

EVALUATING ENGLISH FINAL-SEMESTER TEST MADE BY THE TEACHER AT SMA N 5 BANDAR LAMPUNG

Saghina Meividia Anas, Cucu Sutarsyah, Muhammad Sukirlan

Email : saghinamivdiaa@gmail.com

ABSTRACT

Tujuan penelitian ini adalah untuk menentukan kualitas soal tes akhir semester yang dibuat oleh guru di kelas X SOS 4. Penelitian ini menggunakan pendekatan deskriptif kualitatif dengan data yang diambil dari lembar jawaban siswa dan soal pilihan ganda menggunakan iteman. Hasil penelitian menunjukkan bahwa (1) validitas soal kurang baik, 2) reliabilitas soal seimbang, yaitu 0.492 (3) tingkat kesukaran dari 16 bulir soal (35.56%) baik, 16 (35.56%) sangat sulit, 11 (24.44%) sulit, 1 (2.22%) mudah dan 1 (1.22%) sangat mudah, (4) daya beda pada soal dari 10 bulir soal (22.22%) berkategori tinggi, 5 (11.11%) pada tingkat rata-rata, 2 (4.45%) rendah, dan 28 (62.22%) perlu diganti, (5) kualitas bulir soal dari 95 (42.22%) harus diganti, 113 baik, dan 17 alternatif sangat baik. Kesimpulan dari penelitian ini adalah kualitas soal tes adalah seimbang.

The objectives of this research were to determine the quality of final test at X SOS 4 class. This research applied qualitative descriptive approach which data were taken from the students' answer sheet and the multiple choice question by using iteman. The result of the research showed that (1) The validity of the test was not proper enough to used, (2) The reliability was sufficient, that is 0.492, (3) The level of difficulty consisted of 16 (35.56%) test items considered good, 16 (35.56%) very difficult, 11 (24.44) difficult, 1 (2.22) easy and 1 items very easy, (4) The discrimination power of the test items consisted of 10 (22.22%) high, 5 (11.11%) average, 2 (4.45%) low, and 28 (62.22%) need dropping, (5) The qualities of the alternatives consisted of (42.22%) should be dropped, 113 as good distractors, and 17 as very good distractors. It can be concluded that the quality of the test items was moderate.

Keywords : *evaluating, validity, reliability, level of difficulty, discrimination power*

INTRODUCTION

Test is an important part of teaching learning process that cannot be separated in the implementation of the teaching and learning process itself. In teaching learning process, testing is important to measure the ability of the students. Heaton (1990:5) states that, both teaching and testing are so closely interrelated that it is virtually impossible to work in either field without being constantly concerned with the other. Besides, the result of the test can be used to see how successful the teaching and learning process is implemented. As stated by Woods (2005:25), this is important as testing can often influence the nature of what is taught and how it is taught.

To measure the students' ability, multiple choice testing can be an efficient and effective way. It asks the student to recognize a correct answer among a set of options that include 4 wrong answers called distracters.

As the measurement to evaluate the students' understanding in the teaching and learning process, the test should meet the criteria of a good test. There are several aspects that constitute the criteria of a good test. According to Sulistyo (2007:21) who states that a good test must meet the requirements: reliability, validity, practicality/usability, and economy. If a test does not meet the requirements of a good test, the test will produce biased scores that will not reflect the real ability of the test takers. The test that is developed in this case is known as teacher-made test.

The teacher-made test here is constructed and administered by the teacher. Hence, the items of the test are not analyzed systematically by the teacher. Arikunto (2003:147) states that the teacher-made test is constructed from the items that are commonly not tried out, analyzed, and revised first. Therefore, based on that case, the quality of the teacher-made test is questionable.

Regarding the quality of the teacher-made test, it is obvious that the test items need to be tried out and analyzed. However, most of the teachers are still using manual method to analyze the quality of their test items. Actually, the teachers can use iteman software program to analyze the quality of the test items they have made.

Basically, it is a software that is used to analyze test item and determine which test item is good and which is not. As iteman is considered useful, the teachers are more expected to have an involvement in assessing the multiple choice tests using the item analysis program. However, most of the English teachers were unfamiliar with iteman software program. They lacked of knowledge about how to analyze the test items, to decide the validity, reliability, discriminating power, and level of difficulty, especially by using iteman program.

Besides, most of students rarely got good scores because based on their opinion, the test items were too difficult. It might be assumed that the test items made by the teacher were not considered as a good test based on the criteria. They did not know how to measure the quality of the test items they have made before those items were used. Hence, most of the students did not get good scores on the exam because of its difficulties.

Based on the explanation above, the problem concerning with the analysis of test items was considerably need to be investigated because the test items which will be used for testing the students have to be analyzed before being tested. Therefore, the researcher was interested in analyzing the Final Semester test items made by the teacher using iteman software and qualitative analysis, which were used for the first semester of the first year students in SMA. The purpose of the item analysis is to determine whether the test items are good or not, based on the validity, reliability, discriminating power, and level of difficulty.

METHOD

The object of this current research was teacher-made English final semester test items for the first grade on the first semester at SMAN 5 Bandar Lampung. It consisted consisted of test items and the student's answer sheets. The design of this current research was descriptive and evaluative which described the result of an evaluation on an object based on standard criteria using iteman. There were 45 questions while the number of answer sheets used in this research were around 30. Both of them were analyzed based on standard criteria, i.e., level of difficulty, discriminating power, reliability and validity. This research was intended to evaluate and propose some revisions for final semester test in the first grade in SMAN5 Bandar Lampung. Final semester test items were collected as the data for this research. The tests were used to determine whether there were some revisions for the test items based on the result of analysis while using iteman. The data analysis focused on evaluating the test items to find out whether the test items being used well, should be totally revised, or partially revised, or dropped totally

RESULTS AND DISCUSSIONS

This research was aimed to answer the questions whether the test items made by the teacher in the final test fulfill the criteria of a good test, i.e., validity, reliability, level of difficulty and discrimination power, and to determine whether there are several numbers should be revised or totally dropped after using iteman software program. Several procedures had been done in order to answer the questions. The data had been computed and analyzed to find out the answers.

The result showed that the validity of the test items were not fully valid because it did not fulfill the criteria of face, content and construct validity itself. Furthermore, according to the iteman, the reliability of this research was valid because the alpha showed the range number of 0.492 which means valid. In addition, according to the result of level of difficulty, not all items in the final examination had good level of difficulty. The results showed that the level of difficulty of the test items consisted of 16 (35.56%) test items which were considered good/average, 16 (35.56%) test items which were very difficult, 11 (24.44) test items which were difficult, 1 (2.22) test items categorized as easy and 1 items which was very easy. The question can

be said not good or has low quality because most of the items included in the category the items should be dropped or revised. Problem with these categories can be repaired by replacing the question where some students were able to answer it because it is likely most of the students had comprehended the material in the questions. Regarding to the discrimination power after using iteman, it consisted of 10 (22.22%) categorized as high, 5 (11.11%) classified as average, 2 (4.45%) categorize as low/need revising, and 28 (62.22%) grouped into very low or need dropping, (5) The qualities of the alternatives consisted of 95 alternatives (42.22%) should be dropped, 113 alternatives have good distractor, and 17 alternatives have very good distractor. The researcher indicated that some of the items fulfilled the requirements of the quality of the good test item but some of them did not. There were more than 60% items should be dropped, meanwhile only around 20% items that can be used directly without any revision. This suggested that the teacher should revise many items before using them.

The final research question was, what revision should be made by the teacher on English final semester test items at SMAN 5 Bandar Lampung?. From the result of analysis by using iteman, it was found that 95 options should be revised because they were classified into low category. It can be seen from the following graph below.

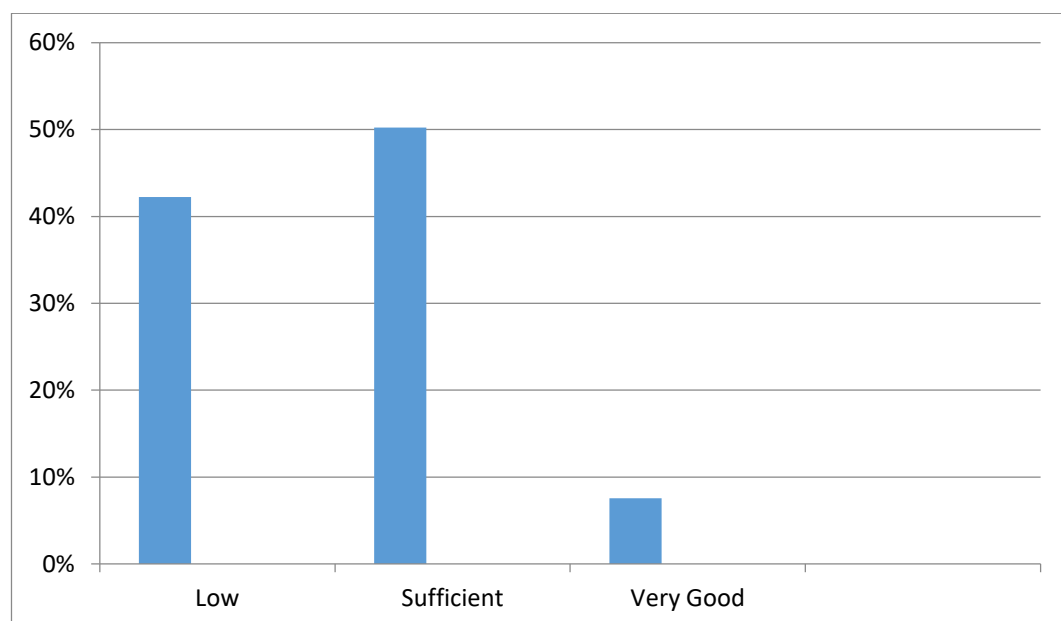


Figure 4.5. Proportion Endorsing (Proportion of the answer items)

Based on the analysis above, it can be concluded that not all the test items have good quality and can be accepted. Less than 50% the items should be revised by the teacher before giving the test item to the student in the examination. A well-developed test should provide an opportunity for students to show their ability to perform certain language tasks. A test should be constructed with the goal of having students learn from their weaknesses, in this way a good test can be used as a

valuable teaching tool according to Rajhy (2014). Therefore, the test can be qualified as a good test item

Conclusions and Suggestions

The validity of the English final test semester test items was tried out, analyzed and revised. However, the validity of the test items should be compared with the current of English curriculum used in SMAN 5 Bandar Lampung. Based on the discussion between the researcher and the English teachers at SMAN 5 Bandar Lampung, each of the items was relevant with the syllabus and some items were not. Concerning with the face, content and construct validity, the test items were not fulfill the requirements. Therefore, the test which is prepared by the English teacher is considered to be not fully valid.

Based on the result in the output data in the iteman, the following results are drawn as follows:

1. The reliability of the English final test semester test items based on the result of the iteman was categorized as average/sufficient because the alpha value (reliability of the test items) was 0.492 which lied between the ranges of 0.401-0.700. It means that the reliability of this English final semest test items was categorized as average/good. the test items were proper enough to be tested to the students because they are reliable
2. The level of difficulty of the English final semester test items were classified into five categories, i.e., average/good, very difficult, difficult, easy, and very easy. Based on the Proportion Correct (level of difficulty), there were 16 out of 45 (35,56%) test items which considered good or average. The question can be said not good or has low quality because most of the items included in the category the items should be dropped or revised. Problem with these categories can be repaired by replacing the question where some students were able to answer it because it is likely most of the students had comprehended the material in the questions.
3. The discriminating power of the English final test items can be classified into four categories, i.e., high, quite average, low/need revising, very low. It indicated that some of the items fulfilled the requirements of the quality of the good test item but some of them did not. There were more than 60% items should be dropped, meanwhile only around 20% items that can be used directly without any revision. This suggested that the teacher should revise many items before using them
4. Proportion endorsing (the qualities of the options) in English semester test items, regarding on the iteman analysis were classified into three classifications, i.e., least/drop, good enough/sufficient, and very good. It was obtained that the

options of the 45 items each of which consist of A, B, C, D, and E totaling 225 options. From the result of analysis by using iteman, it was found that 95 options should be revised because they were classified into low category. It can be concluded that not all the test items have good quality and can be accepted. Less than 50% the items should be revised by the teacher before giving the test item to the student in the examination.

5. Regarding to the interview with the English teachers, it can be concluded that they never analyse the test items before it was used to the students. The teachers had not been familiar with the iteman, hence they were not using it to determine whether the test items they had made were proper enough to used or not.

In line with the conclusions above, the following suggestions are proposed as follows:

1. The teachers should be able to make a proper test items for the students based on the quality of a good test before it is used.
2. The teachers should be familiar with all the terms related to the quality of a good test, such as validity, reliability, prop. Correct (level of difficulty), point biserial (discriminating power), prop. Endorsing (the alternatives/ options), distracters, key answers, alpha and standard deviation.
3. The teachers should be familiar with the iteman program to make them easier in assessing the student's ability.
4. The teachers should be trained to use the item analysis program (iteman) in order to improve the quality of the test.
5. The test items which is made by the teachers should be tried out first, before it is used to the students.
6. The teachers should be trained on how to analyze the test items effectively and efficiently and how to revise the bad test items.
7. The researcher should be able to analyze the other test items, such as Mid Semester test, Final School test (UAS), and National Examination (UN).

REFERENCES

- Arikunto, S. (2003). *Dasar-dasar evaluasi pendidikan (edisi revisi)*. Jakarta: Bumi Aksara.
- Heaton, J. B. (1990). *Writing english language tests*. New York: Longman Inc.
- Rajhy, H. A. A. (2014). Five characteristic of a good language test. *National Journal of Extensive Education and Interdisciplinary*, 2(4), 61-66.

Sulistyo, G. (2007). Tests, assessment, and measurement in English as a second language at schools: An introduction to their basic concepts and principles. Unpublished coursebook. Malang: English Department, State University of Malang.

Woods, C. (2005). *Professional development for teachers: Teaching and assessing skills*. Cambridge: Cambridge University Press.