Article

Students' English Summative Test Analysis : A Case Study at SMA Perguruan Rakyat 1 Jakarta

Mulyadi[1], Siti Nurani[2]

[1] *Guidance and Counseling Study Program, Faculty of Education and Social Sciences, Universitas Indraprasta PGRI, Jalan Nangka No. 58C Tanjung Barat, Jagakarsa, Jakarta Selatan, 12530, Indonesia*

[2] *English Education Program, Faculty of Languages and Arts, Universitas Indraprasta PGRI, Jalan Nangka No. 58C Tanjung Barat, Jagakarsa, Jakarta Selatan, 12530, Indonesia*

A B S T R A C T

The present research aims at analyzing the English summative test at SMA Perguruan Rakyat 1 in the Second Term of 2018/2019 Academic Year. The analysis concerns to difficulty level and discriminating power to know the teacher-made test English summative test at SMA Perguruan Rakyat 1 whether or not the test has good items, especially of its difficulty level, discriminating power and the effectiveness of the distracters. Such research procedures are employed, such as observation, literature review and statistical analysis. Specifically, the result of the research is expected to be beneficial for educational practitioners to have a more-depth insight of English summative test analysis. Furthermore, the result of the research is also expected to give valuable knowledge for those who are interested in analyzing the English summative test writing, mainly to the indicators of its discriminating power and distracters. Lastly, it is also hoped to be as reference for further researchers to draw related concepts of English summative test analysis that is integrated with various disciplines implemented in the classroom contexts.

INTRODUCTION

English is one of the subjects taught in every formal school in Indonesia. Both spoken and written English are the factors which could influence an individual's standard of living. It means that someone who has ability in English, she/he will get job easier. Moreover, English is used in any appropriate field; those are education, technology, art and culture. English has also played an important role in the daily life of Indonesian people for many years. Being aware of this, English was appointed to be a compulsory subject in schools. The English syllabus of curriculum 1994 for Junior High School stated "Pada akhir sekolah lanjutan tingkat pertama, siswa memiliki ketrampilan membaca, menyimak, berbicara dan menulis dalam bahasa Inggris melalui tema yang dipilih berdasarkan tingkat perkembangan dan minat mereka, tingkat penguasaan kosa kata dan tata bahasa yang sesuai". It is obvious that the students of Junior High School must have skills in English, namely listening, speaking, reading and writing. So, the students are expected to be able to communicate in English both spoken and written well.

An English teacher needs to concern on some important components, such as materials, methods, instruments of teaching and evaluation to support students' progress. For identifying the students' achievement in teaching learning activity, a teacher must do an evaluation. The evaluation as stated above is a part of teaching and learning process. In addition, the evaluation cannot be separated from the components in teaching and learning activity. In this case, every teacher needs to do the evaluation to get information to make a decision. Then, he will know the effectiveness of the method he uses in teaching. In evaluation, a teacher needs an instrument to measure what students achieve. Mehrens and Lehman in Purwanto (2001) told that to teach without testing is unthinkable. Based on their argument, test is an important thing in teaching. Without testing, teacher will not get the information about the result of his teaching learning process. Then, the result of the test will be a function of feedback of instructional objectives and developing of a good systematic teaching learning process. In other words, it can be concluded that evaluation helps teachers reach out for the instructional objectives that has been already established.

All educators must have a good test for getting accurate information. Harris (1997) said that all good tests possess three qualities, namely: validity, reliability and practicality. Therefore, Heaton (1988) added that a good test must also have discriminating power and difficulty level. It means that a good test has standard requirements must be met by the test makers. A test existed in schools must be suited to students' need, for instance, Senior High School in Jakarta held an English summative test for the second term of 2018/2019 academic year. The test was made by a team of English teachers but the researchers heard from the students there, the test was easy and another said it was difficult, the others said it was average. For that reason, an analysis of English summative test items is needed to be done. In this case, the analysis is concerning to difficulty level and discriminating power to acknowledge the teacher-made test English summative test at SMA Perguruan Rakyat 1 in the second term of 2018/2019 academic period have good items or not especially the difficulty level, discriminating power and the effectiveness of the distracters. Therefore, the researchers are interested in doing a research entitled "An analysis on English Summative test at SMA Perguruan Rakyat 1 in the Second Term of 2018/2019 Academic Year".

In line with the problems exemplified above, the research formulates its focus on finding the correlation of such factors, such as the discriminating power and the difficulty level in the summative test items. This is conducted to define a more-depth understanding about the test analysis itself as it is helpful for teachers in doing the analysis of the test items of each student. This present research releases much of contribution, that is, be beneficial for educational practitioners in benefiting the summative test in learning English in the senior high school level of education and be as reference for further researchers to raise new concept of analyzing the English summative test, combining with other links of disciplines, primarily in the discussion of other field used in the classroom contexts.

Evaluation is an important part of teaching-learning activity. Through evaluation, teachers are able to plan, to derive and to apply information to make alternative decision. Wrighstone in Purwanto (2001) said that educational evaluation is the estimation of the growth and the progress of pupils towards objectives or values in the curriculum. Evaluation can help teachers finding out some information to measure and to diagnose how far the teaching-learning process that students had been achieved and the improvement of the material in the instructional objectives.

Evaluation is defined as the process of gathering data about the goals, environment, personnel, methods and content of an educational program on the basis of which decisions can be made about the program". According to him, we can see evaluation plays an important role in many facets of the school program. It is clear that evaluation in education is a systematic program that cannot be separated from curriculum. It is not only held in end of a certain program but it also held in the beginning or during that semester program takes place. The information and data concerning is necessary in the object that is evaluated. It can be concluded that evaluation is a process of teaching-learning activities and the result can be useful for diagnosing and measuring how far the students' progress in the class.

In evaluation process, teachers usually use test as a tool to measure the ability of students. Almost each evaluation that teachers had been doing cannot be done well without test. Therefore, both educators and students must know what test is. Test is an indicator to detect a student's ability that was considered from sequence of score which they get. A test is an instrument or a tool of measurement student's in a certain time. Test that is usually held in a school is summative test. Griffin and Nix (1991) stated that in some schools there was no continuous assessment and final marks were to assessment. Many students in these schools learned to cram for exams. Through summative test which held in the end of semester, teacher derives data and information to certify each student to follow the next level or not.

For measuring the conformity, efficiency and consistency of a test, a test must have good characteristics. Harris (1997) described that the test needs to be valid, reliable and practical. The first characteristic of a good test is valid. In this case, it must be seen whether the test used really measures what is supposed to measure. The second good characteristic of test is reliable. Harris (1997) stated that reliability is meant the stability of test scores. A test cannot measure anything well unless it measure consistently. Hence, the consistency is important to know the reliability

of test. Then, practicality of the test is necessary to be concerned. A test could be told to have a good practicality if the test is enabled. Heaton (1988) uttered that a test must be practicable. In other words, it must be fairly straight forward to administer.

The multiple-choice item is one of the item formats. It consists of some options of each item, has a correct answer and it must be chosen by students. Brown (1996) stated that multiple-choice format requires students to examine same language materials and then selected, usually from among three, four or five options, the answer that the best completed a statement or best filled in a blank in the statement. Supported that, Harris (1997) said that these items generally can be answered fairly rapidly, the test writer can include a large number of different tasks (that is, individual items) in the testing score. Finally, in as much as the examinee responds by choosing from several possible answers supplied by the test writer, scoring can be done quickly and involves no judgments as to degree of correctness. It is clear what Harris meant. This item is not only easy and quick to be answered by students but it is also easy and quick to be scored by the teacher. Besides, it usually consists of large number items.

After test is done by students, it is necessary to know whether or not test consists of a good test and also the test could differentiate the upper and lower student or not. To know that case, teacher should analyze it. This activity called "item analysis". As Harris (1997) described that after the test pretest answer sheets have been accumulated, the item should be analyzed to determine their effectiveness in term. The item analysis is necessary to be done. This form is constructed primarily for using the multiple-choice item. Not only Harris but also Madson (1983) told that selection of appreciate language item is not enough by itself to ensure a good test. Each question needs to function properly; otherwise, it can weaken the exam. Fortunately, there are some rather simple statistical ways of checking individual items. This procedure is called "item analysis". It means item analysis is important because we need to see the quality of any content.

Based on the explanations above, they show that the importance of item analysis in improving the quality of a test will ease for the teacher in providing a better test for their students in advance. Heaton (1975) explained that all items should examine from the point of view of their difficult level and their level of discrimination.

Item difficulty is a part of item analysis. It needs to be done because it is good to know which item is too difficult and too easy for students. Heaton (1975) defined the index of difficulty (facility value) of an item simply shows how easy or difficult the particular item proved in the test. It was generally expressed the percentage of the students who answered the item correctly.

Another primary thing in an item analysis is how well it discriminates between weak and strong examinees in the ability being tested. Brown (1996) stated that item discrimination indicated the degree to which an item separated the students who performed poorly. The discrimination of an item relates to its ability to distinguish between more and less knowledgeable students. It shows us the students who performed well or badly in each item. This research is concerned with the English summative test items, which consists of 50 multiple choices with 4 options before that a multiple choice has distracter as incorrect answer. Meanwhile, Tinambunan (1988) said that a distracter is any of the incorrect answer options in multiple choice test items. According to that statement, the researchers understand that the purpose of the distracter is to distract the incorrect answer.

METHOD

Such procedures are taken to collect the data as the followings:

1. Observation
Observation is a method of collecting the information by investigating directly to the analysis of English summative test.

2. Interview
Collecting the data and information by interviewing teachers and students to the research importance.

3. Literature Review
The data and information are collected by reviewing sources or references. In this phase, literature review starts off with the analysis of the previous research from both offline and online sources. The list of books, journals, and websites as the references is remarked in the references section.

4. Statistical Analysis
Collecting and analyzing the data by using statistical calculation in order to get the objective result of the test item analysis. There are two types of item analysis to be explored, that is, item difficulty and item discrimination.

The item difficulty will be analyzed based on Heaton's facility value (1975) by using the formula as follows:

$$FV = \frac{(CU+CL)}{2n} \qquad (1)$$

Namely:
FV = Facility value or the index of difficulty that we are looking for.
CU = Sum of the students from the upper group who answered correctly.
CL = Sum of the students from the lower group who answered.
$2n$ = Total sum of the lower and upper group.

To identify the level of difficulty, the researchers use Sudjana's guideline (1991). If the FV is:

0.0 – 0,30 = Difficult
0,31 – 0,70 = Medium
0,71 – 1.00 = Easy

Meanwhile, the item discrimination will be analyzed based on the proposition of Brown (1996) in Heaton (1975) by the following formula:

$$DP = \frac{(CU+CL)}{N} \qquad (2)$$

Namely:

DP = Discriminating power.
CU = Sum of the students from the upper group who answered correctly.
CL = Sum of the students from the lower group who answered correctly.
N = Number of candidates in one group.

To determine whether an item has good discriminating power or not, can be seen the following criteria:

0.0 – 0.20 = Poor
0.21 – 0.40 = Satisfactory
0.41 – 0.70 = Good
0.71 – 1.00 = Excellent

RESULTS AND DISCUSSION

The English summative test consists of 50 items with 5 options and had been answered by students second grade SMA Perguruan rakyat 1. The scoring and coding had been done by the writers; 1 for the right item and 0 for the wrong item that the responses have answered. After that, the responses are arranged in the table responses tabulation of students. Then, the scores ability of the second grade students in English summative test is obtained from the table of responses tabulation of students in class II. The next step, the data from the table was transferred into a distribution frequency. The scores distribution frequency can be seen in the table as follows:

Table 1 Scores Distribution

No.	Class Interval	Frequency
1.	41-47	9
2.	34-40	14
3.	27-33	5
4.	20-26	5
5.	13-19	2
6.	6-12	1

From the distribution of frequency, the writers finds that there are one score in the interval 6-12, two scores in the interval 13-19, five scores in the interval 20-26, five scores in the interval 27-33, fourteen scores in the interval 34-40 and nine scores in the interval 41-47.

Based on the procedure of the research, a number of respondents who answered the items correctly are found. Those candidates are divided into two groups; they are upper group and lower group. Each group consists of 10 students. In the upper group, no one student got all items correctly. It is found that a student got 48 items, four students got 46 items, a student got 45 items and three students got 44 items and one student got 43 items. Meanwhile in the lower group is only one student got 36 items, one student got 34 items, one student got 32 items and one student got 31 items. One student got 30 items, one student got 29 items and two students got 24 items. Then, one student got 23 items and one student got 18 items.

From the upper and lower groups' data, the writers calculate the index of facility value and discriminating power to get a more-depth information of the data analysis of the test item given to the students. The result can be presented as follows:

1. The Difficulty Level

The English summative test consists of 50 multiple choice items with 5 options. The data from upper and lower group are calculated by using Heaton formula in order to get the level of difficulty. The result of the difficulty was analyzed by using Arikunto's guideline (1988). Below is the computation distribution of the difficulty level observed to identify the test categories.

Table 2 Computation Distribution of the Difficulty Level

No.	Range of Difficulty Level	Category	Frequency	Percentage
1.	0,17 – 1,00	Easy	30	56%
2.	0,31 – 0,70	Medium	19	42%
3.	0,00 – 0,30	Difficult	1	2 %
	Total		50	100%

2. The Discrimination Power

The English summative test consists of 50 multiple choice items with 5 options. The data from upper and lower group are calculated by using Heaton formula in order to get the discriminating power. The result of the discrimination power analyzed by using Sudjana guideline. Below is the computation distribution of the discrimination power:

Table 3 Computation Distribution of the Discrimination Power

No.	Range of Discrimination Power	Category	Frequency	Percentage
1.	0.00 – 0.20	Poor	19	38%
2.	0.21 – 0.40	Satisfactory	12	24%
3.	0.41 – 0.70	Good	18	36 %
4.	0.71 – 1.00	Excellent	1	2 %
	Total		50	100%

CONCLUSION

Having finished the research of English summative test, the writers make some conclusions. Based on the findings in the previous chapter, the writers conclude only one of 50 items is an excellent item, eighteen of them are good items, twelve of them are satisfactory items and nineteen of them are good items. It means 31 items (60%) can be maintained and used in the summative test next period and 19 items (40%) must be revised. These findings of research indicate the English summative test items for the second grade term 2018-2019 academic years at SMA Perguruan Rakyat 1, Srengseng Sawah Jakarta Selatan have good criteria of the difficulty level, the discrimination power and the effectiveness of distracters.

REFERENCE

Arikunto, S. (1988). Teknik Dasar Evaluasi Pendidikan. Bandung: PT Remaja Rosdakarya.

.

Brown, J. D. (1996). Testing in Language Programs. New Jersey: Prentice-Hall. Inc.

Griffin, P., & Nix, P. (1991). Educational Assessment and Reporting. Australia: Harcourt Brace Jovanovich Inc.

Harris, D. P. (1997). Testing English as a Second Language. Bombay: Tata Mc Graw Hill.

Heaton, J. B. (1975). Writing English Testing. Singapore: Longman.

_____. (1988). Writing Testing English Language. London: Longman.

Madson, H. S. (1983). Techniques in Testing. New York: Oxford University Press.

Purwanto, N. (2001). Prinsip-prinsip dan Teknik Evaluasi Pengajaran. Bandung: Remaja Rosdakarya.

Sudjana, N. (1991). Penilaian Hasil Proses Belajar Mengajar. Bandung: Remaja Rosdakarya.

Tinambunan, W. (1988). Evaluation of Students Achievement. Jakarta: DepDikBud