

CHAPTER II

REVIEW OF RELATED LITERATURE

This chapter covers the previous study, the nature of language testing, the function of test, the kinds of test, the criteria of good test and the principle of making multiple choice test items.

A. Previous Studies

The first, previous study was conducted in MTs-Muslimat NU Palangka Raya. This research used descriptive quantitative. The result showed that the problem of English summative evaluation faced by the second years students of MTs-Muslimat NU Palangka Raya were 79,87% of problems in answering the summative evaluation items. The classification of the problems is as follow: the first problem was the students' vocabularies were less. The second, they could not make English sentence. The third problem was the students did not know about the English tenses forms. The fourth problem, the students did know about the English grammar as like; adverb, conjunction, degree comparative, and English proposition. The cause problems of the English summative evaluation were caused by the test did not fulfill as the characteristic of a good test. The first, there were only 17 items as valid items from 35 numbers of item tested. The second, the summative test was in the very poor of reliability that was 0,07. The third, the summative was

impracticability; it was found 4 items were in the confusing meaning. The last, the level of the English summative evaluation items difficulty were very hard.¹

The second, previous study was conducted in MTsN-1 Model Palangka Raya. This research used quantitative. The result showed that the content validity of the English summative test of multiple choice made by the English teacher of MTsN-1 Model Palangka Raya for the second grade student at the first semester in academic year 2008/2009 had fair validity. The reliability of the test was 0,75 so it was high reliability.²

Based on the previous study above, the English summative tests made by the English teacher have to fulfill of criteria of good test. Because of that, the writer knew the quality of summative test made by the English teacher of SMK Kesehatan Borneo Bhakti Husada Palangka Raya.

B. The Nature of Language Testing

Language is a form of language science or linguistic, which is studying about the complications of language.

One of the evaluation instruments is a test. There are many meanings of the test. Test is a trial which is held to know some results from a certain subject which is taken from a student or a group of students.³ By testing teacher can know the ability of learning that students have.

¹Akhmad Halimi, *Summative Evaluation Problems of English Lesson Faced by the Second Year Students of MTs-Muslimat NU Palangka Raya*, Palangka Raya: STAIN Palangka Raya, 2008.

²Noor Amalia Shanti, *The Quality of Multiple Choice Item Made by the English Teachers of MTsN-1 Model Palangka Raya*. Palangka Raya: STAIN Palangka Raya. 2008

³M. Buchari in Suharsimi Arikunto, *Dasar-dasar Evaluasi Pendidikan*, (Jakarta: BinaAksara, 1987), p.199

Test is systematic procedure for observing and describing one or more characteristics of person with the aid of either a numerical or category system.⁴ The evaluation is a systematic process of determining the extent to which instructional objectives are achieved by a student.⁵

By testing, teacher can know the ability of learning that students have. Test is a trial which is held to know some results from a certain subject which is taken from a student or a group of students.⁶

C. The Kinds of Test

There are many kinds of the test which is appropriated by the identified characteristic. Generally, the kinds of the test can be used in many programs. In this case, kinds of the test are: based on organization, based on method of teaching, based on function, based on the form of the test, and based on scoring system.

1. Based on Organization

Based on organization, the tests are divided into standardize test and teacher made test. In this study the writer used teacher made test as the theory.

a. Standardize Test

Standard test measure the common objectives of a wide variety of schools have standard procedures of administration and scoring and provide norms for interpreting the scores.⁷

⁴Anthony J. Nitko, *Educational Test and Measurement an Introduction*, (NewYork. Harcourt Brace Jovanovic, Inc.1983), p.6

⁵E.N. Gronlund, *Measurement and Evaluation in Teaching*, (USA : Mc. Millan Publishing Company, 1985). p.25

⁶M. Buchari M. Ed, *Tehnik-tehnik dalam Evaluasi Pendidikan*, (Bandung, 1980). P.119

b. Teacher Made Test

Test made by the teacher is a test that made by the teacher for assessing the student. This test especially in test result of learning and usually valid for one school, even sometimes for only one class.⁸

2. Based on Method of Teaching

Based on method of teaching, tests divided into norm-referenced test and criterion reference test. It is aimed to give axplanation about the purpose/meaning of the raaw score because it has not been interpreted.

a. Norm-Referenced Test

Norm-referenced test (NRTs) is design to rank pupil in order of achievement, from high to low, so that decision based on the relative achievement (e.g. selection, grouping, grading) can be made with greater confidence.⁹ It means NRT are made to compare test takers to each other. On an NRT driving test, test-takers would be compared as to who knew most or least about driving rules or who drove better or worse. Score would be reported as a percentage rank with half scoring above and half below the mid-point.

b. Criterion Reference Test

⁷ E.N. Gronlund, *Measurement and Evaluation in Teaching (fourth edition)*, New York: Macmillan Publishing Co., Inc. 1981, p. 322

⁸ Slameto, *Evaluasi Pendidikan*. Cet.2. Jakarta: Sinar Grafika Offset, 2001, p. 32

⁹ E.N. Gronlund, *Measurement and Evaluation in Teaching (fourth edition)*, 1981, p. 20

Criterion-referenced test is a group of items which directly measure behaviors that expressed in a set the target of behavioral or of performance objective¹⁰

3. Based on Its Function

Based on the function, the tests are divided into placement, diagnostic, formative and summative test. Its aimed to use and comprehend the test are not combine each other. In this study, the writer used summative test as the theory.

- a. *Placement Test* is a test to measure the students' basic ability and to place them based on their ability in a right level or suitable group.
- b. *Diagnostic Test* is designed to diagnose the students' difficulties and weaknesses during the teaching learning process.
- c. *Formative Test* determines learning progress, provide feedback for reinforcement of learning and correct learning errors.
- d. *Summative Test* is given at the end of a course or unit of instruction and the results are primarily for assigning grades or for certifying pupil mastery of the instructional objectives.¹¹ Another definition, summative test is intended to show the standard that the students have now reached in relation to other students at the same stage. Therefore it typically comes at the end of a course or unit of instruction.¹² It can be conclude that

¹⁰ M. NgalimPurwanto, *Prinsip-Prinsi pdan Teknik Evaluasi Pengajaran* .Cet. 9. Bandung: Remaja Rosdakarya, 2000, p. 27

¹¹ E.N. Gronlund, *Measurement and Evaluation in Teaching (fourth edition)*, New York: Macmillan Publishing Co., Inc. 1981, p. 21

¹²Drs. Wilmar Tinambunan, *Evaluation of students achievement*, (Jakarta:Depdikbud.,1998), p.7-9

summative evaluation is a method of judging the worth of a program at the end of the program active ties. In this study the summative evaluation is focused on the outcome of the student. So, the process of teaching learning is demanded to reach the goal of this evaluation. Whether or not the goal of this evaluation reached. Can be seen after doing the summative evaluation.

4. Based on the Form of the Test

There are two kinds of test based on the form of the test, they are written test and oral test. In this study the writer used written test.

a. Written Test

Written test is a test which are the items and the answers that have been given to the student is in the form of written language.¹³ It means the testers have to answer the test by writing the answers on the paper and it is called the students' answer sheet.

b. Oral Test

Oral test is a test where the questions are from the tester or teacher orally and the testers/students have to answer orally too. Another definition, oral test is testing ability to speak.¹⁴ This test is usually in the form of speaking.

¹³ M.C. Thoha, *Tehnik Evaluasi Pendidikan*, Jakarta: Raya Grafindo Persada, 2003, p. 54.

¹⁴ J.B. Heaton, *Writing English Language Test*, London: New Jersey, 1974, p. 83

5. Based on the Scoring System

Based on the scoring system, the test divided into subjective test and objective test. Based on my preliminary observation, objective test is used in SMK Kesehatan Borneo Bhakti Husada. The writer also interviewed the English teacher. From the interview, the writer knew how the summative test is made.

a. Subjective Test

Subjective test is one of the test kinds for the studying result that has characteristics as bellow:

- 1) The test formed question or command which need answer such as essay or explanation of sentence, generally is long enough.
- 2) The form of question or command demands to the tester for giving explanation, comment, interpretation, comparison, and differentiation and so on.
- 3) The total of the tests are limited about five to ten numbers.
- 4) Generally, the number of the item test is begin with the words: "Explanation...", "Clarify...", "Mention...", "Why...", "How...", or the other words which are similar.¹⁵

b. Objective Test

Objective test is a test which is highly structured and requires the pupils to supply a word or two or to select the correct answer from among

¹⁵Anas Sudijono, *Pengantar Evaluasi Pendidikan*, Jakarta: PT. Raja Grafindo Persada. 2005, p. 100

a limited number of alternatives.¹⁶ Another definition, objective test is a test which its item answerable by choosing the available answer. So, that participant presents the similarity of the data both of student who has true answer or the student who has false answer.¹⁷

From the explanation above, it can be concluded that the objective test is one kind of tests formed by multiple-choice and the students can choose the right answer to answer the item or question.

Some examples of objective test items (including multiple choice items) are:

1. Transformation

John Brown is a very fast runner

John Brown runs

2. Completion

The time is now half ten

3. Combination

Ann did her homework. Then she went swimming (after)

4. Addition (grammar)

YET Haven't you seen this film?

5. Rearrangement

at-poor-look-that-woman-old

6. Correct/incorrect (true/false)

Put a tick if the statement is correct and a cross if it is incorrect.

¹⁶ Wilmar Tinambunan, *Evaluation of Students' Achievement*, Jakarta: Departemen Pendidikan dan Kebudayaan, 1988, p. 55

¹⁷ M.C. Thoha, *Teknik Evaluasi*, p. 66

The train left to time this morning

He's doing some research on modern transport problem.

Although the examples have been confined to the testing of grammar, many of the item types can be used equally successfully to test vocabulary as well as aspects of the reading, writing, listening and speaking skills.¹⁸

7. Multiple-Choice Test

In certain ways, multiple-choice techniques are to testing the same as perspective is to drawing: only through a true appreciation and mastery of these techniques is the would-be test constructor able to throw aside the limitation imposed by them and discover newer and improved techniques of testing. The multiple-choice item is now widely regarded as being one of the most useful of all objective item types. Although it is amongst the most different of all objective item types to construct, it is simple to score and administer.

The following general principles should be observed when multiple-choice items are constructed.

1. Each multiple choice should have only one answer.
2. Only one feature at a time should be tested.
3. Each option should be grammatically correct when placed in the stem, except of course in the case of specific grammar test items.

¹⁸ JB. Heaton, *Language Testing*. London: New Jersey, 1987. P. 14

4. All multiple choice items should be at a level appropriate to the linguistic ability of the testees.
5. Multiple choice items should be as brief and as clear as possible (though it is often desirable to provide short contexts for grammar items).
6. In many tests, items are generally arranged in rough order of increasing difficulty.¹⁹

Generally, the alternative options of multiple-choice are (a), (b), (c), and (d). Just one of correct answer from the all alternative options, it calls the key answer. Other options are incorrect answer, it call the distracter option.

Each distracter, or incorrect option, should be reasonably attractive and plausible. It should appear right to any testee who is unsure of the correct option. Items should be constructed in such a way that students obtain the correct option by direct selection rather than by the elimination of obviously incorrect options.

For most purposes, each distracter should be grammatically correct when it stands by itself, otherwise testees will be exposed to incorrect forms. Unless a distracter is attractive to the student who is not sure of the correct answer, its inclusion in a test item is superfluous. Plausible distracters are best based on: mistakes in the students' own written work, their answers in previous tests, the

¹⁹ Rahmadi Nirwanto, *The Quality of Tests in the Textbook*. Palangka Raya: STAIN Palangka Raya. 2012. P. 40

teacher's experience, and a contrastive analysis between the native and target language.

The distracters should not be too difficult nor demand a higher proficiency in the language than the correct option. If they are too difficult, they will succeed only in distracting the good student, who will be led into considering the correct option too easy (and a trap).²⁰

One further step in the analysis of multiple choice items is to inspect the way each item distracter functioned. How well the alternative functions in a multiple choice item is also important in criterion referenced test.²¹

Determining the effectiveness of distracters is by comparing the students' numbers on upper and lower group who choose the wrong alternative distracters. If an item contains distracters which attracted no one, not even the poorest examines, it is a nonfunctioning choice which will increase the chances that some examines will get the item right by guessing between or among the remaining two or three possibilities.

The effectiveness of distracter defines by checking the frequency with which each distracter is selected by those failing an item.

²⁰ JB. Heaton, *Language Testing*. London: New Jersey, 1987. P.19

²¹ David P. Harris, *Testing English as Second Language*, New York: Mc Graw Hill Book Company, 1969., P. 256

D. The Functions of Test

The function of test in education cannot be discharged from the purpose of evaluation itself. The purpose of evaluation in education is to get the verification data that will showed until where the level of ability and student's success in attainment of curricular. Besides that, it also can be used by all teachers and education supervisor to measure or asses until where the effectiveness of teaching experience, learning activities and teaching methods that used.²²

There are two functions of test:

1. As a means of measuring educative participant. In this relation, a test has function to measure the progress reached by educative participant after the educative participants learned in selected time.
2. As a means to measures the learning program, because through the test, it can be known how far the instruction program have been achieved.²³

In other opinion, the functions of test are:

1. To measure the students performance in classroom.
2. To diagnose the students weaknesses and difficulties.
3. To evaluate the effectiveness of the syllabus as well as the methods and material used.
4. To provide the students with an opportunity to show their ability to recognize and produce correct form of language.²⁴

²²RiniYulia, "*Kemampuan Guru Menentukan Nilai Akhir Mata Pelajaran PAI di SMA Palangka Raya*", *Skripsi*, Palangka Raya: STAIN, 2006, p. 17.

²³Anas Sudijono, *Pengantar Evaluasi Pendidikan*, Jakarta: Raja Grafindo, 2007, p. 67.

²⁴J.B. Heaton, *Writing English Language Test*, London: Longman Group Ltd, 1979, p.2.

E. Criteria of Good Test

In this study, the writer used validity, reliability, index difficulty, and the distracter as the criteria of good test.

1. Validity

Validity of a test is the extent to which it measure what it is supposed to measure and nothing else.²⁵ The validity of a test must be considered in measurement in this case there must be seen whether the test used really measures what are supposed to measure, briefly. The validity of a test is the extent to which the test measures what it is intended to measure. There are four types validity:

a. Face validity

Face validity means the way the test looks to the testiest, teachers, moderators, and administrator. Therefore it is useful to show a test to colleagues or friends in order to discover absurdities and ambiguities of a test.

b. Content validity

Content validity is concerned with the materials that the students have learned. The test should cover samples of the teaching materials given. To fulfill this, the teacher should refer his consideration to the teaching syllabus. Another opinion, content validity depends on careful analysis of the language being tested and of

²⁵JB. Heaton, *Writing English Language test*, (Longman 1998), p. 153

the particular course objectives; the test should be so constructed as to contain a representative sample of the course.²⁶

c. Construct validity

Construct validity deals with construct and underlying theory of the language learning and testing. If the test has construct validity it is capable of measuring certain specific characteristics in accordance with a theory of language and behavior and learning.

d. Empirical validity

There are two kinds of empirical validity: Concurrent validity and predictive validity which depend on whether the test scores are correlated with subsequent or concurrent criterion measures. If we use a test of English as a second language to screen university applicants and then correlate test scores with grades made at the end of the first semester, we are attempting to determine predictive validity of the test. If on the other hand, we follow up the test immediately by having an English teacher rate each student's English proficiency on the basis of his class performance during the first week and correlate the two measures, we are seeking to establish the concurrent validity of the test.²⁷

2. Reliability

A test should be reliable as a measuring instrument. A test cannot measure anything well unless it measures consistently. Another opinion, a

²⁶ *Ibid.* P. 154

²⁷ *Ibid.*,

test cannot be valid unless it is reliable.²⁸ If the test administered to the same students on the different occasion and there is no difference to the results. It can be said that the test is reliable. In other opinion, the test is said can be trusted if it gives the consistent result when it is tested repeatedly.²⁹

3. Index Difficulty

Assumption that is used to get good item, besides has fulfill validity and reliability, it is also has to have the balancing of index difficulty. Few of index difficulty are based on student's ability in answering it, not from the side of the test maker.

Determining item difficulty on a test related to the numbers of students who answer the item correctly. A good test has to be proportional item, which means in a test has to have balancing within easy item, enough and the difficult. The good item consists of a fair number of difficult, fair and easy items. A very easy test will not stimulate students to enrich his effort to solve it. While a test that too difficult will make students hopeless and there no passion to try again because it out of their rich.

Index difficulty item is about 0,00 – 1,00. It means the lowest index difficulty item is 0,00 and the highest is 1,00. Index difficulty 0,00 (P = 0,00) is the guidance for the tester that the item conclude to difficult item. Because, this case the tester cannot answer the question correctly. Index

²⁸J. Charles Anderson, Caroline Clapham and Dianne Wall, *Language Test Construction and Evaluation*, (British: Cambridge University Press, 1995), p.18

²⁹Suharsimi Arikunto, *Dasar-Dasar Evaluasi Pendidikan*, Jakarta: Bumi Aksara, 1999, P. 60

difficulty 1,00 ($P = 1,00$) is the guidance that the item too easy. Because of all the testers can answer correctly.³⁰

F. The Principles of Making Multiple Choice Test Item

Seem like making instrument any items, the procedure of making multiple choice test item started by making latticework of test items. The latticework of test item is the guidance to make test items will not stump from the purpose of test making. The lattice works are:

1. Represent the curriculum.
2. The components are detail, explicit and easy to understand.
3. The items can be made appropriately with indicator and form of items that have been decided.

According to Department of National Education by the Center of Education Assessment, there can be found sixteen principles of writing multiple choice test. The principles are:

1. The item must be appropriate with indicator.
2. The option of answer must be homogeneous and logic.
3. Every item must have a right answer.
4. The main item must be formulated clearly and explicitly.
5. The main formulation of item and option of answer must be in statement form which is needed.
6. The main items may not give guidance to the right answer.

³⁰ Witherington in Anas Sudijono, *Pengantar Evaluasi Pendidikan*, Jakarta: PT. Raya Grafindo Persada, 2007, p. 371

7. The main items may not contain statement that has double negative characteristic.
8. The option of answer formulation must have the same quantity.
9. The choice of answer may not contain statement of “all answers above are right/wrong.”
10. The choice of answer in form of number or time must be arranged based on big-small series value of the number.
11. Picture, graph, table, chart and one kind which are found in the item must be clear and useful.
12. Test item may not depend on the previous choice of answer.
13. Every test item must use an appropriate language based on language principles used in the test.
14. Every test item must use communicative language.
15. May not use the local language if the test items will be used for other places or for national use.
16. Choice of answer may not repeat the word/phrase that is not a meaning unity.³¹

³¹ Muhammad Karwapi, *Kaidah Penulisan Butir Soal Bentuk Pilihan Ganda*, Wordpress.com, accessed on November 08, 2012