

## An Item Analysis of English Summative Test for The First Semester of The Third Grade Junior High School Students in Surakarta

Septi Ika Pradanti, Martono, Teguh Sarosa

**English Education Department  
Teacher Training and Education Faculty  
Sebelas Maret University of Surakarta**

**Abstract:** The aims of this research are:(1) to find out whether the multiple-choice items of English summative test for the third grade junior high school students in Surakarta have fulfilled the criteria of good test or not; and (2)To describe whether the multiple-choice items of English summative test for the third grade junior high school students in Surakarta have fulfilled the criteria of good test item viewed from difficulty level, discrimination power, and distractor evaluation or not. The data were taken from 100 students' answer sheets in five schools in Surakarta. The test item was analyzed by using item analysis technique seen from the score and the language content analysis. The analysis was considered by using three aspects i.e. index of discriminating power, level of difficulty and the distractor evaluation. The research finding shows that (1) the difficulty level shows 13% for very difficult item, 22% for very easy item, and 64% for satisfactory item; (2) only 35% of all test items have appropriate index of discriminating power; and (3) only 49% of the distractors are effective since they were selected by the students. Based on the three criteria of a good test item above, there are only 46% of multiple choice items fulfilling the criteria of a good test items.

**Key words:** Item Analysis, Summative Test, Multiple Choice Item

### INTRODUCTION

Testing is an important part of every teaching and learning. It contributes directly to the teaching learning process used in classroom. It is also useful in programmed instruction, curriculum development, etc. Madsen (1983:3) states well made tests of English can help students in at least two ways. First of all, such can help create positive attitude toward the class. English tests can also give the benefit to the students by helping them master the language. It means that a well English test can give many effects to the students in learning English. Linn and Gronlund (2000:32) define test as an instrument or systematic procedure for measuring a

sample of behavior by posing a set of questions in a uniform manner. Based on that, test is also included as a form of test which can measure how well the students' achieve their second language learning.

“Good evaluation of our test can help us measure students' skills more accurately. It also shows that we are concerned about those we teach (Madsen, 1983:178).”

So, from that statement the teacher must do the evaluation after the test. It is also emphasized by a following quotation from NBPTS (National Board of Professional Teaching Standards):

“Teachers are responsible for managing and monitoring students”

learning. The teachers intend to recognize, know how to create, enrich, maintain, and alter instructional settings to capture and

From the statement by Gronlund, the teacher does the evaluation to establish how successful the teaching process can improve the students' achievement in the instructional objective. As the result, we know the quality of the teacher's instruction and the students' ability. The teacher can use this information to make an improvement for further learning. Therefore, the teacher can increase the quality of learning process and learning outcome as a part of the increasing the quality of education system. Harmer (1998: 22) argues that a good teacher keeps a record of their students' achievements so that they are always aware of how they are getting on. By doing this, they can begin to see when teaching and learning has or has not been successful.

Some students may feel anxious when hearing the word „test“. On the other hand, it can motivate them to improve their learning and help the teacher to evaluate the instructions. Heaton (1990: 6) states that a good classroom test will help the teacher to locate the precise areas of difficulty encountered by the class or by the individual student. It means that by using a good test the teacher is able to recognize the students' weaknesses and difficulties, so they are able to identify and analyze the students' error in mastering English as the target language. In other words, a test will be used to measure the students' performance.

sustain student interest. They use many methods to measure student growth and understanding (Linn and Gronlund, 2000: 35).”

A test which will be appropriate for this evaluation is summative test. Based on Linn and Gronlund (2000:40) summative test is used to assess achievement at the end of instruction. It can be said that the test is designed to determine the extent to which the instructional goals have been achieved and is used primarily for assigning course grades or for certifying student mastery of the intended learning outcomes. Preparation in facing national examination also becomes one reason for the writer to study about this. In the first semester for third grade junior high school students, teacher and student can measure the student's achievement by its result. If the result shows high score or higher than the standard score, the students are ready enough in facing national examination. That is one reason the writer chooses to analyze English summative test for the first semester of the third grade junior high school students. A well of an item analysis is viewed from difficulty level, discrimination level, and distractor evaluation.

For the type of summative test, there are so many types of test items applied in measuring the students' performance. Nevertheless, students seem to be familiar to objective test items e.g., true-false, completion, matching, short answer, and multiple-choice items. The multiple choice item is generally recognized as the most widely applicable and useful type of objective test item. It can more effectively

measure many of the simple learning outcomes measured by the short answer item, the alternative response item, and the matching exercise. In addition, it can measure a variety of the more complex outcomes in the knowledge, understanding and application areas. This flexibility, plus the higher quality of the items usually attained with the multiple choice form, has led to its extensive use in achievement testing. The purpose of writing this article is for reporting about the research has been done. It is covering the problem, method and the result in this study. It can also help the other researchers find new knowledge in doing/studying the research in line material.

## RESEARCH METHODS

The study was conducted in five junior high schools in Surakarta in December 2014. The writer collected 100 students' answer sheets of the third grade junior high school students. It means 20 students' answer sheets as the sample of its each junior high school.

The document that was used in this study was the files or data of related information including the English summative test items and the students' answer sheets. In collecting data, the writer did an observation by visiting the school to ask for the students' answer sheet and test paper of the English summative test for the first semester of the third grade junior high school students in Surakarta to be analyzed.

## RESEARCH FINDINGS AND DISCUSSION

This study was conducted in five junior high schools in Surakarta. It is designed primarily to serve the writer's purpose to describe the appropriateness of the multiple-choice items of English Summative test for junior high school students in Surakarta. Based on the purpose of the study, the writer used descriptive method with quantitative approach in conducting the study.

The findings of the data analysis of English summative test item is aimed to (1) describe the level of difficulty, discriminating power, and the effectiveness of distractors of the multiple-choice items of English summative test for the first semester of third grade junior high school students in Surakarta in 2014/2015 academic year and (2) identify whether the multiple-choice items have fulfilled the criteria of a good test item viewed from level of difficulty, discriminating power, and the effectiveness of distractors.

The findings will be described and discussed in two sub-chapters. The first is item analysis on multiple-choice item by applying item analysis technique. The second one is the research finding presenting the global finding of the analysis. The writer classified the students into three groups, i.e. upper, middle, and lower group. The classification group was determined by selecting the top 27% and the bottom 27% out of 100 students' score which have high and low score in the test. In this case, the total of test score refers to the total number of students taking the test. The writer arranged the students'

score from the upper to the lower. The writer took 100 students as the sample of this research (20 students of each school), so that the upper group consists of 27 students, the middle group consists of 56 students and the lower group consists of 27 students. The upper and lower groups were used in the computation, while the middle was discarded. In analyzing the item, the writer also made a table, showing a list of the students' answer and the spread of selected distractors. This summative test will be analyzed by the score and by the content of language.

This summative test has 50 questions. It consists of 45 multiple choices item and 5 essay items with 29 kinds of materials and 120 minutes for the time limit of the test. The writer will focus on analyzing the multiple choice items of this summative test because the score is gotten from the multiple choices item. It is caused of the answers of essay items are not checked by the teacher. The teacher gives the same score for all of the answers of essay items. The question of summative test is also matched with the blue print. The blue print is getting from MGMP of all English teachers in Surakarta municipality.

The study of summative test in third grade junior high school students is also as measure the students' preparation to face their National Examination (UN). So, by this study the teacher and students can be aware of the question in National Examination in their second semester.

Sudjana (1989:158-159) describes about the advantages the data of summative result. Summative test is

held in the end of semester. The material takes from the RPP and syllabus which is appropriate with the curriculum. The questions usually are made in the objective form. The score of summative test is considered as final result because there is no chance to change or make up the score at that semester. It can be improved in the next semester. Because of that, the data of summative result can be used by the teacher to: (1) Making a students' improvement report after considering the score from formative test and the other students' improvement. (2) Rearranging all of the material subject and sub-material subject after monitoring the score of summative test especially for the material which is the students have not master it. (3) Repairing and renewing the summative test item based on the students' score. (4) Arranging the students' learning program for the next semester based on the student' score on summative test and considering the material which should student master before continuing the material on the next semester.

Sudijono (2005:23) states, *evaluasi sumatif adalah evaluasi yang dilaksanakan setelah sekumpulan program selesai diberikan (berakhir) atau dengan kata lain evaluasi yang dilaksanakan setelah seluruh unit pelajaran selesai diajarkan*. In other words, we can say that the summative test is usually taken in the end of semester after all the material has been done. For this study summative test for junior high school students is UAS. This theory is also the reason for the writer to select the descriptive method with quantitative approach as the study method.

Based on the research and finding that was done by the researcher, there were 29 kinds of material in the summative test for third grade of junior high school students in Surakarta. The material were covering transactional/interpersonal dialogue, short functional text, procedure text, report text, and also arranging the words into sentence. The material in the summative test consists of 6 transactional/interpersonal dialogues, 7 short functional texts, 7 report texts, 4 procedure texts, and 1 arranging the words.

In this summative test, there are some based competences and standard competences which appear as the based information in doing the test. They are reading in writing skills. In the multiple choice item, two of those skills appear in the form of filling the cloze text and also arranging the word/sentence become good sentence/paragraph.

Based on Standard Kompetensi Lulusan in the National Examination of SMP and Mts, the competency of those skill are (1) Reading (*membaca*) *memahami makna dalam wacana tertulis maupun pendek baik teks fungsional maupun esai sederhana berbentuk deskriptif* (descriptive, procedure and report) *dan naratif* (narrative and recount) *dalam konteks kehidupan sehari-hari*; (2) Writing (*menulis*) *mengungkapkan makna secara tertulis teks fungsional pendek dan esai sederhana berbentuk deskriptif* (descriptive, procedure and report) *naratif* (narrative and recount) *dalam konteks kehidupan sehari-hari*. The explanation of reading based on its is *memahami makna test tulis fungsional*

*pendek seperti pesan pendek, pengumuman, kartu ucapan, label, iklan, brosur, surat pribadi, dan esai sederhana berbentuk deskriptif* (descriptive, procedure and report) *dan naratif* (narrative and recount) *dalam konteks kehidupan sehari-hari dengan cara mengidentifikasi gagasan utama (judul, tema, pokok pikiran), informasi rinci tersurat, informasi tersirat, rujukan kata, makna kata, frasa dan kalimat (termasuk kosa kata tata bahasa dan ciri kebahasaan lainnya yang terkait dengan jenis teks dan tema)*. Whereas, the explanation for the writing skill is *mengungkapkan makna dalam bentuk teks tulis fungsional pendek seperti pesan pendek, pengumuman, kartu ucapan, label, iklan, brosur, surat pribadi, dan lain-lain, dan esai sederhana berbentuk deskriptif* (descriptive, procedure and report) *dan naratif* (narrative and recount) *dalam konteks kehidupan sehari-hari dengan cara menyusun kata/kalimat acak menjadi kalimat/ paragraf yang padu dan melengkapi kalimat/paragraf/ teks rumpang*.

Based on the analysis, the researcher finds that there are 22.22% of the test items which can be answered easily by the students. it means that there are 10 easy items. While, there are 29 satisfactory items on this summative test (64.44%). Those difficult items and easy items should be revised before they are used again. The last, this the table shows that there 6 difficult items which percentage value 13.33%.

Based on the calculation of discriminating power index, it can be concluded that there are some items which need to be revised. There are 6

items (13.33%) which is categorized as very poor in level of discriminating power, 10 items (22.22%) as poor in discriminating power, 16 items (35.55%) as satisfactory item in discriminating power and the last is 13 items (28.88%) as good in discriminating power. Here, satisfactory item still dominate the value of discriminating power. Next, it is followed by good value of discriminating power as (35.55%). Last, for the effectiveness of distractors was found that this test has 89 distractors which work functioning as intended. We call it functioning distractor because the distractor can attract more students from lower group than upper group. Next, there are 20 distractors which are functioning poorly. It is because the students from lower and upper group choose the distractor almost in the same equal or only different for 1 amount. Not all the distractors work as intended. There are some reasons why this happens.

Finally, item analysis helps teacher to improve multiple-choice items since this type of test items are objective, easily quantified, and calculated using a certain formula. Moreover, the analysis shows the strengths and the weaknesses of the test items. From the analysis, the teacher can find students' achievement and students' difficulty in mastering a certain subject or topic. Besides, teacher can also consider which items need to be revised, discarded, or saved for the next tests. By using item analysis, the teacher can also revise test items, which are constructed by the teacher.

## **CONCLUSIONS AND SUGGESTIONS**

The researcher makes some conclusions based on the data analysis in the previous chapter. The conclusions are as follows (1) related to the level of difficulty, discrimination index, and the effectiveness of distractors, it can be concluded that (a) The test items have not had proportional level of difficulty, since the spread of the difficulty level is drawn toward the normal curve. The proportion of the difficulty level shows the percentage of 13.33 % for very difficult item, 64.44 % for satisfactory item, and 22.22 % very easy item. (b) Totally, there are 29 out of 45 items have appropriate index of discriminating power. Then, there are still 16 items which have inappropriate index of discriminating power (6 items should be omitted and 10 items are needed to revise). (c) It is clear that most of the distractors are effective since there are 66% of all distractors which are selected by the students. (2) Based on the result of the item analysis, not all English summative test items fulfill the criteria of a good test item. Only 46% of the total items fulfill all criteria. It means that 21 items have a good index of discriminating power and level of difficulty with effective distractors. Then, there are 35% or around 16 items, which still have some weaknesses. These items have fulfilled only one or two of the criteria. While the rest of the items i.e., 8 items or around 17% of all items do not fulfill any criteria of a good test item.

Then as the suggestion in constructing test item, the teacher should consider the three aspects of good test

item. Firstly, the teacher should consider the students' ability to get proportional level of difficulty. The item should examine the material which have been learnt by the students. Second, the test items should be able to discriminate the upper and the lower students. A confusing stem or question can be difficult for the students in answering the question and influence the value of discrimination index. It means that in constructing the question the teacher should use clear instruction and language structure. The teacher should avoid any discrimination index that values negative.

Last, the teacher should also be careful in constructing distractors, since the distractors have main role in multiple-choice item. Confusing distractors or ineffective distractors will provide a low discriminating power and make the items become too easy or too difficult. In constructing the distractors, the teacher should use clear language structure and vocabularies, which are familiar to the students. The last, before taking up using the test items, the test-publisher should have some research or organize a tryout. The test items which are constructed should be based on the syllabus, so that the test items are able to examine the students' ability in desired direction.

## BIBLIOGRAPHY

- Cangelosi, James S. 1995. *Merancang Tes Untuk Menilai Prestasi Siswa (Terjemahan)*. Bandung: Penerbit ITB.
- Harmer, J. 1998. *How to Teach English: An Introduction to the Practice of English Language Teaching*. New York: Longman.
- Heaton, J. B. 1990. *Writing English Language Tests*. New York: Longman
- Linn, Robert L. and Norman E. Gronlund. 2000. *Measurement and Assessment In Teaching 8th Edition*. New Jersey: Prentice-Hall, Inc.
- Madsen, H. S. 1983. *Techniques in Testing*. New York: Oxford University Press.
- Sudijono, Anas. 2005. *Pengantar Evaluasi Pendidikan*. Jakarta: PT Raja Grafindo Persada.
- Sudjana, Nana. 2008. *Penilaian Hasil Proses Belajar Mengajar*. Bandung: PT Remaja Rosdakarya.
- Thoha, M. Chabib. 1991. *Teknik Evaluasi Pendidikan*. Jakarta: Rajawali.