

## **AN ITEM ANALYSIS OF ENGLISH SUMMATIVE TEST FOR THE TENTH GRADE STUDENTS OF SMA MUHAMMADIYAH 3 JAKARTA IN THE 2013/2014 ACADEMIC YEAR**

**Anisa Fitriani Lailamsyah**

Universitas Bina Sarana Informatika, Jakarta

**Fitri Apriyanti**

Universitas Bina Sarana Informatika, Jakarta

fitri.fap@bsi.ac.id

### **Abstract**

This research is intended to analysis of the English summative test for the tenth grade students of SMA Muhammadiyah 3 Jakarta in the 2013/2014 academic year. The research method used to analyze the item of English summative test for senior high school are quantitative and qualitative method. Quantitative method was used to analyze data of the facility value and discriminating power of each item. The qualitative method was used to describe and analyze the quality of test items.

**Keywords:** Item Analysis, English Summative Test, Tenth Grade Student

### **INTRODUCTION**

Evaluation is an important part in teaching and learning process. It gives big contribution for the teaching and it provides an information about the students' progress. Evaluation result can be used as basis of improving a learning teaching process and of making a decision about the students. Gronlund and Robert point out, "Evaluation is a systematic process of collecting, analyzing and interpreting information to determine the extent to which pupils are achieving instructional objectives." Hence, evaluation is the systematic process to know how far the students can achieve the instructional objectives after the teaching and learning process.

Furthermore, collecting data is the first step of evaluation. In order to collect the data, a teacher needs an instrument. One of the evaluation instruments that are commonly used in educational world is test. Brown explains, "Test is a method of measuring a person's ability or knowledge in a given domain." It means a test is designed to quantify the achievement and to measure student's progress.

The types of test which are used by teachers are formative and summative tests. Formative test is used to know students' achievement at the specific of study or the

end of the chapter. Summative test is used to measure students' achievement at the final of study.

In order to get accurate data, a test should have good criteria. There are validity, reliability, and practicality. A test can be valid if the test can measure what is supposed to measure. It can be reliable if the result of the test is the same even though the test is administered to the same level students in the next time. It can be practical if it is easy to do and to administer. Beside the three criteria above, the quality of a test is also influenced by the quality of individual items. In order to see such item quality, teacher has to do item analysis.

Analysis item is the process to identify the quality of each item and to find which one of items that has good quality, needs to be revised even needs to be removed. There is the procedures of analysis an item, the following information; Difficulty level of the item and Discriminating power of the item. After doing the procedures, teacher will get information about the success of the test.

It has been stated above that ideally a teacher must be sure that the test has fulfilled the criteria of a good test that includes the quality of the test items. However, analysis

test item has not become an academic habit in SMA Muhammadiyah 3 Jakarta.

SMA Muhammadiyah 3 Jakarta always conducts evaluation by using summative tests at the end of each semester. However, the items of test have never been analyzed whether it has the criteria of a good test or not. So, the teacher has not known the quality of the items.

Considering this fact, the writers choose the tenth grade students of SMA Muhammadiyah 3 Jakarta 2013/2014 academic year as a place to do this research. The writer will conduct the study entitled "An Item Analysis of the English Summative Test for the Tenth Grade Students of SMA Muhammadiyah 3 Jakarta in the 2013/2014 Academic Year."

In the teaching and learning process evaluation is very important for the teacher and the students. It is the process to make decision concerning the students' achievement in an instructional progress.

According to Purwanto, "Educational evaluation is the estimation of the growth and progress of the pupils towards objectives or value in the curriculum." The key words are *estimation of the growth* and *progress*. *Estimation of the growth* means knowing how far the students master in achieving instructional and *progress* means the development of students in achieving instructional. It can be summarized that evaluation is used to measure the extent to which students have mastered the instructional objective.

On the other hand, Gronlund and Robert says that, "Evaluation is the systematic process of collecting, analyzing, and interpreting information to determine the extent to which pupils are achieving instructional objectives." The key words are *systematic process*, *determining the extent*, and *achieving instructional objectives*. It means that evaluation is a systematic process to define how far instructional objectives are achieved.

Another definition comes from Bloom in Daryanto's book. He says, "Evaluation, as we see it is the systematic collection of evidence to determine whether in fact certain changes are taking place in learners as well as to determine the amount or degree of

change in individual students." It means by doing evaluation a teacher can identify how far the change of individual students relates to the learning process. Through evaluation, a teacher is able to get information whether the teaching and learning process that has been carried out is successful or not.

Based on the explanation above, the writer concludes that evaluation is systematic process through collecting and analyzing information or data to make valid decision of students' progress in achieving the instructional objective.

In general, educational evaluation has several purposes. The evaluation is conducted by teacher to their students. According to Murwani it consists of four objectives, they are : (1) to give feedback to the teachers for improving the quality of teaching and learning process and as a basic to improve assessment program; (2) to know students learning achievement; (3) to place the students in the appropriate learning situation, based on their ability and characteristics; and (4) to know students psychological background and learning difficulties. It can be summarized that the objectives of evaluation in education are related to the effort for enhancing the quality of teaching and learning process.

From the explanation above, it can be concluded that the result of evaluation can be used as a base to improve the quality of teaching and learning process. The improvement of teaching and learning quality can be obtained if the teacher knows the student's strength and weakness.

Test is one of instruments which are commonly used to do an evaluation. According to Reynolds et al, "Test is a device or procedure in which a sample of an individual's behaviour is obtained, evaluated, and scored using standarized procedures." The key words are *device*, *score* and *standardized procedures*. *Device* means an instrument of set of items which are used, *score* means test which has given it will produce a score as a result, and standardized situation means the same difficulty level of the items. It means that with test teacher will judge students' progress from the score as a result in the test

which is given to each student with the same of difficulty level.

Hopkins and Antes state, “A test is an instrument, device, or procedure that purposes a sequence of tasks to which a student is to respond- the results of which are used as measures of a specified trait.” The key words are *instrument, device* and *procedure*. The test in this definition is used in educational setting to measure students’ specified trait by using an instrument, device or procedure that usually consists of a set of questions or tasks to be responded by the students.

Moreover, Bachman states “A Test is a particular type of measurement that focuses on eliciting a specific sample of performance.” The key words are measurement, eliciting a specific sample of performance. Measurement means to know the extent of students’ achievement, eliciting a specific sample of performance means gather the students’ result. From the key words above the writer summarizes that test is used to assess the students’ extent in achieving instructional objectives.

Based on the definitions above, the writers conclude that test is set of items which are given for students as assessment of students’ progress in the learning and will make score as a result of the response.

There are many tests that teachers carry out in the classroom. According to Harmer, based on the function, the test is divided into four types. First of all is placement test. Placing new students in the right class in a school is facilitated with the use of placement test. It means that, the placement test can evaluate students’ skill level and placing the students in the right class.

Secondly, it is diagnostic test. It can be used to expose students difficulties, gaps in their knowledge and skill deficiencies during a course. Thus, diagnostic test is intended to diagnose learning difficulties and gap in students’ knowledge in the subject during teaching and learning activity.

Third is proficiency test. It gives a general picture of a learner’s knowledge and ability, they are frequently used as stages people have to reach if they want to be admitted to a foreign university, get job, or

obtain some kind of certificate. Thus, the proficiency test can help students to apply to universities abroad, apply for a job or get other important certificates.

The last one is achievement test. Achievement test is designed to measure learners’ skill progress in relation to the syllabus they have been following. In other words, this test can be used to measure skill progress of students that they have mastered a subject matter based on the syllabus.

There are two types of test which belong to achievement test. They are formative test and summative test. A teacher should know how to make a good test. Test which is organized should have some good characteristics. According to Arikunto in Sudjiono’s book, “*Sebuah tes yang dapat dikatakan baik sebagai alat pengukur harus memenuhi persyaratan tes, yaitu memiliki: 1) validitas, 2) reliabilitas, 3) praktikalitas*”. A test is valid when it measures what is intended to measure. It can be reliable when a test can generate a consistent and stable result and it is practical if it is easy in administration.

Based on explanation above, the writer summarizes that there are three criteria of a good test; validity, reliability and practicality. So, before the teachers make a test, they have to consider three criteria of good test in order to get accurate information from the test which is given.

Item analysis is an important point to know the quality of each item. According to Sumarsono, “*Analisis butir soal adalah suatu proses yang dilaksanakan sesuai prosedur dan langkah tertentu dengan tujuan untuk mengidentifikasi butir-butir soal mana saja yang efektif atau berkualitas sehingga dapat dipakai untuk pengesanan dan butir-butir soal mana saja yang lemah sehingga perlu ditinjau dan direvisi.*”

The key words are *process, based on procedure, identifying, test items, effective, and weak*. It can be concluded that item analysis is the systematic procedure to identify whether the test items are effective or not to be administered.

Another definition comes from Brown and Hudson. They state that “Item analysis is usually done for purposes of selecting which items will remain on future revised

and improved version of the test.” The keywords are *selecting, revised, and improved version*. From these keywords above, it means from doing item analysis teacher will choose the item which must be revised or changed to be a good one.

From the definitions above, the writer summarizes that item analysis is conducted to know the quality of items which are tested and choose the item which must be revised or changed to be better for future test.

According to Hopkins and Antes, there are some procedures of item analysis that should be done by the teacher, as follow: (1) Ordering the scored papers from high to low by total score; (2) Choosing the number of papers for criterion groups; (3) Counting the number of responses marked for each alternative; (4) Recording the data from the student responses; (5) Calculating which include: a difficulty index for each item and a discrimination index for each item; (6) Calculating a discrimination index for distracter alternatives for multiple-choice items. The writers will use the procedures of item analysis stated by Hopkins and Antes above because the procedures are easier to be applied.

The first indicator to analyze test item is index difficulty level or facility value. This indicator is to know the difficulty level of items which are tested. According to Sumarsono “*Suatu set soal yang baik adalah set soal yang komposisi jumlah butir soal mudah, sedang dan sukar tersusun secara memadai.*” The key words are *a set of test, number of easy, medium, and difficult test items*. It can be inferred that the test items should be balance between the easy and difficult items. To count the facility value of items, the writer adopts the following formula from Heaton:

$$FV = \frac{\text{Correct U} + \text{Correct L}}{2n}$$

The discriminating power is used to know the items which are in different levels between upper and lower group. Purwanto states that

“*Daya pembeda suatu soal ialah bagaimana kemampuan soal untuk membedakan siswa-siswi yang termasuk kelompok pandai (upper group) dengan siswa-siswi yang termasuk kelompok kurang (lower group)*”.

From the statement above, the purpose of discriminating power is to find out the ability of items to differentiate the upper from lower group. If the upper group answer the items correctly and the lower group answer the items incorrectly, it means that the items can be categorized as good ones because it can distinguish the upper from the lower group. If the items are answered correctly both by upper and lower groups, it means that the items cannot be categorized as the good ones because it cannot distinguish between the upper and lower students. To measure the item discrimination, the writer adopts the following formula from Heaton:

$$DP = \frac{\text{Correct U} - \text{Correct L}}{n}$$

There are some procedures for analyzing the weak items according to Sumarsono: (1) Copy the item; (2) Identifying the competency to measure (follow the competency/ indicator determined in the KTSP); (3) Showing the item data; (4) The item details: The form of stem: whether it is statement or interrogative, etc. The length of stem: how many words there are. Numbers of option: how many options there are. The form of option. The length of options; (5) Analyzing the weakness of item.

## METHODOLOGY

The objectives of this research are to find out the good items, to store the items in item storage, to analyze the weak items and to revise them. The instruments which are used to get the data are the students’ answer sheets of English summative test for the tenth grade of the second semester in SMA Muhammadiyah 3 Jakarta.

## FINDINGS

The English summative test of SMA Muhammadiyah 3 Jakarta consists of 4 section : (1) Reading section, there are 35 items. The numbers are 1, 2, 3, 4, 5, 11, 12, 13, 14, 15, 21, 22, 23, 25, 26, 27, 28, 30, 31, 32, 33, 35, 36, 37, 38, 39, 40, 41, 42, 43, 45, 46, 48, 49 and 50; (2) Vocabulary section, there are 7 items. The numbers are 8, 16, 20,

24, 34, 44 and 47; (3) Structure section, there are 5 items. The numbers are 10, 17, 18, 19 and 29; (4) Writing section, there are 3 items. The numbers are 6, 7 and 9.

The writers try to get the result of students' score from SMA Muhammadiyah 3 Jakarta in aiming to analyze the English summative test. All the result of the students' scores are described below:

**Table 1. The list of upper group student name's based on the highest score until the lower**

No	Students' Name	Score
1	DD	88
2	ER	86
3	FH	84
4	DS	78
5	A	76
6	JP	74
7	MF	74
8	BT	72
9	DN	72
10	MF	72
11	PS	72

**Table 2. The list of lower group student name's based on the highest score until the lower**

No	Students' Name	Score
1	YA	58
2	MH	56
3	UW	54
4	NP	52
5	RR	52
6	AS	46
7	AZ	42
8	IZ	42
9	RB	40
10	YA	24
11	MH	18

**Table 3. The Category of the Facility Value of English Summative Test**

Range of	Category	Frequency	%
0,00 – 0,30	Difficult	4	8
0,31 – 0,70	Medium	26	52
0,71 – 1,00	Easy	20	40
Total		50	100

## REFERENCES

Cecil R. Reynolds et al. 2010. *Measurement and Assesment in Education*. New Jersey: Pearson Education.

**Table 4. The Category of Item Analysis of Discriminating Power**

Range of	Category	Frequency	%
Under 0,00	Discarded	2	4
0,00 – 0,19	Revised	15	30
0,20 – 0,29	Marginal	9	18
0,30 – 0,39	Good	5	10
0,40 - above	Very good	19	38
Total		50	100

After analyzing the English summative test items for the tenth grade students of SMA Muhammadiyah 3 Jakarta, it is discovered: based on the category of difficulty level, there are 4 items (8%) are categorized difficult, 26 items (52%) are categorized medium and 20 items (40%) are categorized as easy items. Moreover, based on the category of discriminating power, there are 19 items (38%) are categorized very good items, 5 items (10%) are categorized good, 9 items (18%) are categorized marginal and 15 items (30%) are categorized poor or need to be revised. There are also 2 items (4%) that have to be discarded because the discriminating power is negative.

There are some factors that make the item becomes poor. For instance, the item is too difficult or easy, the distractor is not attractive and the material has not been taught before. In addition, there are some mistakes in typing words in the items. Although it is not the main cause of the poor items, it can make students confused.

## CONCLUSION

There are some factors that make the item becomes poor. For instance, the item is too difficult or easy, the distractor is not attractive and the material has not been taught before. In addition, there are some mistakes in typing words in the items. Although it is not the main cause of the poor items, it can make students confused.

Drs. M. Ngalim Purwanto. 1991. *Prinsip-Prinsip dan Teknik Evaluasi*. Bandung: PT. Remaja Rosdakarya.

- H. Douglas Brown. 2001. *Teaching by Principles an Interactive Approach to Language Pedadogy, Second Edition*. New York: Addison Wesley Longman.
- H.M. Daryanto. 2001. *Evaluasi Pendidikan*. Jakarta: Bumi Aksara
- James Dean Brown and Thom Hudson. 2002. *Language Testing*. Cambridge: Cambridge University Press.
- Jeremy Harmer. 2001. *The Practice of Language Teaching*. England: Longman.
- Lyle F. Bachman. 2004. *Statistical Analyses for Language Assessment*. New York: Cambridge University Press.
- Ngalim Purwanto. 2007. *Prinsip dan Teknik Evaluasi Pendidikan*. Bandung: Rosakarya.
- Prof. Drs. Anas Sudijono, 2006. *Pengantar Evaluasi Pendidikan*. Jakarta: PT. Raja Grafindo Persada.
- Sigit Sumarsono. 2012. *Metode Riset Pendidikan Bahasa*. Jakarta: UHAMKA Press.