

# CHAPTER I

## INTRODUCTION

### 1.1. BACKGROUND OF THE STUDY

In a foreign language teaching and learning process, evaluation is so closely related to the teaching process that it is impossible to work in one field without involving the other. Evaluation becomes an important part of the teacher's daily activities since without some kinds of evaluation, it would be impossible to teach.

An evaluation used in any teaching learning activity is not for burdening the students or teachers but it is used as a means for accelerating the process itself. A good teacher must be able to present material in more efficient ways in order that students can learn better. In order to do this, he must know a great deal about his students. He needs to know something about their abilities, their past achievements, their interests, their strengths and their weaknesses. Consequently, a teacher is continuously seeking answers to questions such as : "Which concepts should I teach first?", "How long will it take to complete this unit?", "What kinds of assignment will help the students the most?" and evaluation will help the teacher to answer the questions. By doing some evaluations, a teacher is able to increase his own

effectiveness by making adjustments in his teaching to enable certain groups of students or individuals in the class to benefit more. A good classroom evaluation will also help the precise areas of difficulty encountered by the class or by some students. A good evaluation can also be useful in curriculum development if they are constructed properly to measure the extent to which objectives are being realized.<sup>1</sup>

To become a good evaluation instrument, there are some criteria needed and one of them is validity. Gronlund states in his book "Measurement and Evaluation in Teaching" :

When one is selecting or constructing an evaluation instrument, the most important question to ask is "To what extent the results serve the particular uses for which they are intended?". This is the essence of validity.<sup>2</sup>

Validity is a measurement term used to describe the appropriateness of an information-gathering instrument or of the information which has been obtained. Whenever we obtain information, some errors will occur. The errors can be within the information-gathering instrument

-----

<sup>1</sup> William J. Micheels and M. Ray Karnes, Measuring Educational Achievement, New York : McGraw-Hill Book company, Inc., 1950, p.99

<sup>2</sup> Norman E. Gronlund, Measurement and Evaluation in Teaching, London : Collier Macmillan Publishers, 1981, p. 65

(inappropriate content, the difficulty of the items, ambiguity, the lack of amount of information), within the information-gathering process (test administration, scoring and recording), or within the individual being evaluated (trait instability, response to the evaluation task, test-taking ability and health).

The 1989 Structure I test which was observed by the writer is a language achievement test. In doing such a kind of test, it is not doubted that a student also employs his language aptitude. In other words, a student's language achievement actually depends on his innate language capacity namely the bigger innate language capacity a student has, the more language achievement he will get. Consequently, a language achievement test can be used to measure achievement and aptitude of a student at the same time and it means a language achievement test can be used to predict a student's success in his future language achievement.

Since validity is a very important quality in constructing an evaluation and there are many factors that can influence the validity of an evaluation, the writer is interested in making a research about the validity of an evaluation and because of limited time, she made a study on only one kind of validity that is predictive validity of the 1989 Structure I test of the English Department of

the Teacher Training College of Widya Mandala Catholic University. She chose the predictive validity since she found that good students sometimes got low scores and slower students got high scores in some kinds of test. In this study, she chose a structure test since she regarded that in the structure subject, there was a structural syllabus in which language items were selected, sequenced and graded to the simplest until the most difficult one for example the Simple Present Tense will be taught first before The Simple Past Tense.

In measurement theory, it is said that individuals have true scores. When we set out to measure a person's intelligence, we assume that he has some real amount of intelligence which differs from the amount other individuals have. Whether we are measuring height or weight, intelligence or personality or speaking ability, we assume that the individuals we are observing have some amount of the ability being measured. The information from the observations we make is used to estimate the individual's true ability or true score. However, the information from the observation contains error which means the observed score is not equal to the true score. In other words, we err when we measure both physical characteristics and psychological characteristics and in fact, psychological measurements are much less direct and

subject to even greater error.

An erroneous score can be either negative or positive . When we measure, we sometimes obtain a score which is larger than the person's true score and we sometimes also obtain scores which are less than a person's true score. No matter how carefully we follow the rules for the prevention of error to have a valid information, the information we obtain will always contain some errors. In other words, no score and no evaluative information is perfectly valid. Terry D. Brink in his book "Evaluation - a practical guide for teachers" states that:

The idea is to obtain all the data we can about the error in any information so that we can use that information to best advantage when making judgements and decisions. The first strategy is to obtain empirical evidence.<sup>3</sup>

Since any information, especially psychological measurement contains errors, the writer wants to know whether the results of the 1989 Structure I test can be used as a tool to predict the students' achievement in 1990 Structure II test namely the students who got high scores in 1989 Structure I test would also get high scores in 1990 Structure II test and the ones who got low scores

-----

<sup>3</sup> Terry D. Tenbrink, Evaluation - a practical guide for teachers, United States of America : McGraw-Hill, Inc., 1974, p. 35

would get low scores also in 1990 Structure II test.

## 1.2. STATEMENT OF THE PROBLEM

On the basis of the background discussed earlier, the writer formulated the statement of the major problem as follows : "Does the 1989 Structure I test have a high degree of predictive validity ?". In attempt to find the answer to the question, the major problem is broken down into the following minor problems :

1. Does the 1989 Structure I test have a significant correlation to the 1990 Structure II test ?
2. To what extent, do the estimated scores of 1990 Structure II test deviate from the actual scores ?
3. Is the regression significant in playing a role in predicting the students' achievement in the 1990 Structure II test ?

## 1.3. OBJECTIVE OF THE STUDY

Derived directly from the above formulated problems, the major objective of this study is to know whether the 1989 Structure I test has a high degree of predictive validity and the minor objectives are as follows :

1. to find whether the 1989 Structure I test has a

significant correlation to the 1990 Structure II test.

2. to find the extent of deviation of the estimated scores of 1990 structure II test from the actual scores .
3. to find the significance of the regression in playing a role in predicting the students' achievement in the 1990 Structure II test.

#### 1.4. SIGNIFICANCE OF THE STUDY

One of the best ways to check on an evaluation whether it is good or not is to determine whether the scores got by the testees can estimate their innate language ability. Since language tests only show the testees' performance and not their competence, only the degree or the extent of their innate language ability can be drawn. By knowing the degree of the testees' innate language ability, we can predict the success of their education achievement in the future. For example, from a testee's college entrance examination score, we can estimate the degree of his innate language ability so that we can know whether he will be able to pass all the target lessons or not. Therefore, the writer has made this study with one hope that these findings of this study will help English teachers, especially those who teach Structure I

and II in the English Department of the Teacher Training College of Widya Mandala University in their efforts to construct a good test by interpreting scores got by students in a specific test to see whether the scores can be used indeed to predict their education achievement in the future.

### 1.5. THE SCOPE AND LIMITATION OF THE STUDY

Realizing how broad the topic of this study would be, the writer felt it necessary to limit the study only to a certain scope. Although actually the results of structure test can be taken from Structure I, Structure II, Structure III and Structure IV test, in this study the writer only limits her study to the results of Structure I test and the results of Structure II test.

Since the main objective of this study is to see the degree of predictive validity of a specific test, the writer limits her study on the 1989 Structure I final semester test. To see whether the test has predictive validity or not, the writer had to correlate the test with another test administered at different time. Because of the need of the correlation and to be sure that there was a significant correlation, the writer took 1990 Structure II final semester test as the other test. The reason why she chose the 1989 Structure I final semester test as the

specific test, not the 1990 Structure I final semester test, was because the scores of 1991 Structure II final semester test had not been available when the writer took the data in April 1991 - the 1991 final semester test was held in July 1991. As a result, the writer regarded the scores of 1989 Structure I test and the scores of 1990 Structure II test as the newest data she could get. In addition, she chose final semester test - not mid term test - with the consideration that the material of final semester test represents the material of mid term test.

The English Department of Teacher Training College of Widya Mandala University holds D-3 and S-1 program. From the experience that the writer has got, the number of the students taking S-1 program is bigger than the students taking D-3 program. Realizing this fact, the writer took students of S-1 program to get a spread of scores of the group.

#### **1.6. THE ASSUMPTIONS OF THE STUDY**

Attempting to answer the questions stated in the statement of the problem, the writer assumes that the students' scores on the 1989 Structure I final semester test and on the 1990 Structure II final semester test are valid representatives of their achievements in their language components since the scores are obtained from two

scorers.

The writer also assumes that the respective teachers of the structure subjects are qualified in their field as the writer had experienced by herself that those teachers taught well and the writer got much additional knowledge about English structure from them.

In addition, the test is done under a controlled situation so the students being tested are assumed to do the test by themselves.

#### 1.7. THEORETICAL FRAMEWORK

This study is based on some theories and principles. The language testing theory states the nature and the functions of a good language testing in a language teaching learning process. A good test has some criteria and one of the criteria is that a good test should have predictive validity - the extent to which a test is related to some other valued measure of performance in the future. In order to predict the success of future education achievement, there must be correlation between the predicting variable and the predicted variable and cognitive learning theory will discuss about this. Furthermore, regression analysis and correlation theory will state the use of statistical analysis in predicting something.

### 1.8. THE DEFINITION OF KEY TERMS

To help the readers understand the ideas presented in this thesis, the writer feels necessary to include the definition of the following key terms.

Predictive Validity. According to Gronlund, predictive validity is the degree or the extent to which a test is related to some other valued measure of performance in the future.<sup>4</sup>

Structure. Structure, according to Hans P. Guth, is the study of how words work together in sentence.<sup>5</sup> French confirms that :

Structure of language is kinds of materials that go to make up language and the way the materials are put together, arranged and used to build up sentences, which as statements, questions, commands and exclamation from the living language and are the means of expressing what we feel, want to do, talk about and so forth.<sup>6</sup>

Final Semester Test. It is an achievement test

-----

<sup>4</sup> Norman E. Gronlund, Constructing Achievement Test, London : Prentice-Hall, Inc., p.133

<sup>5</sup> Hans P. Guth, Words and Ideas, California : Wordsworth Publishing Company Inc., p. 489

<sup>6</sup> French, The Structure of English, London: The English Language Book Society and Oxford University Press, 1963, p. 7

administered to the students at the end of the semester to measure the students' mastery of a particular teaching-learning material.<sup>7</sup>

Correlation. According to the Heritage Illustrated Dictionary of English Language, correlation is "A causal, complementary, parallel, or reciprocal relationship, especially a structural, functional or qualitative correspondence between two comparable entities". In statistics there are two kinds of correlation, namely the 'Positive Correlation' which is defined as the simultaneous increase or decrease in value of two numerically valued random variables, and the 'Negative Correlation', the simultaneous increase in the value of one and decrease in the value of the other of two numerically valued random variables.

Correlation in this study refers to the relationship between the scores of 1989 Structure I test and the scores of 1990 Structure II test of the English Department of the Teacher Training College of Widya Mandala University.

Correlation coefficient. In the Heritage

---

<sup>7</sup>Gronlund, Norman E., op. cit, p.19

Illustrated Dictionary of English Language, Correlation Coefficient is defined as ' A measure of the interdependence of two random variables that ranges in value from -1 to +1, indicating perfect negative correlation at -1, absence of correlation at 0, and perfect positive correlation at +1.

Regression Analysis. This is an analysis making use of a score of an individual on one variable as a means of determining or estimating a score of the same individual on another variable.<sup>8</sup>

#### 1.9. THE ORGANIZATION OF THE THESIS

This thesis consists of five chapters. The first chapter is the introduction. The second chapter is review of the related literature. The third one talks about the methodology of research. The obtained data will be analyzed in chapter four. Finally, the writer ends her thesis with summary, conclusion and recommendation that are stated in chapter five.

---

<sup>8</sup> Helen M. Walker and Joseph Lev, Elementary Statistical Methods, New York : Holt, Rinehart & Winston Inc., 1969, p. 190