

## CHAPTER II

### LITERATUR REVIEW

#### A. Evaluation

##### 1. Definition of Evaluation

There are several definition of Evaluation. Commonly, evaluation is defined as systematic acquisition and assessment of information to provide useful feedback about some object. These are several definions about evaluation are delivered by experts, as follows:

- a. Evaluation is a process that is deliberately planned to get information or data, based on the data and then try to make a decision. (Purwanto, 1992:3)
- b. Evaluation is a going process that encompasses a much wider domain. (Douglas, 2004:4)
- c. Evaluation is assessing (but done by measuring first). It means that evaluation is an activity to give judgement about certain result that should be done after measuring the result. (Arikunto, 2006:3)

In this research, evaluation is understood as a process to determine how far the students can absorb teaching's material. In other side, evaluation also give feedback especially to the teachers related their teaching method. Whether it is good enough or needs improvement.

##### 2. Kind of Evaluation

There are some kinds of evaluation in the sector of school's education which covers 3 components, as follows:

a. Evaluation toward teaching's program

Evaluation toward teaching's program is series of actions which are done to find out how well the success of the program is. According to Cronbach and Stufflebeam in Sudaryono (2012:41), evaluation toward teaching program is an effort to give an information which will delivered to someone who will make the decision. Thus, the final of evaluation should be good in order to give the accurate result to the student's parent.

Evaluation toward teaching's program is also aimed at finding out the target of achievement. For determining how far the target of the program can be achieved, it has correlation with the purposes which has been programed. For example, the teacher has predicted that there will be more than eight students who get score 10 in the test. However, in fact there are only 3 students who get 10, so the level of teacher's achievement is only  $\frac{3}{8} \times 100\%$ , that is 37%.

b. Evaluation toward process of teaching's implementation

Evaluation toward process of teaching's implementation covers:

- 1) The appropriateness between learning process which is going on and the teaching program which has been decided.
- 2) Teachers' readiness to implement teaching program, and the readiness of the students to follow teaching learning process.
- 3) Students' interest and attention to follow teaching learning process.

- 4) The role of guidance for the students.
- 5) The communication from both teacher and students during teaching learning process.
- 6) Giving motivation to the students.
- 7) Giving the task to the students.
- 8) An effort to decrease the negative effect which is created from school's program.

c. Evaluation toward teaching result

Evaluation teaching's result covers:

- a) The evaluation toward students' competence in understanding the school objective which is mentioned in the school's program.
- b) The evaluation toward students' achievement in understanding the general purpose of learning.

3. The Principles of Evaluation

Ir. Sudaryono states that there are some principles which should be watched while doing the evaluation. Ir. Sudaryono states that there are 7 principle. The 7 principles are as follows:

a. Continuity

This principle means that a good result evaluation is when evaluation itself can be done continuously. The teachers must give evaluation every time the teacher finish giving material to the students in order to get a good conclusion.

b. Comprehensive

Evaluation which is done by the teacher should cover students' attitude, cognitive domain, affective domain, and psychomotor of the student.

c. Objectivity

Instrument in evaluation should be free from subjectivity or personal bias of the teacher.

d. Validity

Validity means the instruments in evaluation can be used to measure the aspect which will be measured. For example, if the test is aimed to describe students' achievement, the teacher should like to be able to interpret the scores as relevant and representative sample of domain to be measured. If the results are to be used to predict students' success in some future activities, the teacher should like their interpretations to base on as accurate an estimate of future success as possible.

e. Reliability

Reliability means that evaluation or measurement can be used to measure the students' competence which is given in different time and it is free from bias or error. In the end, the result of evaluation which is done shows the accuracy of the result.

f. Criteria

Criteria used in evaluation are absolute standard (acuanpatokan) and relative standard (acuan normal).

g. Useful

An evaluation which is done should be useful for both students and teacher.

4. The Purpose of Evaluation

There are several purposes why evaluation should be done by the teacher. At least, there are six purposes of evaluation related to teaching learning process.

The six purposes are as follows:

1. To assess the achievement (attainment) related to the objective of the study.

There is a relation between the purpose of the study, method of evaluation and the way of how students master the material. How teacher evaluate the students will determine the way or the method how students master the material. On the contrary, the purpose of evaluation will determine the method of evaluation which is used by the teacher.

2. To measure various aspects of learning.

Aspects of learning can be divided as cognitive, psychomotor, and affective. All the aspects of study should be evaluated in appropriate proposition.

3. As the medium (means) to find out how well students understand the material.

Because the students may have different background, they also may understand the material in different way. Of course, the teacher have to

know which student who already understand about the material and others who don't understand the material well.

4. To motivate the students.

Evaluation also can be used to motivate students as well. the result of evaluation will stimulate students' action, and if the result of evaluation is good it will create spirit or encouragement to maintain or increase students' motivation.

5. To prepare the information for guidance and counseling.

Information about students' problem like ability, quality, social adaption, reading ability, and the result of the study will be very important thing to guide students' carrier. Determine students' interest is the best way to help the students, especially for their future.

6. To change the result of evaluation as the base of change curriculum.

There is a relation between evaluation and instruction. It is because evaluation is part of instruction. Besides that, between curriculum and instruction also have relation because instruction can be used as one of the important parts of curriculum.

## **B. Test**

### 1. Definition of Test

Test is one of the instruments which is used to do evaluation. Test is used to measure how well the student has understood the lesson or the material which is delivered by the teacher. The definition of the test according to Linn and Gronlund in Sudaryono is an instrument of systematic procedure for measuring a sample of behavior by posing a set of questions to the students, because a test is a form of assessment, test

also answer the questions how well the individual performs either in comparison with others or in comparison with domain of performance task (Sudaryono,2012:101). Generally, teacher uses a test to improve the result of teaching. Through test, teacher is able to know the information about students' level in mastering the lesson. Besides that, test is also used as a form which contains students' achievement during teaching learning process.

## 2. Form of Test

The items used in classroom tests are typically divided into two general categories, there are;

### a. Objective Test Item

Objective test item is highly structured and requires the students to supply a word or two or to select the correct answer from a number of alternatives. The objective test item includes a variety of different types, those are:

#### a) Short answer

Short answer questions are open-ended questions that require students to create an answer. Short answer questions are commonly used in examination to assess the basic knowledge and competence (low cognitive levels) of a topic before more in-depth assessment question are asked on the topic.

Example of short answer:

#### **Announcement**

The school drama club is going to hold the "Damarwulan" casts audition. For you who are interested to join the audition, please register yourselves in our office on Friday, March 14th 2014 at 1 p.m.

The requirement:

The latest photographs of postcard size (2)

The audition will be held on Saturday, March 15th 2014 at 9 a.m. The list of the selected candidates will be announced in a week.

We do appreciate your participation, Thank you

Notes: no charge for the registration.

The leader of casts audition,  
Mr.Arif Hertanto

Who is the announcement for ?

b) Completion

The completion item requires the student to answer a question or to finish an incomplete statement by filling in blank with the correct word or phrase

Example of completion:

Water will froze up at temperature ..... Degree of Celsius.

Water will froze at temperature 100 degree of Celsius.

c) True-False or Alternative Response

True-False test item offers a series of statements each of which is to be judged as true-false.

Example of true-false;

T                      F        He not a lazy boy.

T                      F        We are all Indonesians.

## d) Matching

Matching is an objective test consisting of two sets of items to be matched with each other for a specified attribute.

Example of matching;

1. Quickly	1. Adjective
2. Jump	2. Adverb
3. Dog	3. Conjunction
4. And	4. Noun
5. She	5. Pronoun

## e) Multiple choice

The multiple choice item consists of two parts: (a) the stem, which identifies the question or problem and (b) the response alternative. Students are asked to select the one alternative that best completes the statement or answers the question.

Example of multiple choice;

In which of the following sentences do the subject and verb disagree?

- a. When they win, they are happy.
- b. Politics are hard to understand.
- c. The majority is always right.
- d. One or the other is to be elected.

### b. Subjective Test Item

Subjective test item permits the students to select, organize, and present the answer in essay form. Subjective test item is also usually called as essay question. It is commonly viewed as a single item type. A classification of essay question, those are;

#### a) Extended response type

Extended response item give almost complete freedom to the students in making the answer or response. It permits the students to decide which fact they think are most pertinent, to select the students own method of organization, and to write as much as seems necessary for a comprehensive answer.

#### b) Restricted response type

Restricted response type tend to limit the nature, length, or organization of students' answer or response.

### 3. Kind of Test

Generally in teaching education there are four kinds of test those are; Placement test, Diagnostic test, Summative test, Formative test. The explanation of each kinds of test as follow:

#### a. Placement test

Placement test is concerned with the students' entry performance and typically focuses on questions such as the following:

- 1) Does the student possess the knowledge and skills needed to begin the planned instruction?

- 2) To what extent has the student already mastered the objectives of the planned instruction?
- 3) To what extent do the pupils' interest, work habits, and personality characteristic indicate that one mode of instruction might be better than another?

The goal of placement test is to determine the position in the instructional sequence and the mode of instruction that is most beneficial for each student.

b. Formative test

Formative test is used to monitor learning process during instruction. This is to provide continuous feedback to both the students and teacher concerning on learning successes and failure. Formative test depend heavily on specially prepared test for each segment of instruction (e. g. , unit, chapter). These are usually mastery tests that provide direct measures of the intended learning outcomes of the segment.

c. Diagnostic test

Diagnostic test is used to identify students' strength and weaknesses. It is intended primarily to ascertain what further teaching's necessary. The main aim of diagnostic test is to determine the causes of persistent learning problems and to formulate remedial action.

d. Summative test

Summative test is design to determine the extent to which the instructional objectives have been achieved and is used primarily for assigning course grades or for certifying students' mastery of the intended learning outcomes. Summative test typically comes at the end of a course (or unit) of instruction.

4. The function of Test in Teaching

A test is a very important aspect in teaching. Many experts are interested in explaining the function of the test. Heaton states that, a test which set out to measure students' performance as fairly as possible without any way setting traps for him can be effectively used to motivated students. A well-constructed classroom test will provide the students with an opportunity to show his ability to recognize and produce any forms of the language. Provide that details of his performance or given as soon as possible after the test, the student should be able to learn from his errors and consolidate the pattern taught. In this way a good test can be valuable device. From the above theories, so test can be function as motivator, controller or indicator of teaching learning process. That's way the existence of the test is needed to support the students achievement. Test also has function as a feedback, as a motivation tool, as diagnose and remedy. These are the explaining of each function of the test:

a. Test as Feedback

Test can be used as feedback for students and teachers. For students, the results of the test show their learning improvement. If someone gets good mark, he will continue his learning program and in another side if a student get bad mark, he has to repeat his learning. For the teacher, the test helps them to provide more appropriate instructional guidance for individual students as well as for the class as a whole.

b. Test a Motivation Tool

Test given motivation to students to learn. Students who get low scores should be made it as whips for them study hard in the next test. For the students who get high scores, test can motivate them to maintain and to improve their achievement.

c. Test a Diagnose and Remedy

Test in here is used to identify the students' difficulties. Teacher know his students' difficulties through the items that the students' cant' do.

5. Organizing Test

It is necessary to make a good planning before designing the test. There are some principles that the teacher must notice before designing the test. Sukewi states that;

- a. Test that is designed should measure students' ability that has correlation with teacher's purposes in the instructional objective.

- b. Test is to measure the representative sample or students' achievement and teaching material that has been taught.
- c. It is designed to get the beneficial of the result.

#### 6. Criteria of a Good Test

The characteristic of a good test are, as follows;

##### a. Validity

When constructing or selecting test and other evaluation instruments, the most important question is to what extent will the interpretation of the scores be appropriate, meaningful, and useful for intended application of the result? Validity refers to the appropriateness of the interpretations made from test scores and other evaluation result, with regard to a particular use.

According to Nurgiyantoro (2001:103), the validity of a test is the extent to which it measures what it is supposed to measure and nothing else. There are four types validity:

##### 1. Face validity

Face validity means the way the test looks to the testiest, teachers, moderators, and administrator. It can be seen from the outfit prepared for the testiest. For example, for listening exam, the equipment must be loudspeaker, radio or cassette, and for reading exam, the equipment must be paper and so on.

##### 2. 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 validity, teachers should refer his consideration to teaching syllabus. All the test item or questions in the test must represent all the materials which have been given by the teachers.

### 3. 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.

### 4. 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. In this case of concurrent validity the relationship between a test and a currently available criterion is assessed, while in the case of predictive validity the criterion does not become available until a later date. Empirical validity is usually expressed in the form of correlation coefficient, but sometimes a test of significance between the scores of criterion groups is reported instead.

In this research, the writer will analyze the validity of Final test in the first grade students of SMA N 1 PurworejoKlampok in academic year 2014/2015 through content validity.

## b. Reliability

Reliability is the most important characteristic of evaluation result. Reliability provides the consistency that makes validity possible and indicates how much confidence the teacher can place in students' result. Reliability refers to the consistency of measurements that is how consistent test scores or other evaluation results are from measurement to another.

According to the American Psychological Association, Standards for Educational and Psychological Testing (Washington D. C. APA) there are several method in determining the reliability of the test.

### 1. Test-Retest Method

To estimate reliability by using test-retest method is by measuring its stability. The teacher gives the same test twice to the same groups with anytime interval between tests, from several minutes to several years. The result of the test then will be correlated and the correlation coefficient provides a measure of stability that indicates how stable the test result are over the given period of time.

The longer time interval is between test and retest, the more the result will be influenced by changes in the students characteristic being measured and the smaller the reliability coefficient will be.

## 2. Equivalent-Forms Method

Estimating reliability by using equivalent – form method uses two different but equivalent form of the test. The teacher gives two forms of the test to the same group in closed succession then the results are correlated. The correlation indicates the degree to which both forms of the test are measuring the same aspects of behavior.

## 3. Split-Half Method

Reliability can also be estimated from a single administration of a single form of a test. The teacher gives test once, then the result is divided in half for scoring purposes. The usual procedure is scoring the even-numbered and the odd-numbered items separately. The procedure is done for each student, which, when it correlated, indicates the degree to which consistent results are obtained from the two halves of the test.

The preceding reliability coefficient is determined by correlating the scores of two half-test. To estimate the scores' reliability based on the full-length test, the Spearman- Brown formula that is usually applied :

Reliability on full test;

2 times correlation between half-test

1 plus correlation between half-test

#### 4. Kuder- Richardson Method

Estimating reliability by using Kuder- Richardson method is similar with split half method, that these formulas provide a measure of internal consistency. However, in Kuder – Richardson method, it does not require splitting the test in half for scoring purposes.

In this research, the write will analyze the reliability of Final test in the first grade students of SMA N 1 PurworejoKlampok in academic 2014/2015 by using split half method.

##### c. Practicality

The third characteristic of a good test is practicality or usability in the preparation of a new test. The teacher must keep in mind a number of very practical considerations which involves economy, ease of administration, scoring and interpretation of result.

Economy means the test should not be costly. The teachers must take into account the cost per copy, how many scores will be needed, ( for the more personnel who must be involved in giving and scoring a test, the more costly the process becomes). How long the administering and scoring of it will take, choosing a short test rather than longer one. Ease of administration and scoring means that the test administrator can perform his task efficiently.

### C. Item Analysis

There are several meanings of what item analysis are. Item analysis refers to the process of collecting, summarizing, and using information about individual test items especially information about pupils response to items. Item analysis is an important and necessary step in the preparation of good multiple choice test, because of this fact, it is suggested that every teacher who uses multiple choice test data should know something of item analysis. How it is and what it means.

For the teacher made test, the followings are important uses of item analysis: determining whether an item functions as teacher intended, feedback to students about their performance and as a basis for class discussion, feedback about students' difficulties, area for curriculum improvement, revising the item and improving item writing skill.

Item analysis usually provides two kinds of information on items:

1. Item facility, which helps the teacher decide if the test items are at the right level for the target group, and
2. Item discrimination, which allows the teachers to see if the individual items are providing information on students' abilities consistent with that provided by other items on the test.

Item facility express the proportion of the people taking the test who got a given item right. (item difficulty is sometimes used to express similar information, in this case the proportion that got an item wrong). Where the test purpose is to make distinctions between students, to spread the students out in

terms of their performance on the test, the items should be neither too easy nor too difficult. If the items are too easy, then students with different levels of ability or knowledge will all get them right, and the differences in ability or knowledge will not be revealed by the item. Similarly if the items are too hard, then able and less able students alike will get them wrong and the item will not help the teacher in distinguishing between them.

Analysis of item discrimination addresses a different target: consistency of performance by the students across items. The usual method for calculating item discrimination involves comparing performance on each item by different groups of the test takers: those who have done relatively poorly. For example, as items get harder, teachers would expect those who do best on the test overall to be ones who in the main get them right. Poor item discrimination indicates a signal that an item deserves revision.

If there are a lot of items with problems of discrimination, the information coming out of the test is confusing, as it means that some items are suggesting certain candidates that are relatively better, while other individuals are better, no clear picture of the students' abilities emerges from the test. (The scores, in other words, are misleading and not reliable indicators of the underlying abilities of the candidates) such a test will need revision.