

CHAPTER II

REVIEW OF RELATED LITERATURE

This chapter presents a review of theories related to the research topics. It is including the meaning of the test, the type of the test based on the way of arrangement, the criteria of good test, and the rules of students' school test scoring.

A. The Meaning of Test

The nature of a test is a series of questions by the evaluator or tester verbally and in writing that must be answered by the participants or testee. The answer given is either true or false. The test is a systematic procedure to observe or describe one or more characteristics of a person by using a standard numeric or system category. A systematic test is procedural in nature, meaning that it should be a step from planning and implementation to reports and follow-ups.

Sheeba in his article mentioned that there are differences between test, evaluation, assessment, and measurements. A test or quiz is used to test someone's knowledge of something to determine what he or she knows or has learned. Testing measures the level of skill or knowledge that has been searched.⁹

From evaluation, Sheeba said that it means the process of making judgments based on criteria and evidence.¹⁰ When a teacher has completely made a set of plannings, then they are done in the teaching and learning process, and some testings and assessment are conducted, at last an evaluation of the teaching and learning activities is made. The Teacher collects many evidences of the students' achievement such as the results of assignments, students' portfolios, and other observations and results the teacher recorded. Thus, the evaluation is at the final step. Foyewa also stated that evaluation is one of the ways through which

⁹ Sheeba, "Importance of Testing And Evaluation in Teaching and Learning", *International Journal of Society and Humanities*, Vol. 11 No. 1 (2017).

¹⁰ Ibid.

feedback can be obtained from the learners on what the teachers had taught them.¹¹

Assessment is the process of documenting knowledge, skills, attitudes and beliefs, usually in measurable terms. The goal of assessment is to make improvements, as opposed to simply judgment. In an educational context, a teacher makes an assessment as the process of describing, recording, scoring, collecting, and interpreting all information about students' learning. In Indonesia, the government has made a recent system. That is an authentic assessment system in the 2013 curriculum. According to Gulikers, Bastiaens and Kirschner, an authentic assessment is an assessment requiring students to use the same competencies, or combinations of knowledge, skills, and attitudes that they need to apply in the criterion situation in professional life.¹²

Measurement is beyond its general definition. It refers to a set of procedures and principles of how to use the procedures in educational tests and assessments. The basic measurement in evaluations is in the form of raw scores, standard scores, percentile ranks, and class rank.

Thus, based on those differences, it is clearly stated that the meaning of a test has grown from the need for evaluation in the learning process. To know the students' improvement, the teachers carry out the test. Then they continue to do measurements, assessments, finally they end in the final step called evaluation.

B. The Type of a Test Based on the Way of Arrangement

A test has its difference in the forms. Based on Djiwandono, it can be divided into standardized and teacher-made tests. Standardized tests are compiled based on the requirements and strict procedures to produce a good test in accordance with what has been planned. Meanwhile the teacher-made tests are the

¹¹ Foyewa, "Testing And Evaluation In English Language Teaching – A Case Of O Level English In Nigeria", *International Journal of English Language Teaching* Vol.3, No.6, pp.32-40, September 2015.

¹² J. Gulikers et al. "The First International Conference on Enhancing Teaching and Learning Through Assessment", Polytechnic University, Hong Kong, 2005.

tests prepared by the teacher as an evaluation of the implementation of the daily tasks. They have been in progress, including evaluating students' progress.¹³

1. The Standardized Test

According to Supardi, this test is developed by a team of experts, or prepared by a special institution that organizes a professional test. A number of requirements and a long process are needed in making this test. The stages include making a blue print, some trial phases, revision, and studying the characteristics of a good test. An analysis of a test such as the level of difficulty, discrimination, validity and reliability both rational and empirical validity, reliability in the sense of stability and homogeneity is also required.¹⁴

According to Djiwandono, the blueprint contains important aspects and dimensions of the underlying basis in the initial start-making test. There should be described the steps that must be taken namely test planning, test writing, critical a review of test, test trials, test revision, the final form of test.¹⁵

From these explanations, a standardized test may be used not only for test participants but also possibly for many people. As well as in the field of English language, widely known as TOEFL (Test of English as a Foreign Language), it is managed and developed by ETS (Educational Testing Service) located in Princeton, New Jersey, United States. Its popular product is iBT TOEFL (TOEFL Internet-Based Test) via the Internet. In Indonesia, TOEFL becomes entrance test in university. Besides, there is also an IELTS (The International English Language Testing System) coming from Cambridge University, England, although there are differences in the basics and the details of implementation. TOEIC (Test of English for International Communication) and GEPT (General English Proficiency Test) also include to standardized test. Such James M. Sims who stated that in Taiwan, university students will be tested their English

¹³ M.S. Djiwandono, *Tes Bahasa: Pegangan Bagi Pengajar Bahasa* (Jakarta: Indeks, 2011).

¹⁴ Supardi, *Penulisan Autentik Pembelajaran Afektif, Kognitif, dan Psikomotor Konsep dan Aplikasi* (Jakarta: PT. Raja Grafindo, 2015), 9.

¹⁵ M.S. Djiwandono, *Tes Bahasa: Pegangan Bagi Pengajar Bahasa* (Jakarta: Indeks, 2011).

proficiency which the form of test is TOEIC or GEPT in order to be graduated.¹⁶ According to Arikunto, the uses of standard assessment instrument items are as follows:¹⁷

- a. Comparing learning achievement with individual or group
- b. Comparing the level of student skills in various fields of study for individual or groups
- c. Comparing the students' achievement of various schools or classes
- d. Studying the development of students in a particular period

2. The Teacher-Made Test

Compared with the previous standard tests, teacher-made tests are much simpler in terms of its implementation procedure. All tests are made by teachers themselves where the teachers are assumed to already have the knowledge about how to make a good test. In making the test, teachers must also understand the condition of the ability of students, the level of achievement, their advantages and disadvantages, and the appropriateness with the materials they have got. With knowledge of the content of the teaching materials, teachers may arrange tests compatible with the teaching materials that have been presented in teaching and learning activities.

The type of generally made by teachers is midterm test and final test. The obligation of making midterm and final tests is also accompanied by others for instances, daily assignments, homework, short tests / quiz, even more unlimited other tests. According Supardi, the uses of a test or questions made by the teacher are as follows:¹⁸

¹⁶ James M. Sims, "A Valid and Reliable English Proficiency Test: A model from a University Language Program in Taiwan", *English as a Global Language Education (EaGLE) Journal*, Vol. 1 No.2 (2015) 91-125.

© Foreign Language Center, National Cheng Kung University & Airiti Press Inc.
DOI:10.6294/EaGLE.2015

¹⁷ Suharsimi Arikunto, *Dasar-dasar Evaluasi Pendidikan* (Jakarta: Bumi Aksara, 2013).

¹⁸ Supardi, *Penulisan Autentik Pembelajaran Afektif, Kognitif, dan Psikomotor Konsep dan Aplikasi* (Jakarta: PT. Raja Grafindo, 2015), 80.

- a. To determine how well students have mastered the materials given within a certain time.
- b. To determine the position of students in a class or a group
- c. To obtain a value.

C. Criteria of Good Test

A test can be categorized as a good test from the quality from its analysis results, which are categorized as quantitative and qualitative analyses. For the former, a quantitative analysis covers reliability, index of difficulty levels, and discrimination power. This quantitative analysis is divided into two points, namely test and item analyses. For the latter, a qualitative design includes as content validity, material direction, construction direction, and language direction.

1. Reliability

In research, reliability is a consistent measurement after being performed repeatedly to the same subject and in the same condition. The instrument is reliable when it produces consistent results. A test is said to be reliable if the score obtained by the test taker is relatively similar despite being done repeatedly, according to Rasyid.¹⁹

The types of reliability are various, depending on the model implementation of that test. The types of reliability are test-retest reliability, equivalent-forms reliability, split-half reliability, Cronbach Alpha reliability, and Kuder-Richardson reliability either KR 20 or KR-21. The following is the review of the varieties of reliability.

a. Test-retest reliability

This method can be used when the test is repeated for the same number of participants as mentioned by Djiwandono.²⁰ Test participants get two same tests, namely pre- and post-tests. Repetition is carried out after a certain period of limit,

¹⁹ Fathor Rasyid, *Metodologi Penelitian Sosial Teori dan Praktik* (Kediri: STAIN Kediri Press, 2015).

²⁰ M.S. Djiwandono, *Tes Bahasa: Pegangan Bagi Pengajar Bahasa* (Jakarta: Indeks, 2011).

so that participants do the same test without relying on their memory. Therefore, the intervals between the tests are not supposed to be too short or too long time, as it may give impacts on the amount of the improving scores during the learning process.

Results obtained are in the form of two scores from the implementation of the pre- and post-tests. Both series of tests are calculated in terms of correlation using the Pearson product-moment. The Manual formula is as follows according Sugiyono.²¹

$$r - xy = \frac{\Sigma (X - X1) (Y - Y1)}{N SxSy}$$

Note:

r-xy: Correlation Pearson product-moment

Σ : Total

X: each student's scores on tests of X

Y: each student's score on the test Y

X1: The average score of the test X

Y1: The average score of the test Y

Sx: standard deviation of the test X

Sy: the standard deviation of the test Y

N: Number of participants to the two tests

b. Equivalent-forms reliability

This method has the requirement namely the availability of two similar tests and they are presented to a group of participants using the same test. What is meant by similar test is the same test in various substantive aspects except the test item. The similarity is on the principal aspects of the test, the type of test, the content, the number of test items, even test item sequences and the level of difficulty.

The second test can be held shortly after the completion of the first test form. All scores are collected and calculated using the formula of the degree of correlation with the Pearson product-moment.

²¹ Sugiyono, *Metode Penelitian Kuantitatif, Kualitatif, dan R&D* (Bandung: Alfabeta, 2017), 354.

c. Split-half reliability

In the split-half reliability calculations, tests are held once without repetition. After the scores of all participants are calculated, then the test-maker divides the scoring list into two parts with the same number of test items. If a test is tested twice, a split half method merely requires a test only once. It is divided into two: odd and even items. While, Arikunto mentioned that the test can also been divided into two namely the previous and last items.²² The formula is as follows.

$$r - xy = \frac{nx r}{(N - 1) r + 1}$$

Note:

r - xy: reliability tests complete

r: correlation coefficient between the two parts of the test, the results of the Pearson product-moment formula

n: adjustments that need to be done, namely the number of parts of the test

d. Cronbach Alpha reliability

As well as Split-half reliability, researchers can also apply a variant calculation namely Cronbach Alpha formula. This method requires two equal parts of a test, but the calculation is taken only in the form of average scores, the standard deviation of the first and second parts, and the standard deviation of the entire test.

$$\alpha = \left(1 - \frac{S_1^2 + S_2^2}{S_s^2} \right)$$

Note:

α : Reliability Cronbach Alpha entire test (coefficient alpha)

S_1^2 : The standard deviation of the test piece to-1

S_2^2 : The standard deviation of the test part 2

S_s^2 : The standard deviation of the entire test

... .2: squared

²² Suharsimi Arikunto, *Dasar-dasar Evaluasi Pendidikan* (Jakarta: Bumi Aksara, 2013)..

e. Kuder-Richardson reliability either KR 20 or KR-21

Calculation of reliability in this method is only for one-time implementation of the test. The answers to the test items are totaled. True is represented by 1, false by 0. This method has two types: the KR 20 or KR-21.

The KR-20 type method is the first and it generates a higher level of reliability and is accurate although it is rather complicated. Therefore, the development of the KR-21 formula is much simpler. Based on Sugiyono, here is how to calculate the type of KR-21 formula.²³

$$K - R21 = \frac{k}{k-1} \left(1 - \frac{X(k-X)}{kS^2} \right)$$

Notes:

k: the number of items

X: the average score

S: standard deviation

2. Index of Difficulty

A good test item is that the assessment is not too easy and not too hard. Students are trained in taking a test with a balanced level of difficulty. If it is too easy then the students more easily do it without any stimuli to think critically. Instead, if the test item that is too hard it will cause students to be hopeless and they do not have any spirit to try again because it is out of the material scope.

An index of difficulty shows whether a question is difficult or easy. The value of index of difficulty is from 0.00 to 1.0. If it appears that the value is 0.00, it means that the test is too difficult. If the value is 1.0, the test is too easy. To calculate the difficulty of test items, the formula from Rohmad is adopted.²⁴ The items of instrument considered to be good is where the level of difficulty is in the middle level (the difficulty index is between 0.30 up to 0.70).

²³ Sugiyono, *Metode Penelitian Kuantitatif, Kualitatif, dan R&D* (Bandung: Alfabeta, 2017), 361.

²⁴ Rohmad, *Pengembangan Instrumen Evaluasi dan Penelitian* (Yogyakarta: Kalimedia, 2017), 247.

$$P = \frac{JSB}{JS}$$

Note:

P: Index of Item Difficulty

JSB: The number of students who answered correctly

JS: The total number of student participants test

3. Discrimination Power

Discrimination power according to Rohmad is the ability to distinguish between students with good or high scores (upper group) and low ones (lower group).²⁵ Here is how to find the discrimination index formula based on Rohmad.²⁶

$$D = \frac{BT}{JT} - \frac{BR}{JR}$$

Note:

D = Item discrimination index

BT = Total of upper group answer correctly

JT = Total of upper group

BR = Total of lower group answer correctly

JR = Total of lower group

The testees (test takers) are divided into upper and lower groups. This division is based on the result of right answers given by testees to whole test. Testees are sorted from the total right answers in the higher to the total right answer in the lower. If the number of all testees is less than 100, the grouping can be carried out by separating all testees into two groups (each group is 50%). If the number of testees is more than 100, the upper group can be taken from 27% upper testees (the rank). The lower group is taken from 27% lower testees (the rank), therefore, each group represents upper and lower groups, respectively.²⁷

The criteria of a good test can be identified through a qualitative analysis. A qualitative analysis of the instrument items is divided into moderator and panel techniques. A moderator technique is the discussion technique, where there is one

²⁵ Ibid., 249.

²⁶ Ibid., 250.

²⁷ Ibid., 251.

person as an intermediary. The experts can be only such teachers who teach the materials, linguists, curriculum makers, or assessment experts to discuss together the test analyzed.

4. Validity

Validity according to Grondlund is simply associated with the suitability test as a measurement tool with the main targets that need to be measured.²⁸ The validity of the measuring instrument is not solely related to the position of the measuring instrument as a tool, but especially on the suitability of the results, in accordance with the objectives of the measuring instrument.²⁹ For example, if a teacher as a test-maker intends to measure students' writing skills, the teacher should focus his/her measurement on the ability to express themselves in writing, rather than speaking. The suitability and relevance of a test with the ability of test-takers is what is meant by validity.

Basically, the validity is not manifold, but that it varies in accordance with the way of proving existence of validity. Its existence can be identified through the study of content validity, construct validity, and criterion validity.

a. Content validity

To prove content validity can be done by arranging grids compared with tasks required in doing a test and test items or the curriculum contents.

b. Construct validity

Validity refers to the concordance between the results of the measuring instrument and the ability to measure. Rasyid stated that construct validity can be done by identifying and pairing items about the intended goals which means to reveal the extent of the cognitive aspect. Like content validity, construct validity that determines the level should also be done by basing itself on a lattice of measuring instruments.³⁰

²⁸ Norman E. Grondlund, *Measurement and Evaluation in Teaching* (New York: Macmillan Publishing Company, 1985).

²⁹ J.H. McMillan, *Educational Research: Fundamentals for the Customer* (Boston: Pearson Education, 2012).

³⁰ Fathor Rasyid, *Metodologi Penelitian Sosial Teori dan Praktik* (Kediri: STAIN Kediri Press, 2015).

c. Concurrent validity

The validity of test can be proved by connecting validity or linking test scores which are assessed with a score of similar tests and characteristics. This is carried out almost at same time.

Thus, a researcher uses content validity in order to know how the content of the test is appropriate with the purpose of English try-out test by using national examination (grids). The form is presented in the Tab 2.1. below based on Dea Cindrakasih as stated in the previous study.³¹

Table 2.1. The Content Validity in Standard Competence

No.	Basic Competence	Related Items	Total	Percentage

5. The Qualitative Panel Technique

A qualitative panel technique is a means to analyze a form of tables and checklists for an essay test form, and a multiple choice test form based on Abdul Ghofur and Djemari Mardapi, where it contains materials, construction, language / culture, and key answers / guidelines scores. The preparation needed is the questions making the grid, the reference curriculum / syllabus, the source book, the assessment guidelines, and English dictionary.³²

A qualitative analysis emphasizes the point where a test item instrument can be in a narrative form, completion, multiple choice, assessment of affective, and psychomotor domains. Particularly in this study, the researcher chose a multiple choice test. In the rules stated by Supardi, the analysis should be carried out

³¹ Dea DB Cindrakasih, "Analysis of English Try Out Test of The Ninth Grade Students at Mtsn Probolinggo", *Journal Ilmiah Edukasi & Sosial*, Vol. 9 No.1 (2018), 28-37. URL: <http://jiesjournal.com/index.php/jies/article/view/122>, 31.

³² Abdul Ghofur and Djemari Mardapi, *Pedoman Pengembangan Penilaian* (Jakarta: Departemen Pendidikan Nasional, 2004), 91-92.

before the items were tested and the results must be correct and precise.³³ In short, a test may be qualified as a good instrument if it is provided with strong supporting qualitative analysis. Three points are in the item instrument, namely material direction, constructive direction, and language directions. Dealing with the form, it can be seen in Chapter Three.

D. Rules of Students' School Test Scoring

In the pandemic year 2020, the implementation of test for the third students is different. The final score is different from that of the previous year that used a national examination held by the central government. According to the Minister of Education and Cultures of the Republic of Indonesia, Nadiem Makarim, published e-official letter about an announcement letter containing the rules of students' school test below:³⁴

1. The school test for graduation in the form of a test that gathers students must not be performed except the one that has been performed before this released announcement letter.
2. The school test may be carried out in the scores from transcript portfolio and achievement previously obtained, the tasks, online tests, and/or other distant assessments.
3. The school test is compiled to encourage meaningful learning activities, and not necessary to measure overall curriculum.
4. Schools that have held school test may use the school tests to determine students' graduation. For schools that have not held school tests, they should obey several rules as follow:
 - a. Elementary graduation is determined based on last five semesters (the fourth grade, the fifth grade, and the sixth grade odd semester). The

³³ Supardi, *Penulisan Autentik Pembelajaran Afektif, Kognitif, dan Psikomotor Konsep dan Aplikasi* (Jakarta: PT. Raja Grafindo, 2015).

³⁴ Surat Edaran Nomor 4 Tahun 2020 Tentang Pelaksanaan Kebijakan Pendidikan Dalam Masa Darurat Penyebaran Corona Virus Disease (COVID- 1 9), Jakarta: Pusdiklat Pegawai Kementerian Pendidikan dan Kebudayaan, 2020.

scores of the odd semester in the sixth grade may be used as the addition of graduation scores.

- b. Junior High School and Senior High School graduation is determined based on scores of the last five semesters. Even semester scores of the ninth and the twelfth grades may be used as addition of graduation scores.
- c. Vocational School graduation is determined based on the transcript, practices, portfolios, and practice scores during last five semesters. Even semesters scores in the previous year may be used as the addition of graduation scores.