested and those who do not.

2.2.4.2 Discrimination Index

Discrimination index refers to the ability of an item to distinguish among the students on the basis of how well they know the material that has been tested. It is also important to note that the difficulties of items can identify the comprehension ability among the students. As well as item difficulty, the student will also differ into two groups; upper and lower group. All of the groups will be a representative for 27% from the whole student who took the test. The higher the percentage.

The higher the percentage of the discrimination index means the ability of the items to differentiate the comprehension ability among students is high. The discrimination index is range between -1 to +1. An item with positive discrimination is the one that is more frequently answered correctly by students who scored high on the test as a whole than by students who scored low. Arikunto (1999) stated that a good test item is the one having discrimination index between 0.4 until 0.7.25 Discrimination index can be obtained by subtracting the number of students in a lower group who get the item right (L) with the number of students in the upper group who get the item right (U). After that the result will divide with the half total number of student both in lower and upper group. The formulation of discrimination index as shown below:

$$DI = \left(\frac{U_R}{N_U}\right) - \left(\frac{L_R}{N_L}\right) Or\ P_U - P_L$$

Where:

DI: Discrimination index

U_R: The number of pupils who answered the item correctly in the upper group

N_L: the number of pupils who answered the item correctly in the lower group

N_U: the number of pupil in the upper group

N_L the number of pupil in the lower group

$PU = \frac{U_R}{N_U}$: the proportion of pupil who answered the item correctly in the upper group

$PL = \frac{L_R}{N_L}$: the proportion of pupil who answered the item correctly in the lower group

Table 2.2.5

The Classification of Discrimination Index (Arikunto,1999)

DP	QUALITY	RECOMMENDATION
1.0 - 0.70	Excellent	Retain
0.70 - 0.40	Good	Possibilities of improvement
0.40 - 0.20	Mediocre	Need to check/review
0.00 - 0.20	Poor	Discard/review in depth
< -0.01 Worst Definitely Discard	Worst	Definitely discard

A test has discrimination index if more of the upper group answer the questions correctly than the lower do. By analyzing the discrimination index of each item, it will show the information that helps the teacher in identifying the flaws, giving further explanation about material, and also giving the feedback in learning materials.

2.2.4.3 Distractor Effectiveness

To diagnose the problem shown in difficulty and discrimination index, the proportion of student choosing the distracter can also be calculated to assess how the distracters are functioning. It can be compared with the proportion of students choosing the correct response. The distracter is considered as effective if it is chosen by the most of students in a lower group. If the distractor is mostly chosen by the upper group, it can be said that the distracter did not function as it should be. One of the objectives of item analysis is to know about the answer distribution to a subject in alternative answers. Through distracter efficiency, teacher may know the number of students who answered correctly, which distractor is too showy and make it easier for students not to vote, the misleading distracter and the distractor who managed to attract lower group students. Surapranata (2006) explained in his book that distractor is considered as an effective (functioning) if it is selected at least by 5% of examinees, and if it is chosen by less than 5% of examinees, it means that the distracter is categorized as an ineffective (non-functioning) distracter. This principle is based on Tarrant (2009) theory and Arikunto (2006) theory.

2.2.5 Types of test items

An item is the basic unit of language testing. Brown (1996) stated that the item is the smallest unit that produces distinctive and meaningful information on a test or rating scale. The item used in classroom tests are commonly divided into two broad categories: 1) The objective item and 2) The essay test.

2.2.5.1 Objective Item

Nouman (1977) stated that objective test items can be used to measure a variety of knowledge outcomes. The most generally useful is the multiple-choice item, but other item types also have a place. Following simple but important rules for construction can improve the quality of objective test items. In constructing an achievement test, the test maker may choose from a variety of item types. One of them is referred to as objective item. This kind of item test can be scored objectively. Furthermore, equally competent scorers can score them independently and obtain the same results.

The objective item can be classified into two types, which are selection-type test item and supply-type test item. Here, the researcher limited the study on the selection-type test item. Because the type test used in this research is selection-type test item. There are many kinds of selection-type test items. They are multiple choice items, true-false items and matching items. Then, the researcher only focus on the multiple choice items.

🞦 Multiple Choice

- Multiple choice items are made up of an item stem, which present a problem situation, and several alternatives, which provide possible solution to the problem. The options usually of, a, b, c or d. that will be counted correct, and the distractors, which are those choices that will be counted as incorrect. Nouman (1977) said that the multiple-choices item plays such an important role in the objective testing of knowledge outcomes that it will be treated first and in considerable detail

The term options refer to collectively to all the alternative choices presented to the students and includes the correct answer and the distractors. These terms are necessary for understanding how multiple-choice items function.

Designing multiple choice item test, the test maker should be consider in some ways. Based on Damien (2014) there are 18 basic rules in designing multiple choice item test. They are:

a) Design each item to measure an important learning outcome.

b) Present a single clearly formulated problem in the stem of the item.

c) Put the alternatives at the end of the question, not in the middle.

d) Put as much of the wording as possible in the stem.

e) Eliminate unnecessary wordiness.

f) Avoid negatively worded stems. "Which of the following is not........."

g) Avoid requiring personal opinion. Other item types are more suitable for this.

h) Avoid textbook wording.

i) Do not have linked or clued items.

j) All options should be homogeneous.

k) All options should be plausible.

l) Put repeated words in the stem, not in the options.

m) Punctuation should be consistent.

n) Make all options grammatically consistent with the stem of the item.

o) List options vertically.

p) Other options logically.

q) Use the option "all of the above" sparingly.

r) Use the option "none of the above" sparingly.

- Multiple choices have some advantages. Wilmar (1988) writes the advantages of multiple choices as follow:

a) The multiple choice item can be used for subject matter content in any different level of behaviour, such as ability to reason, discriminate, interpret, analyse, infer, and solve problems.

b) It has less chance for the students to guess the right answer than the true false item does because it is followed by four or five alternatives.

c) One advantage of the multiple choice items over the true-false item is that students also know what is correct rather than only know that statement is incorrect.

- The weaknesses of multiple choice items according to Brown (2004) are:

a) The technique tests only recognition knowledge

b) Guessing may have a considerable effect on test scores.

c) The technique severely restricts what can be tested.

d) It is very difficult to write successful items.

e) Washback may be harmful.

f) Cheating may be facilitated.

Essay Test

According to Airasian (2012), essay questions give students the greatest opportunity to supply and construct their own responses, making them the most useful for assessing higher-level thinking processes such as analysing, synthesizing and evaluating. The essay question is also the primary means by which teachers assess students' ability to organize, express and defend ideas. The main limitations of essays are that they are time-consuming to answer and score, and they place a premium on writing ability. On the other hand, Damien (2014) explained that essay tests are the best measure of students skills in higher - order thinking and written expression.

CHAPTER V

CONCLUSION AND SUGGESTION

5.1 Conclusion

The objective of this study is to find empirical evidence of whether or not the summative test items have good characteristic of a test in terms of difficulty level, discriminating power, and distracter effectiveness. Therefore, the formulation of the problem for this study is do the test items of English summative test used for seventh grade students of SMP Negeri 3 Ungaran has fulfill criteria based on difficulty level, discriminating power, and item distracter or not. The test consists of 40 multiple-choice items in which there are four distracters in each item. In this research, the writer uses quantitative research as a method of her study. Therefore, the writer analyzes the existing data by referring to some related theories to describe the difficulty level of the items, item discrimination index and the effectiveness of the distracters.

Based on the analysis and data interpretation from the previous chapter, the writer would like to conclude the quality of English Even Summative Test item which have been tested at the seventh grade of SMP Negeri 3 Ungaran as follows; from 40 items, there are 15 (37,5%) acceptable items to meet the criteria of index difficulty. Besides there is 1 (2.5%) too difficult item and there are 24 (60%) unacceptable items due to the easiness of the items.

For the discriminating index, the writer found there are 16 (40%) poor items, 1 (2.5%) items has worst result, which has to be discard, and 12 (30%) acceptable items and the last there are 11 (27.5%) mediocre items. Moreover for the distractor effectiveness, the writer found there are 51 (42.5%) distracters with effective distractor and 69 (57.5%) distracters with in-effective distractor.

Almost all of the items are easy in difficulty, and it also has many ineffective distracters. While, ineffective distracters make the question easier to answer and the result of an easy question is the gap between lower and upper group become smaller. In conclude, it can be said that the English summative test did not test real ability of students comprehension; and from the difficulty index, discriminating index and the distractor efficiency perspective of English summative test for seventh grade of SMP Negeri 3 Ungaran did not meet the criteria of effective and acceptable test.

5.2 Suggestions

Based on what the writer found in the data analysis and interpretation, it can be understood that the test maker of this English even summative test cannot construct ideal test items, in contrast the test maker go as she pleased. It will be better for the test maker to understand how to construct an ideal yet effective item test before conducting or even writing a test. In order to help further teaching-learning process especially in conducting a test, therefore, the writer makes suggestions as the following:

1. The test maker should use correct grammar, punctuation, capitalization, spelling, and avoid trick items. Therefore, there will be no misinterpreting among the students and teacher.

2. If the test maker uses item bank as his or her resource, it is better to recheck from the content, stems, grammatical consistencies, and the distractors. So that ambiguity, question that are look alike, and miskeyed response in the stem will not happen again in the future.

3. The test maker should arrange the items started from easiest to difficult in sequence consider as students who taking the test. Therefore, the test maker will know which part of items that the student knows very well and which one is not.

4. The test maker should do an item analysis after conducting a pilot test in order to know whether the test that she, or he made has classified as a good test or not. Moreover, because the test is for standardized test, the items which are too hard or too easy are better not used in the constructing the test.

5. It is better for the test makers to follow the guidelines in writing items of multiple-choice by the educational theory.

REFERENCES

Ahmann, J. Stanley and Marvin D. Glock. (1997). *Evaluating Pupil Growth*.Boston: Allyn and Bacon Inc.

Airasian, Peter W. and Michael K. Russel. (2008). *Classroom Assessment*. New York: Mc Graw Hill, 6th edition.

Allison, Desmond. (1999). *Language Testing and Evaluation: An Introductory Course*. Singapore: Singapore University Press.

Arifin, Zainal. (2011). *Evaluasi Pembelajaran: Prinsip, Teknik dan Prosedur*. Bandung: PT. Remaja Rosdakarya.

Arikunto, Suharsimi. (1999). *Dasar-dasar Evaluasi Pendidikan*. Edisi Revisi. Jakarta; Bumi Aksara.

Amelia, Rika. (2010). *An Analysis of the English Summative Test Items in terms of Difficulty Level for the Second Year Students; A case study of MTs. Darul Ma'arif Surakarta*, (Semarang: Semarang State University)

Andrian Dwi Prayoga, (2011). *An Analysis on Difficulty Level of English Summative Test for Junior High School at Odd Semester 2010/2011 of SMP 13 South Tangerang (A Case Study at The Second Grade of SMPN 13 Tangerang)*, Research Paper at State Islamic University Syarif Hidayatullah Jakarta. Jakarta.

Hikmah Lestari, (2011). *An Item Analysis on Discriminating Power of English Summative Test (A Case Study of second year of SMPN 87 Pondok Pinang)*, Research Paper at State Islamic University Syarif Hidayatullah Jakarta. Jakarta.

Brown, H. Douglas. (2003) *Language Assessment: Principle and Classroom Practice*. San Francisco: Longman.

Burton, Steven., et.all.,(1991) *How to Prepare Better Multiple-Choice Test Items: Guidelines for University Faculty*. New York: Brigham Young University Testing Services.

Djiwandono, M. Soenardi. (2008). *Tes Bahasa (Pegangan Bagi Pengajar Bahasa)*. Jakarta: PT. Indeks.

Haladyna, Thomas M. (2004). *Developing Validating Multiple-Choice Test Items*. New Jersey: Lawrence Erlbaum Associates Publisher.

Haladyna, Thomas, et.al. 2002.*"A Review of Multiple-Choice Item-Writing Guidelines for Classroom Assessment"*, Applied Measurement in Education, 15.

Hingorjo, Mozaffer Rahim,et.al. (2012).*"Analysis of One-Best MCQs: the Difficulty Index, Discrimination Index and Distractor Efficiency"*, Journal of Pakistan Medical Association. 62.

Hoshino, Yuko. (2013).*"Relationship between Types of Distractor and Difficulty of Multiple-choice Vocabulary Tests in Sentencial Context",* Springer; Language Testing in Asia, 13.

Karmel, Louis J. (1978). *Measurement and Evaluation in the Schools.* London: The Macmillan Company.

Kings, Kelly V, et.al., (2004). The Distractors Rationale Taxonomy: Enhancing MultipleChoice Items in Reading and Mathematics. San Antonio: Pearson Inc.

Knapp, Thomas R. (1971). *Statistics for Educational Measurement*. Scranton: The Haddon Craftsmen,Inc.

Madsen, Harold S. (1983). *Techniques in Testing.* Hongkong: Oxford University Press.

Marsh, Colin. (2010). *Becoming a Teacher: Knowledge, Skills and Issues*. Frenchs Forest: Pearson, 5th edition.

Moore, Kenneth D. (2012). *Effective Instructional Strategie*s : From Theory to Practice. Washington DC: SAGE Publication Ltd.

Noll, Victor H. (1965). *Educational Measurement*. Boston: Houghton Mifflin Company.

Nunnally, Jum C. (1974). *Educational Measurement and Evaluation*. New York; McGraw-Hill. Inc.

Nurkancana, Wayan, P.P.N Sunartana. *Evaluasi Pendidikan*. Surabaya: Usaha Nasional, 1986.

Ory, John C. and Katherine E. Ryan. (1993)*Tips for Improving Testing and Grading Surviving Skills for Scholar)*. Los Angeles; SAGE Publication. Inc.

Sukardi, H.M. (2009). *Evaluasi Pendidikan: Prinsip & Operasionalnya*. Jakarta: Bumi Aksara.

Surapranata, Sumarna. (2006). *Analisis, Validitas, Realibilitas, dan Interpretasi Hasil Tes*. Bandung: Remaja Rosdakarya Offset.

Sudaryono. (2012). *Dasar-dasar Evaluasi Pembelajaran*. Tangerang: Graha Ilmu.

Tarrant, Marrie et.al. (2009).*"An Assessment of Functioning and Non-functioning Distractors in Multiple-Choice Questions"*, BMC Medical Education, 9.

Wick, John W. and Donald L. Beggy, (1971). *Evaluation for Decision-Making in the Schools*. Michigan: Houghton Mifflin.

Youngblood, Joseph. (1981).*"Improving Teaching and Testing through Item Analysis".* Jstor; College Music Symposium , 21.

Zimaro, Dawn M. (2004). *Writing Good Multiple Choice Exams*. Austin: University of Texas.