

AN ANALYSIS OF ENGLISH TEACHER-MADE TESTS

Nana Erna¹, Baso Jabu², Kisman Salija³

¹²³Faculty of Language and Education, State University of Makassar, St. Gunungsari Baru. Bonto Langkasa, Makassar, Indonesia

¹nanaerna@gmail.com

²basojabu@gmail.com

³kismansalija@unm.ac.id

Abstract: This writing highlights the English teachers-made tests. This thesis is focused on the quality of the multiple-choice test that was made by the English teachers. This study was conducted in 2018 with 2 teachers and 40 numbers of questions in junior high school. Mixed method was used as the method of this study. The data were collected through interview and validation sheet that has been validated by the expert. The quantitative data were analyzed by using ANATES program version 4.0 and qualitative data were analyzed by using validation sheet. The findings of interview revealed that the teachers' procedures in constructing a test were separated into three main stages. They were preparing test specification, writing, and analyzing. Then, the result of analysis based on the principles of writing a test using validation sheet in form of a multiple-choice item that is appropriate with the material, construction, and language. It was found that several questions still need to be revised. The findings of analysis based on the items of analysis such as the reliability obtained is not reliable. The test has the level of difficulty in category too easy 15%, easy by 35%, medium by 42.5%, and difficult by 7.5%. Discrimination index only 40%, it can be categorized good. Then, the powers of distractor were 50% functioning.

Keywords: *analysis of questions, teacher-made tests.*

1. Introduction

Evaluation has the crucial role in every education system. It can show how far the progress of educational outcome is. One of the evaluations' tool is testing. The result of teaching without testing will be useless because the objectives of learning can be evaluated by having a test. The test is constructed to find out the achievement of the students in teaching and learning process. Then, from the result, it can be determined that the learning objectives are successfully obtained or not. 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 (Heaton, 1988).

According to Woods (2005), testing can often influence the nature of what is taught and how it is taught. By giving a test, the teacher can perceive how far the students understand the material that is delivered by the teachers and how well the materials are presented. The test cannot be expected only to follow the teaching process, but also it should be supportive of good teaching and exert a corrective influence on bad teaching (Jabu, 2008). In this case, the teacher has an obligation to deliver material well and also has an authority to assess the success of learning (Nakarin, 2009).

But nowadays, the teacher who constructs the items of the test is lack of the techniques and skill in constructing it. Arikunto (2003) stated that the teacher-made test is constructed from the items that are commonly not tried out, analyzed, and revised first. In addition, the items of the test are directly used without analyzing the quality. It is less attention such as; (1) the material being tested is non-essential, (2) not drafting the scoring guidance

on the description. With these conditions, of course, the quality is "unpleasant". People never think that the students cannot answer the questions correctly not because they did not understand the material that has been taught, but they did not understand the items of the test that was written by the teachers. Because testing is very essential, teachers as a test constructor should be able to construct a good test, in another word, the teachers have to be a master in techniques or procedures for evaluating the students. The good test must be valid and reliable.

Based on the previous explanation, the main objectives of this research were:

1. To identify the teachers' procedures in constructing the tests for the students.
2. To analyze the quality of the English teacher-made tests based on the principles of writing a test.
3. To analyze the quality of the English teacher-made tests based on the validity, reliability, level of difficulty, discrimination index, and distractors.

2. Literature review

The test is a systematic procedure. The test items are arranged according to specific ways, principles, administrative procedures and scoring should be clear and specific, then each student taking the test must receive the same points and in similar conditions. The tests measure the behavior. The test items require the student to show what is known or what the student is learning by answering the items or doing the tasks desired by the test. another word, the test is a set of items which is a sample of the population that measures certain behaviors in the form of skills, knowledge, intelligence, talent and so on where in the

implementation of students are encouraged to give a maximum appearance.

Tests are not only used to find out students' attribute, tests also can help the teachers. Tests are enabling teachers to increase their own effectiveness by making adjustments in their teaching to facilitate a certain group of students or individual in the class to benefit more (Heaton. 1988). So, the teachers can see the reflection of their teaching techniques that they used in teaching and learning process.

Therefore, the teacher must be a master in techniques how to write a good test to the students. Specifically, for multiple-choice tests, because multiple-choice easy to check but difficult to construct. According to Hughes (1989), stages of test construction is divided into three main stages. The writer can start with writing specification for the test, writing the test, and the last step is pretesting. But when the teachers construct based on those stages, they have to regard about the principles to write a good test. The principles of writing the test is a guide or guidelines that need to be followed by the writer in order to result of the test has a good quality. A good test quality is able to filter the necessary information and function optimally. The principles of writing multiple choice questions are divided into three things, namely Material, Construction, and Language Depdiknas (2008), which are described in various ways as follows: 1) Problems must be in appropriate with the basic competencies and indicators contained in the curriculum, 2) The choice of answers should be functioning, homogeneous, and logical, 3) Each question should have only one right answer, 4) Basic questions should be formulated clearly, 5) The formulation of the questions and the choice of answers should be only a necessary statement,

6) The point of questions do not give directions to the right answer, 7) Basic questions do not contain a statement that is a double negative, 8) Pictures, graphs, tables, diagrams that is contained in the questions must be clear and functional, 9) The length formula of the answer choice should be relatively same, 10) The choice of answer should not contain the statement, "all of the answers above are wrong", or "all of the answers above are correct", 11) Category answers form of numbers or time must be arranged in order of the size of the numeric values, or chronological time, 12) The item does not depend on the answer to the previous question, 13) Each question must use correct English language structure, 14) Use communicative language, so that it is easy to understand, 15) Do not use the local language, if the questions will be used for other regions or national, 16) options do not repeat the word or phrase that is not a unity of understanding. Put the word on the subject.

3. Research Method

A mixed-method research design was implemented in this study. Mixed method research design collects and analyses both quantitative and qualitative data (Lodico, Spaulding and Voegtler, 2006). This research aimed to obtain systematic information related to the quality of the tests as objective and clearly than to identify teachers' procedures in constructing a multiple-choice test for their students by using the interview. Qualitative analysis fulfilled by using a format of validation sheet or expert review. It distributed to identify the quality of the tests based on the principles of writing a test. The principles are the appropriateness with the material, construction, and language. While the quantitative analysis fulfilled by using

ANATES program version 4.0 Then, the result included the validity, reliability, level of difficulty, discrimination index, and distractors power. This study was conducted at SMPN 36 Makassar, South Sulawesi. The object involved in this study was the examination test or final test in the odd semester 2017/2018 academic year that made by the English teachers who teach in the third grade and the students answer sheets. The final test consists of 40 items. The researcher took 100 answer sheets from 210 students answer sheets of the third grade.

4. Findings and Discussion

1. The procedures of teachers in constructing teacher-made tests.

Based on the result of the interview was obtained, the first stage is the teachers prepared the test specification before writing the real test for their students. And their test specification appropriate with some aspects namely standard competences, syllabus, learning objectives, and materials. Not only prepare a test specification but they prepare item bank as well for their final test. But founded on the analysis not all items were appropriate with those aspects.

The second stage is writing a test item. Based on the result of the interview found that sometimes the teachers combined test item that has been chosen from the item bank and textbook, internet, etc. Before writing the items, the teachers choose the material, write test specification and then construct a test.

The third stage is analyzing. The analysis process actually should be started by test tryout. Based on the teachers' response, sometimes they analyzed it but sometimes they did not analyze it. They only focus on the material that appropriates with their syllabus, standard competence, and learning objectives.

Consequently, they rarely analyzed their test. It is not suitable with the theory from Hughes (1989) and Sukardi (2008) stated that a good test must be valid and reliable. To know the items are valid and reliable, the writer must be analyzed it.

2. English teacher-made tests based on the principles of writing a test.

After analysis of English teacher-made tests for the third grade in SMPN 36 Makassar using format validation sheet that includes material, construction and language found that there were some items that need to be repaired. The test items made by the English teachers that are evaluated by the appropriateness of the materials divided into 3 aspects. The first aspect is the appropriateness with the indicator. There were 37 questions from 40 that appropriate with an indicator on the syllabus and the teachers' test specification. It can conclude that 92,5% of the questions were appropriate. And there were 3 questions inappropriate or 7,5%. Inappropriate items are 10, 11 and 12. The second aspect is the material that is asked in accordance with the competence (urgency and continuity). There were 37 proper questions. It can be accomplished that 92,5% were proper. While there were 3 improper questions, they are 10, 11, and 12. It can be calculated 7,5% improper items.

The last aspect is the questions only have one correct option. The result of the analysis indicates that all of the questions or items only have one correct option. In another word 100% questions were appropriate. According to Hasan (2008) in developing assessment instruments, teachers generally prepare test specification independently without being reviewed, revised and validated. Therefore, the

ability of teachers in preparing it needs to be continuously evaluated and improved.

There are some aspects of the test constructions which are described in various ways. Basic questions should be formulated clearly. Founded on the result, it shows that 92,5% of items were appropriate while inappropriate items indicated only 7,5%. Inappropriate items are 2, 3 and 8. The use of the appropriate letter option. This aspect revealed 40 questions or 100% that were suitable. The use of the right punctuation (full stop and comma) signify 85% proper and 15% improper. The use of letters in front of the options was correct. This aspect evidence only 90% right and 10% wrong. The use of the optional letters of all items was not appropriate because using lower case should use capital letters and the used of clue answering questions still need to be added so that the test participants are not confused for answering. Correspondence between questions with answer keys. In this aspect was quite good compared to other aspects because the calculation shows 97,5% items correct.

For constructions, aspects such as basic questions do not contain a statement that is a double negative, the choice of the answer should not contain "all the answer above are wrong" or "all of the answers are correct", and the item does not depend on the answer to the previous question. All of those aspects were proper. It can be decided 100% that has been appropriate. The next aspect is pictures, graphs, tables, diagrams, that is contained in the questions must be clear and functional. There were only 33 questions from 40 that utilized pictures and table then all of the questions were appropriate or 100%. The questions are number 1 until 33.

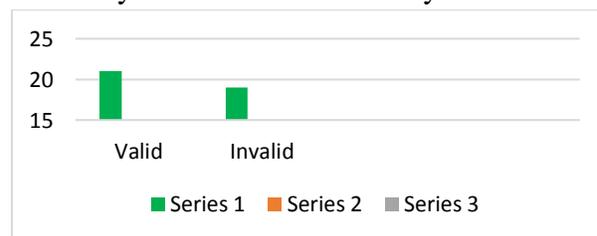
In terms of language, the aspects do not use local/regional languages and the choice of answers does not repeat words or phrases that are not a single unit of understanding. It has been 100% compliant. Aspects of each question using the language in accordance with the English grammar as much as 90% appropriate, then used communicative language as much as 97,5%. One of the factors that affect the quality of the problem is the teachers who compiled the question. According to Purnomo (2007), the special ability requirements possessed by the teachers in preparing the test appropriate with the standard is expert in the material to be tested, able to generate ideas, understand the characteristics of test participants and master in a technique of writing the questions.

3. *The quality of English teacher-made tests based on item analysis*

a. **Validity**

The validity of the content includes aspects of the material, construction, and languages based on the analysis of the test specification procedures. According to Gronlund (1976) in preparing a test to measure the students learning outcomes, content validity is the most important, because it can measure the entire material that has been taught. In this case, there is a certain skill that cannot be measured using multiple-choice test. Founded on the result of analysis by using ANATES program version 4.0.

Figure 4. 1 The result of English teacher-made test analysis based on the validity.



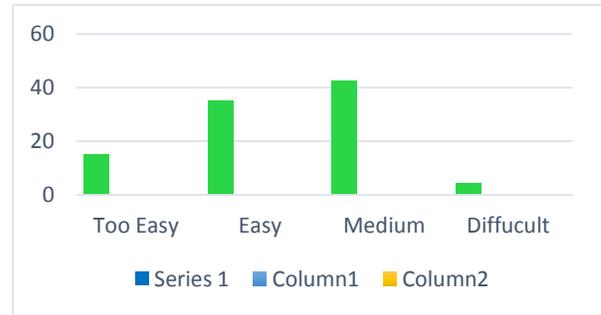
b. Reliability

The reliability of the evaluation sets to a stability evaluation tool or measuring tools in reporting the results of its size. Measuring tools consistent results in different situations is called a reliable measuring instrument (trustworthy). According to Madsen (1983), factors affecting the reliability of evaluation tools or measuring devices are the number of questions, homogeneity, problem-solving, the diversity of conditions, the level of difficulty, where the evaluation questions that are too difficult or too easy to cause lower levels of mistrust about the questions; The value of the coefficient of reliability of the test is interpreted by using a benchmark according to Sukiman (2012). If R is equal to or greater than 0.70 means the test results of the study have had high reliability (= reliable). Then, if R is smaller than 0.70 means the test results of the study have not had high reliability (= unreliable). Based on the result of the teacher-made final test in the odd semester for third grade of SMPN 36 Makassar based on the reliability presents that it has 0.63. It indicated the test unreliable.

c. Level of Difficulty

There are some criteria used to determine the type of difficulty level. Level of difficulty is appropriate with the provisions established that if the index level of difficulty from 0.00 to 0.30 classified difficult questions, from 0.31 to 0.70, 0.71 classified medium, and then 1.00 classified easy questions (Sukiman, 2012).

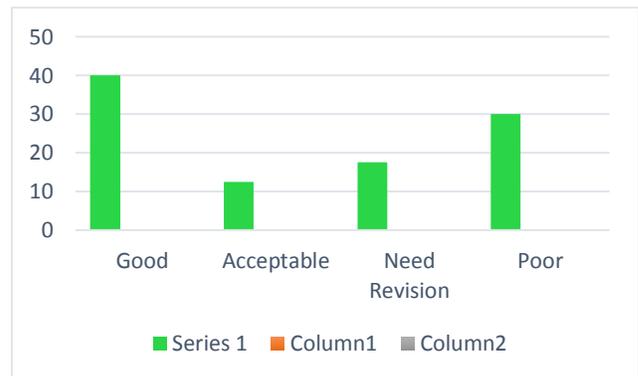
Figure 4. 2 The result of English teacher-made test analysis based on the level of difficulty.



d. Discrimination Index

According to Baron (1958), the distribution of difficulty item is the instrument must be able to distinguish the group of students who are good and the group of students who are less intelligent. Then, according to Sukiman (2012), there are some criteria for the discrimination index. Discrimination index coefficients are 0.40 to 1.0 is good; 0.30 to 0.39 is acceptable; 0.20 to 0.29 need revisions and -1.00 to 0.19 is poor.

Figure 4. 3 The result of English teacher-made test analysis based on the discrimination index.

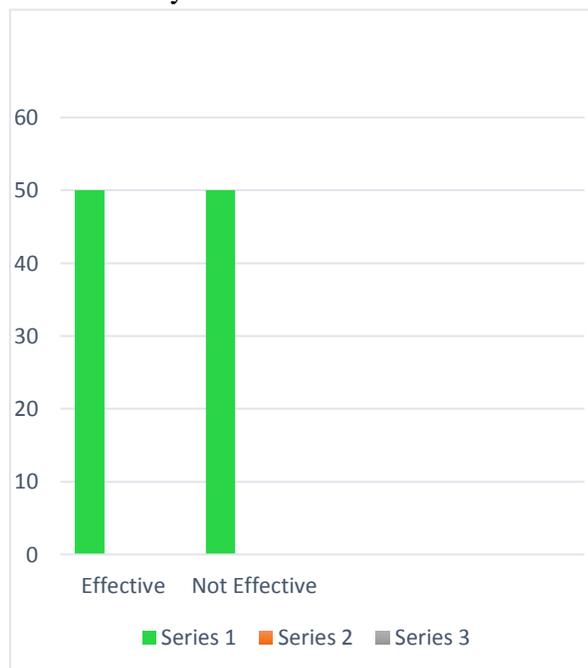


e. Distractors

The distractor analysis is an important part in item analysis because distractors' quality can influence the students' performance in test item. Basically, distractors are aimed to distract the less knowledgeable from the correct answer. The distractor needs to be both plausible and attractive. Consequently, they

can do the job properly. According to Ebel (1979), distractor power aims to mislead those who are less able (do not know) to be distinguished by the able (to know more). The distractor power is functioning properly if it has been chosen by more than 5% of the followers of the test ($p > 5\%$) and if it is less than or equal to 5% ($p \leq 5\%$) means the distractor power is not functioning properly.

Figure 4.4 The result of English teacher-made tests analysis based on distractors



5. Conclusion

Based on the result of data analysis and finding in the previous chapter, the researcher put forward the following conclusion:

The teachers' procedures for constructing a test consist of three main stages. The stages are preparing, writing, and analyzing the test. In preparing the test, the teachers construct a test specification before constructing a real test for the students, but their test specification is not appropriate with the standard of a good test specification. In writing a test, they were re-writing a test that has been tested in the

students' daily examination, but some of the item tests were constructed by themselves. While analyzing the test, they only changed the ambiguous words to simple words which are familiar to the students without analyzed the validity, reliability, difficulty level, discrimination index, and distractors.

The quality of English teacher-made test if reviewed based on the principles of writing a test still need to be revised about three aspects such as material, construction, and language on some items of questions.

The result of test analysis shows that no tests are perfectly good. Reliability analysis result obtained for multiple choice test is 0.63 that met with unreliable criteria. In terms of difficulty level, it is categorized into some classification. It shows that too easy category 15%, in easy category 35%, medium category as much 42.5%, and 7.5% difficult category. The discrimination index indicates that the test consists of 16 numbers discrimination features in the range 0.40 to 0.1 were well categorized. Acceptable discrimination consists of 5 numbers or 12.5%. Need revision categories were 17.5% and poor as much 30%. The effectiveness of distractors about 50% that has a well-functioning or effective.

References

- Arikunto, S. (2003). *Prosedur Penelitian Suatu Pendekatan Praktik*. Jakarta: PT. Rineka Cipta.
- Depdiknas. (2008). *Panduan Penulisan Butir Soal*. Jakarta: Depdiknas.
- Ebel, R. (1979). *Essential of Education Measurement. Third Edition Englewood Cliff*. New Jersey: Prentice Hall, Inc.
- Groundlund, N. E. (1976). *Measurement and Evaluation in Teaching*. New York: McMillan Publishing.

- Heaton, J. (1988). *Writing the English Language*. London: Logman.
- Hughes, A. (1989). *Testing for Language Teachers*. Cambridge: Cambridge University.
- Jabu, B. (2008). *English Language Testing* . Makassar: Badan Penerbit UNM.
- Lodico, M. G. D. T. Spaulding and K. H. Voegtle (2006). *Methods in Educational Research: From Theory to Practice*. San Fransisco: Jossey.
- Nakarin, P. (2009). Building Effective English Teacher-Made Test: A major KPI for Thai ESP Teachers. *Journal of Educational Measurement*, 185-202.
- Sukiman. (2012). *Pengembangan Sistem Evaluasi* . Yogyakarta: Insan Mandiri.
- Woods, C. (2005). *Professional Development for Teachers: Teaching and Assessing Skill in Foreign Language*. Cambridge: Cambridge University Press.