

CHAPTER II

REVIEW OF RELATED LITERATURE

This chapter presents the review of some theories about Teaching English as Foreign Language, Excellent Class, The Evaluation System and Testing Language Skill.

A. Teaching English as Foreign Language

English as Foreign Language (EFL) is the language which is used by the people of other country or society. According to Patel and Jain (2008: 35), Foreign Language is the language where the secondary environment is not observed and the people of linguistically foreign societies use such language. Moreover, Gebhard (1996: 2) states that by EFL, English as studied by people who live in places where English is not the first language of the people. So, people studied English in other place where English is the first language. The goal of EFL in the educational system is to pass English entrance exams to enter good high school and universities. As such, much of the teaching is directed at making students' able to analyze and comprehend English, so they can pass entrance examinations, no necessary to preparing them to communicate in English. In EFL, teacher and students have some characteristic that the teacher more concern in how the student can speak or communicate by English in class. Dealing with Gebhard (1996: 3-4) states as follows:

In EFL settings, teachers are consistently concerned with ways to get students to speak English in class, ways to use authentic language teaching materials, having to teach to test, having too little time with

students, and getting students to take on more responsibility for their learning. EFL students hear and read in English is in the classroom. EFL students leave the classroom; they can enter any number of situations in which they can use English.

B. Excellent Class

According to *Permendiknas No. 20 tahun 2003*, the citizen who has intelligent potential and special aptitude have right to get the excellent education. Excellent class is created by classifying students according to academic abilities or the average of IQ as a tool to motivate students, to create fair competition between them. This can be done for school generally having diversity and learners. The development of a class of seed technically also demanded the power of adequate professional as special teachers' class to teach math, science, English, psychology, and sport. In addition, the time of study is provided more, because in this class is provided extra hours.

There is grouping system in excellent class at MTsN Aryojeding. That grouping system is aptitude grouping based on the student's talent and ability. Furthermore, there is matriculation program that used to make same about the student's perception to the subject will be learned. Moreover, there is psychology testing that used to know the student's intelligence level, talent, and creativity.

In the excellent class there are some facilities that different with the regular class. Such as, air conditioner, LCD projector, television, etc. That is to support the students' activity in teaching and learning process which in this class is provided extra hours. The class starts from 07.00 am until 03.50 pm including rest and pray time.

C. The Evaluation Systems

An evaluation is done to know how far students' progress and understanding toward the implementation of learning activity. The evaluation commonly called the assessment. In teaching and learning process commonly there are two kinds of evaluation that is formal and informal evaluation. Formal evaluation can be called as traditional. Meanwhile, informal evaluation called as authentic. Dealing with evaluation Brown (2001:402) states as follows:

Informal evaluation is involved in all incidental, unplanned evaluative coaching and feedback on tasks designed to elicit performance, but not for the purpose of recording results and making fixed judgments about a students' competence. Informal evaluation: evaluating students in the process of "forming" their competencies and skills in order to help them continue that growth process. It gives learners information about how they are progressing towards goals and what the next step in the learning process might be. On the other hand, formal evaluations are exercises or experiences specifically designed to tap into a storehouse of skills and knowledge within short time limit. They are systematic, planned sampling techniques constructed to give teacher and student an appraisal of student achievement. They occur at the end of lesson, unit, or course and therefore attempt to measure/ summarize what a student has grasped. They are focus on product of learning: objectively observable performance that may be evaluated somewhat independent of the process that a student has transferred to reach the end product.

1. Traditional/ Formal Evaluation

In designing formal evaluation should consider: a) Strategies for test-takers. It is useful strategies for taking the test. With some preparation in test taking strategies, learners can allay some of their fears and put their best foot forward during a test, b) Face validity means in the students' perception, the test is valid. Teacher can help to foster it with a test that is clearly doable within the allotted time limit, items are clear and uncomplicated, directions are clear, tasks

that are familiar and relate to their course work and a difficulty level that is appropriate to them, c) Authenticity means the tasks themselves need to be tasks in a form that students have practiced and feel comfortable with, d) Washback, to enhance washback is to provide a generous number of specific comments on test performance. The prompt return of written test with feedback is therefore very important to intrinsic motivation. In other words, take some time to make the test performance an intrinsically motivating experience through which a student will feel a sense of accomplishment and challenge.

Furthermore, evaluation in teaching and learning process covers listening, speaking, reading and writing. According to Hughes (1989:134), in evaluating listening involves macro-skills and micro-skills. The macro skills include listening for specific information, obtaining gist of what is being said, following directions and following instruction. The micro-skills include interpretation of intonation patterns and recognition of function of structure. The techniques are possibly used in testing listening is multiple choice, short answer, information transfer, note taking, partial dictation and recordings and live presentation. For speaking, the material tested include dialog and multi-participant interactions including operations of language functions such as expressing, narrating, eliciting, directing and reporting. The techniques are interview, questions and request for information, interaction with peers, role play, respond to tape recordings and imitation. In evaluating reading involves the macro and micro-skills. The macro-skills are scanning text to locate specific information, skimming text to obtain general idea and identifying argument. The micro-skills

are identifying referents of pronouns, using context to guess meaning of unfamiliar word, and understanding relation between parts of text. The techniques are multiple choices, true/false, completion, short answer, guided short answer, summary cloze, information transfer, identifying order of events, referents, arguments or topic and guessing the meaning. For writing, there are three things that should be considered: a) set writing tasks that are properly representative such as asking students to describe, to explain and to argue, b) the tasks should elicit samples of writing which truly represent the students' ability, c) the samples of writing can and will be scored reliably means the test-takers should not give too many choices of writing tasks and they should perform the same tasks to make scoring easier.

In the process of teaching and learning there is test of learning result. According to Purwanto (2009: 66) the test of learning result is the achievement test. This test is used to measure the student's achievement of the material that taught by teacher or the material that learned by students. Based on the function, the test of learning result divided to four aspects. There are formative test, summative test, diagnostic test and proficiency test.

a. Formative Test

Brown (2004: 6) states that formative assessment is evaluating students in the process of "forming" their competencies and skills with the goal of helping them to continue that growth process. This test is used to identify student's achievement about the material after following the learning process. Formative

test also called daily test because it tested in the end of material or unit instruction.

The key to such formation is the delivery (by the teacher) and internalization (by the student) of appropriate feedback on performance, with an eye toward the future continuation (or formation) of learning. For all practical purposes, virtually all kinds of informal assessment are (or should be) formative. They have as their primary focus the ongoing development of the learner's language. So when you give a student a comment or a suggestion, or call attention to an error, that feedback is offered in order to improve the learner's language ability (Brown, 2004: 6)

b. Summative Test

Summative test is used to identify student's achievement about some materials or main discussions that taught by teacher in the certain time. Dealing with evaluation Brown (2004: 6) states as follows:

Summative assessment aims to measure, or summarize, what a student has grasped, and typically occurs at the end of a course or unit of instruction. A summation of what a student has learned implies looking back and taking stock of how well that student has accomplished objectives, but does not necessarily point the way to future progress. Final exams in a course and general proficiency exams are examples of summative assessment.

c. Diagnostic Test

It is to identify the problem of students in learning process. Diagnostic test is used to diagnose the specified aspects of a language. In the pronunciation might diagnose in the phonological features of English are difficult for the student and should become a part of curriculum. Usually, such tests offer a

checklist of features for the administrator (often the teacher) to use in pinpointing difficulties.

d. Placement Test

As the name, placement test is used to divide students based on the certain character or their proficiency in some levels. It is to provide information which will help to place students at the level of the teaching program most appropriate to their abilities.

Recent years have seen a burgeoning of assessment in which the test-taker performs responses on a computer. It called Computer-Based Testing (CBT). Some computer-based tests (also known as "computer-assisted" or "web-based" tests) are small-scale "home-grown" tests available on web-sites. Almost all computer-based test items have fixed, closed-ended responses; however, tests like the Test of English as a Foreign Language (TOEFL) offer a written essay section that must be scored by humans (as opposed to automatic, electronic, or machine scoring). Brown (2004: 14) states as follows:

A specific type of computer-based test, a computer-adaptive test, has been available for many years but has recently gained momentum. In a computer-adaptive test (CAT) each test-taker receives a set of questions that meet the test specifications and that are generally appropriate for his or her performance level. The CAT starts with questions of moderate difficulty. As test-takers answer each question, the computer scores the question and uses that information, as well as the responses to previous questions, to determine which question will be presented next. As long as examinees respond correctly, the computer typically selects questions of greater or equal difficulty. Incorrect answers, however, typically bring questions of lesser or equal difficulty. The computer is programmed to fulfill the test design as it continuously adjusts to find questions of appropriate difficulty for test-takers at all performance levels. In CATs, the test-taker sees only one question at a time, and the computer scores each question before selecting the next one. As a result, test-takers cannot

skip questions, and once they have entered and confirmed their answers, they cannot return to questions or to any earlier part of the test.

2. Authentic/Informal Evaluation

There are several authentic evaluations that can be applied in teaching and learning process: a) Self- and peer-evaluation, a conventional view of language pedagogy might consider self-and peer-evaluation to be an absurd reversal of the teaching-learning process. Successful learners extend the learning process well beyond the classroom and the presence of a teacher or tutor, autonomously mastering the art of self-evaluation. The advantage of self-and peer-evaluation: speed, direct involvement of students, the encouragement of autonomy, and increased motivation because of self-involvement in the process of learning. It can be applied to evaluate oral production, listening, writing and reading skills, b) Conferences, conferencing has become a standard part of the process approach to teaching writing such as conversation about a draft. The advantage is one-on-one interaction between teacher and student such that the specific needs of a student can receive direct feedback. Through it, a teacher can assume the role of a facilitator and guide, it is also intrinsic motivation, students can feel that the teacher is an ally who is encouraging self-reflection and point students toward further development, c) Portfolio, it is a purposeful collection of students' work that demonstrates to students and their efforts, progress, and achievements in given areas. It is include essays, poetry, book reports and others, d) Cooperative test construction, one of the most productive of the various alternative evaluation procedures sees students directly involved in the

construction of test. Cooperatively produced task could engender a good deal of intrinsic involvement in the process of reviewing and selecting items for the final forms of the tasks. It means teacher discuss with their students before giving tasks to them so that they will be more interactive intensively, energetic and responsible toward their task (Brown, 2001: 415-419).

D. Testing Language Skills

In the teaching and learning process, the target of language testing is testing language skills (reading, speaking, listening and writing).

1. Testing Reading

Reading skill is a receptive skill. The task of teacher is to set reading tasks which will result in behavior that will demonstrate their successful completion. Hughes (1989: 136) states that The challenge for the language tester is to set tasks which will not only cause the candidate to exercise reading skill, but also result in behavior that will demonstrate the successful use of this skill.

In the reading test, there are reading macro-skills and micro-skills. For reading macro-skills are scanning text to locate specific information, skimming text to obtain general idea, identifying stages of argument, and identifying examples presented in support of an argument. The reading micro-skills are identifying referents of pronouns, using context to guess meaning of unfamiliar words, and understanding relations between parts of text.

The reading texts can be taken from textbook, newspaper, magazine, novel, letter, academic journal, etc. The texts can be in the forms of newspaper report, advertisement, editorial, etc.

The technique that might be used to test reading skills are multiple choice, true/ false, completion, short answer, guided short answer, summary cloze, information transfer, identifying order of events, identifying referents, guessing the meaning through the context.

- a.) Multiple Choices. The test-takers provide evidence of successful reading by marking a mark against one out of a number of alternatives.
- b.) True/ False. The test-takers should respond to a statement by choosing one of the two choices, true or false.
- c.) Completion. The students are required to complete a sentence with a single word.
- d.) Short answer. It is in the form of questions and requires the students to answer briefly.
- e.) Guided short answer. It is the alternative of short answer in which students are guided to have the intended answer. They have to complete sentences presented to them.
- f.) Summary cloze. A reading passage is summarized by the tester, and then gaps are left in the summary for completion by the test takers. This is really extension of the guided short answer.
- g.) Information transfer. One way to minimize demands on writing by test-takers is to require them to show successful completion of a reading task by supplying simple information in a table, following a route on a map, labeling a picture, and so on.

- h.) Identifying order of events, topics, or arguments. The test-takers can be required to number the events, etc.
- i.) Identifying referents. One of the “micro skills” listed previously was the ability to identify referents.
- j.) Guessing the meaning through the context. This is another micro skill mentioned early.

The above techniques are among the many techniques of testing reading. In scoring reading test, Hughes (1989: 155) suggested that errors of grammar, spelling or punctuation should not be penalized, as long as it is clear that the test-taker has successfully performed the reading task which the item set. The function of a reading test is to test reading ability. To test productive skills at the same time simply makes the measurement of reading ability less accurate.

2. Testing Speaking.

Hughes (1989: 113) states that the purpose of teaching speaking is to develop the ability in interaction successfully in that language and therefore, speaking involves comprehension as well as production.

The materials tested for speaking test include dialog and multi-participant interactions including operations of language function such as:

- a.) Expressing: thanks, requirements, opinions, comment, attitude, confirmation, apology, want/need, information, complaint, reasons/justifications.
- b.) Narrating: sequence of events

- c.) Eliciting: information, directions, service, clarification, help, and permission.
- d.) Directing: ordering, instructing, persuading, advising, warning.
- e.) Reporting: description, comment, decisions.

There are several formats that can be used to test speaking ability, such as:

- a.) Interview: the most obvious format for testing speaking. There are two or more test-takers that do conversation, as informant and reporter.
- b.) Question and request for information: for questions and request, yes or no questions should be avoided.
- c.) Pictures can also be used to elicit descriptions: series of picture (or video sequences) form a natural basis for narration.
- d.) Interaction with peers: two or more test-takers may be asked to discuss a topic, make plans, etc.
- e.) Role play: students can be asked to assume a role in a particular situation and the tester can act as an observer.
- f.) Response to tape-recordings: the students can be achieved through presenting only with the same audio- or video-tape recordings.
- g.) Imitation: the test-takers hear a series of sentences, each of which they have repeat in turn.

Scoring will be valid and reliable only if clearly recognizable and appropriate descriptions of criteria levels are written and scorers are trained to

use them. Description of speaking proficiency usually deals with accents, grammar, vocabulary, fluency, and comprehension as in the following examples taken from Hughes (1989: 131-132):

Proficiency Description

Accent

1. Pronunciation frequently unintelligible.
2. Frequent gross errors and a very heavy accent make understanding difficult, require frequent repetition.
3. Foreign accent requires concentrated listening, and mispronunciations lead to occasional misunderstanding and apparent errors in grammar or vocabulary.
4. Marked foreign accent and occasional mispronunciations which do not interfere with understanding,
5. No conspicuous mispronunciation, but would not be taken for a native speaker.
6. Native pronunciation, with no trace of foreign accent.

Grammar

1. Grammar almost entirely inaccurate phrases.
2. Constant errors showing control of very few major patterns and frequently preventing communication.
3. Frequent errors showing some major patterns uncontrolled and causing occasional irritation and misunderstanding.
4. Occasional error showing imperfect control of some patterns but noweakness that cause misunderstanding.
5. Few errors, with no patterns of failure.
6. No more than two errors during the interview.

Vocabulary

1. Vocabulary inadequate for even the simplest conversation.
2. Vocabulary limited to basic personal and survival areas (time, food, transportation, family, etc.)
3. Choice of words sometimes inaccurate, limitations of vocabulary prevent discussion of some common professional and social topics.
4. Professional vocabulary adequate to discuss special interests; general vocabulary permits discussion of any non-technical subject with some circumlocutions.
5. Professional vocabulary broad and precise; general vocabulary adequate to cope with complex practical problems and varied situations.
6. Vocabulary apparently as accurate and extensive as that of an educated native speaker.

Fluency

1. Speech is so halting and fragmentary that conversation is virtually impossible.
2. Speech is very slow and uneven except for short or routine sentences.
3. Speech frequently hesitant and jerky; sentences may be left uncompleted.
4. Speech is occasionally hesitant, with some unevenness caused by rephrasing and groping for words.
5. Speech is effortless and smooth, but perceptibly non-native in speech and evenness.
6. Speech on all professional and general topics as effortless and smooth as a native speaker's.

Comprehension

1. Understands too little for the simplest type of conversation.
2. Understands only slow, very simple speech on common social and touristic topics; requires constant repetition and rephrasing.
3. Understands careful, somewhat simplified speech when engaged in a dialogue, but may require considerable repetition and rephrasing.
4. Understands quite well normal educated speech when engaged in a dialogue, but requires occasional repetition or rephrasing.
5. Understands everything in normal educated conversation except for very colloquial or low-frequency items, or exceptionally rapid or slurred speech.
6. Understands everything in both formal and colloquial speech to be expected of an educated native speaker.

3. Testing Listening

The testing of listening involves listening macro-skills and micro-skills.

The macro-skills of listening include listening for specific information, obtaining gist of what is being said, following directions, and following instructions. The micro-skills of listening include interpretation of intonation patterns and recognition of function of structures.

There are several types of texts that can be used for listening test such as monologue, dialogue, or multi-participant. Those types can be in the forms of announcement, talk or lecture, instructions, directions, etc.

The source of listening test materials can be recordings of radio broadcast, television broadcast, teaching materials, or even recording of native speakers made by ourselves. There are some techniques that are possibly used in testing listening.

- a.) Multiple choices: this technique has some advantages and disadvantages. For listening test, the problem is greater because the test-takers should listen to passage while reading the alternatives/options. Therefore, the options must be short and simple.
- b.) Information transfer: it can involve such activities as the labeling of diagrams or pictures, completing forms, or showing routes on a map.
- c.) Note taking: the ability to take notes while listening to a lecture is in question; this activity can be quite realistically replicated in the testing situation. Test-takers take notes during the talk, and only after the talk is finished do they see the items to which they have to respond.
- d.) Partial dictation: it may be possible to administer it when no other test of listening is practical, although partial dictation may not be an authentic listening activity.
- e.) Recording and live presentation: the great advantage of using recording when administering listening test is there is uniformity in what is presented to the test-takers. It is fine if the listening activity is listened in a well-maintained language laboratory or in a room with

suitable equipment. If the conditions cannot be obtained, then a live presentation is preferred.

4. Testing Writing

The best way to test people's writing ability is to get them to write directly. There are three things that should consider developing a good test for writing as Hughes (1989: 83) states:

1. We have to set writing tasks that are properly representative of the population of tasks that we should expect the students to be able to perform.
2. The tasks should elicit valid samples of writing which truly represent the students' ability.
3. It is essential that the samples of writing can and will be scored validly and reliably.

The process of scoring can be done either holistically or analytically. Holistic scoring uses a variety of criteria to produce a single score. Its advantage of scoring is very rapid. The following is an example of holistic scoring provided by Cohen (1994: 327-328):

Holistic Scoring

- 5 The main idea is stated very clearly, and there is a clear statement of change of opinion. The essay is well organized and coherent. The choice of vocabulary is excellent. There are no major or minor grammatical errors. Spelling and punctuation are fine.
- 4 The main idea is fairly, and change of opinion is evident. The essay is moderately well organized and is relatively coherent. The vocabulary is good, and there are only minor grammatical errors. There are few spelling and punctuation errors.
- 3 The main idea and a change of opinion are indicated but not so clearly. The essay is not well organized and is somewhat lacking in coherence. The vocabulary is fair, and there are some major and minor grammatical errors. There are a fair number of spelling and punctuation errors.
- 2 The main idea and change of opinion are hard to identify in the essay. The essay is poorly organized and relatively incoherent. The use of vocabulary is weak, and grammatical errors appear frequently. Spelling and punctuation errors are frequent.

- 1 The main idea and change of opinion are absent in the essay. The essay is poorly organized and generally incoherent. The use of vocabulary is very weak, and grammatical errors appear very frequently. Spelling and punctuation errors are very frequent.

Other method of scoring is analytic, which require a separate score for each of a number of aspects of a writing task. The following is an example of analytic scoring provided by Cohen (1994: 328-329):

Analytic Scoring:

Content

- 5 – Excellent: main ideas stated clearly and accurately, change of opinion very clear
- 4 – Good: main ideas stated fairly, clearly and accurately, change of opinion relatively clear
- 3 – Average: main ideas somewhat unclear and inaccurate, change of opinion somewhat weak
- 2 – Poor: main ideas not clear or accurate, change of opinion weak
- 1 – Very poor: main ideas not at all clear or accurate, change of opinion very weak

Organization

- 5 – Excellent: well organized and perfectly coherent
- 4 – Good: fairly well organized and generally coherent
- 3 – Average: loosely organized but main ideas clear, logical but incomplete sequencing
- 2 – Poor: ideas disconnected, lacks logical sequencing
- 1 – Very poor: no organization, incoherent

Vocabulary

- 5 – Excellent: very effective choice of words and use of idioms and words forms
- 4 – Good: effective choice of words and use of idioms and word forms
- 3 – Average: adequate choice of words but some misuse of vocabulary, idioms and word forms
- 2 – Poor: limited range, confused use of words, idioms, and word forms
- 1 – Very poor: very limited range very poor knowledge of words, idioms, and word forms

Grammar

- 5 – Excellent: no errors, full control of complex structure
- 4 – Good: almost no errors, good control of structure
- 3 – Average: some errors, fair control of structure

- 2 – Poor: many errors, poor control of structure
- 1 – Very poor: dominated by errors, no control of structure

Mechanics

- 5 – Excellent: mastery of spelling and punctuation
- 4 – Good: few errors in spelling and punctuation
- 3 – Average: fair number of spelling and punctuation errors
- 2 – Poor: Frequent errors in spelling and punctuation
- 1 – Very poor: no control over spelling and punctuation

The choice between holistic and analytic scoring depends on the purpose of testing (Hughes, 1989: 105). If diagnostic information is required, then analytic scoring is essential. If the scoring is carried out by a small group people, then holistic rubric may be most appropriate. When scoring is conducted by a heterogeneous, less well-trained people or in a number of different places are probably used for. Whichever is used, multiple scoring is suggested.

5. Previous Study

This part presents previous study related to the study. The previous study is conducted to observe *Kajian Model-Model Evaluasi Program Pendidikan* by Aman in 2009 from Faculty of Economic and Social Science, Yogyakarta University. The objective of study was to describe: 1.) models of evaluation's program that developed now days, (2.) how are the advantages and weaknesses of the evaluation's program in education, and (3.) model of evaluation's program that suitable to evaluate the education's program comprehensively.

The result of this research showed that models of evaluation's education that developed in evaluation science are CIPP, Summative-Formative model, Goal Free Oriented, Goal Oriented, Krick Patrick, Black Box, UCLA, EKO, and EPBI. Every evaluation model has advantages and weaknesses based on the evaluation program that evaluated by model of evaluation. Krick Patrick model

has some advantages, such as: 1.) more comprehensive because it's covering cognitive, skill and affective aspect. 2.) The object of evaluation is not only the result of learning but covering process, output and outcomes too. 3.) More applicable for level of class because it doesn't require a lot of hand in evaluation process. While some weaknesses are: 1.) Not too attention in the input, even though the successful of output in the learning is also affected by the input. 2.) To measuring the impact is difficult to do and beyond the reach of teachers and school. While, CIPP model has some advantages, such as: more comprehensive because the object of evaluation is not only for the result but also covering contexts, input, process and result. Moreover, the weaknesses of this model are the application of this model has low level because to measure the contexts, input or result in general meaning will involve a lot of hand and need more times and costs. While the evaluation model that suitable in short term is Krick Patrick model and for long term is using CIPP model.

The research has similarities and differences with the previous study. The similarities were about the previous study also to describe about models of evaluation system in teaching and learning process and used descriptive qualitative approach. The differences were about the object of research. The previous study was conducted the research in wide area. It meant that the area not only in one place, like the present research. Moreover, the previous study also discuss about how the advantages and weakness of the evaluation's program in education and model of evaluation's program that suitable to evaluate the education's program comprehensively. The present research was discuss about

the evaluation system and the implementation of evaluation system that used by the teacher in EFL for excellent class.