



CHAPTER I

INTRODUCTION

In the field of education, evaluation is very important because it is the process of judging the effectiveness or worth of an education experience as measured against instructional objectives. It seeks an answer to the question, "Of what value is this measure of amount, status, or progress when compared with the instructional objectives?"¹ The value of an educational test to the classroom teacher is directly proportional to the extent to which the results obtained are translated into improved teaching practice by the teachers and improved learning condition for the pupils.²

In general, tests are used as sources of information for supervisory instructions; as aids to school surveys; and as research equipment.³ ⁴ However, in the field of language testing, six different purposes are aimed at : 1) to determine readiness for an instructional program; 2) to classify or place individuals in appropriate language

¹ Schwartz et al., Evaluating Student Progress. (N.Y. : Longmans Green and Co., 1957), p. 1.

² Green, Measurement and Evaluation in Secondary School (n.d. & n.p.), p. 89.

³ Green, ibid.

⁴ B. Spolsky, et al., "Preliminary Studies in the Development of Techniques for Testing Second Language Proficiency" Language Learning, 3 (August, 1968), p. 79.

classes; 3) to diagnose the individual's specific strengths and weaknesses; 4) to measure aptitude of learning; 5) to measure the extent of student achievement of the instructional goals; and 6) to evaluate the effectiveness of the instruction.⁵ Among these purposes, only two - to determine readiness for an instructional program, and to diagnose the individual's specific strengths and weaknesses - are directly related to teacher-made tests. Such tests test general proficiency to indicate "what an individual is capable of doing now," (as the result of his cumulative learning experiences) and "may also serve as a basis predicting future attainment."⁶

A good test should meet the criteria of validity, reliability, objectivity, comparability, sampling adequacy, administratability, economy and utility.⁷ However, the most important characteristic of any appraisal technique is validity.^{8,9,10} This refers to "the extent to which the results of an evaluation serve to particular uses of which

⁵D.P. Harris, Testing English as a Second Language (N.Y.:McGraw-Hill Book Co., 1969), pp. 2 - 3.

⁶Harris, ibid.

⁷Baron, Evaluation Techniques for Classroom Teachers (N.Y.:McGraw-Hill Book Co., 1958), p. 14.

⁸Scharwartz, op.cit., p. 52.

⁹Green, op.cit., p.3.

¹⁰N.E. Grorland, Measurement and Evaluation in Teaching (N.Y.:Macmillan Co., 1965), p.59.

it was intended." ¹¹ There are three main types of validity; content validity, statistical validity, and psychological and logical validity. Among them, statistical validity can be achieved indirectly by correlating the obtained scores with other known tests which have already been generally accepted as valid measurements, and which can serve as criteria for validation. ^{12,13}

Unfortunately, in the field of language teaching, many new methods are being experimented with in all parts of the world; and many of them are considered successful. The methods of testing, however, has so far failed to match the success of teaching methods, "partly because many of the skills being taught are difficult to measure, and partly, perhaps, because the purposes of teaching and testing are insufficiently co-ordinated by those who teach and those who examine."¹⁴ Carroll¹⁵ states that there are two ways in testing a language. One is by the "Integrative approach," (attempting to assess overall proficiency at once) and the other is by the "discrete-point approach" (attempting to break-up knowing a language into a number of distinct points making up each

¹¹ R. Lado, Language Testing (N.Y.: Longman Green and Co., 1961), p.30.

¹² Green, op.cit., p. 30.

¹³ J.B. Gerberich, Specimen Objective Test Items (N.Y.: Longman Green and Co., 1956), p. 24.

¹⁴ Spolsky, op.cit., p.59.

¹⁵ J.B. Carroll, et al., "An Investigation of Cloze items in the Measurement of Achievement in Foreign Languages" CEEBRD Report (Mass: Harvard University, 1959), cited by Spolsky, op.cit.

skill.) Robert Lado, was one of the pioneers who developed discrete-point tests in the form of "multiple-choice tests".¹⁶ This type is testing very well-known and is now used in every school in Thailand, as well as in other parts of the world.

The multiple-choice test has many advantages but also many limitations which need to be considered. Brown gives four basic advantages of multiple-choice tests as : specific content, preciseness of problems posed, rapid and accurate marking, and purity of content.¹⁷ These virtues make the multiple-choice tests generally superior in reliability and validity - two very important test characteristics - to composition tests.^{18,19} Lado suggests that objective tests, or multiple-choice tests, have two important values; saving time and scoring with speed and ease.²⁰ There can be no doubt that they test more fully and economically the elements of a language than do other forms of testing. However, he says "Unless they are designed with great skill, objective tests can give the appearance of validity without in fact being valid." It may be possible for an examinee to score highly on a multiple-choice test, for example,

¹⁶Lado, op.cit.

¹⁷J. Brown, Objective Tests : Their Construction and Analysis (London : Longman Green and Co., 1966), p. 142.

¹⁸Harris, op.cit., p. 50.

¹⁹J.M. Ruch, The Objective or New Type Examination (California: Berkley Book Co., 1929), p. 112.

²⁰Lado, op.cit., p. 112.

without, in fact, being able to express himself in the spoken or written language at a level appropriate to his age, experience and education level. That is, a multiple-choice test which achieves objectivity and validity in testing the elements of a language is not automatically testing the skill of an individual to use those elements or to manipulate them in situation.²¹ In addition to this, two basic limitations of any multiple-choice tests (or any type of objective test) are that it does not afford any opportunity to the examinee to display his powers of ordering his thought in a logical or effective sequence, and the test writer must concern himself the elimination of cultural bias.²²

At the present, the author considers the multiple-choice test an uneconomical test, especially for Thailand now where paper and printing material shortage is a national problem. Paper and printing materials cost us approximately triple price now compared with the price four years ago, with the price continuing to rise.^{23,24} Testing English language in Thailand aims primarily at its grammatical points at all levels. It is a "must" for teachers to test their students in

²¹J. Brand, "Some Aspects of the Testing and Teaching of English Language in Schools" AIT (Bangkok:English Language Center Publication, 1970), p.5.

²²Brand, ibid., p.8

²³"Printing Material Shortage", Siam Ruth, November 11, 1973, p.3.

²⁴"The Price of Printing Material in Near Future," Thai Ruth, January 29, 1974, p.1 and 16.

this aspect because it is required at both school and college levels.^{25,26} This requirement means that multiple-choice tests have to be long or have to be used very often if they are short, because there are a lot of grammatical points to be tested in each level. Unfortunately, most of the teacher-made multiple-choice tests in Thailand in all subjects have very short life spans indeed. They are usually used once or twice and thrown away. Very few teachers are interested in analyzing the test items for future use.²⁷ As a result, a very large sum of money is wasted yearly and increasingly. No one knows how much is wasted each year on teacher-made multiple-choice tests that are not actually valid due to their lack of both feedback from test analysis and of essential item-writing skills. When measured in money, the waste may run into millions of baht. But, when measured in human terms, the waste becomes appalling; the costs of stunted education of students, the frustrations and negative attitudes arising from their failure to learn, and the contributions to the drop-out rate with its consequences of unemployment and poverty result in a staggering drain on the nation's resources.²⁸

²⁵ English Curriculum Guide for M.S. 4-5 (Bangkok : Kuruspa Publisher, 1971), pp. 1 & 3.

²⁶ English Syllabus for the Certification of Education Level (Bangkok : Kuruspa Publisher, 1971), pp. 3 - 4.

²⁷ "Evaluation in Secondary Schools," Report on the First Seminar in Supervised Professional Experiences. (Bangkok : Samsen School, 1973), p. 31.

²⁸ J.R. Bormuth, "Comparable Cloze and Multiple-choice Comprehension Test Scores" Journal of Reading, 10(February, 1967), p. 291.

Therefore, as a means of solving some of the problems of multiple-choice test construction, its high cost, and limitations of expression, the author wants to introduce the "cloze test" to be used alternatively to multiple-choice test, if it is possible, in testing the structure of the English language. This implies that the grammatical points will be measured by using reading comprehension as a means.

A cloze test is a test constructed according to cloze procedure which was introduced by Taylor in 1953.²⁹ It is valuable because it is highly reliable and valid and can be easily and objectively constructed and scored.³⁰ The ability to read with understanding can be measured rather well by cloze test. At the intermediate grade level, Bormuth³¹ has shown that correlation between cloze test scores and multiple-choice test scores on seven different types of comprehension was very high. He states that "...the comparable scores could be said to be highly reliable."³² Taylor³³ found that the correlation between AFQT (Armed Forces Qualification Test) scores and pre-cloze, any-word

²⁹W.L. Taylor, "Cloze Procedure: A New Tool for Measuring Readability," Journalism Quarterly, 30 (1953), pp. 415-433.

³⁰Bormuth, op.cit., p. 291.

³¹Bormuth, op.cit., p. 295.

³²Bormuth, op.cit., p. 297.

³³W.L. Taylor, "Recent Developments in the Use of 'Cloze Procedure'" Journalism Quarterly, 33(1956), p. 45.

deletion test scores was .70 and post-cloze, noun-verb-adverb deletion test scores was .80. Rankin³⁴ obtained a correlation of .78 between the scores of a post-cloze, noun-verb deletion test and a criterion. Jenkinson³⁵ found a correlation of .82 between a pre-cloze, any-word deletion test and a reading comprehension test at the adult level. Friedman³⁶ obtained correlations ranging from .90 to .91 between 20 pre-cloze, any-word deletion tests and 20 reading comprehension tests. For non-native speakers, Oller, et al³⁷ found that by using the acceptable word scoring method, and by using Darnell's cloze-entropy technique, cloze test scores have yielded correlations with the total scores on the UCLA ESLPE (University of California at Los Angeles English as a Second Language Placement Examination) and the TOEFL, ranging between .83 and .89.

For the reliability of cloze tests, all of the studies utilizing any-word deletion have reliabilities between .80 and .98 when measured by split-half and Kuder-Richardson formula 21. Oller³⁸ states

³⁴E.F. Rankin, "The Cloze Procedure - Its Validity and Utility" Fifth Yearbook of the National Reading Conference, 8(1959), pp.131-144, cited by E.F. Rankin, Jr., "The Cloze Procedure - A Survey of Research" Fourteen Yearbook of the National Reading Conference, (1965), p.136.

³⁵M.E. Jenkinson, "Selected Process and Difficulties in Reading Comprehension" (Unpublished doctoral dissertation, the University of Chicago, 1965) cited by Rankin, Jr., ibid.

³⁶Friedman, cited by Rankin, Jr., ibid.

³⁷J.W. Oller, "Cloze Tests of Second Language Proficiency and What they Measure," Language Learning, 23(June, 1973), p. 111.

³⁸Oller, ibid.

that cloze tests have more advantages than multiple-choice tests because they have very high internal consistency, can be constructed and administered very easily by "a person who is unskilled in the technicalities of constructing multiple-choice tests," and can be standardized (or criterion referenced) by simply comparing scores of non-native speakers with scores achieved by native speakers. In one study with Conrad, he concluded that "...the cloze method is a very promising device for measuring ESL proficiency,"³⁹ In addition, it noteworthy that in a latest study Hauska⁴⁰ obtained correlation coefficient of $\rho = 1.00$ when he compared reading comprehension scores with cloze scores.

Therefore, the author thinks that a cloze test might possibly be used alternatively to a multiple-choice test for testing English grammatical points.

Purposes of Study

General Purpose

The purpose of this study was to find out if the cloze test

³⁹J.W. Oller and C.A. Conrad, "The Cloze Technique and ESL Language Learning, 22(December, 1971), p. 285.

⁴⁰J.T. Houska, "The Efficacy of the Cloze Procedure as a Readability Tool on Technical Content Material as Used in Industrial Education at the High School Level," Dissertation Abstracts International, 3 (February, 1972), p. 4500-A.

could possibly be used alternatively to the multiple-choice test for testing English grammatical points. The main reason behind this study was to save the expense on printing material and the time for preparing a valid structure test.

Specific Purposes

The specific purposes of this study were:

1. To develop and validate a cloze of English structure test by correlating cloze test scores with multiple-choice structure test scores.

2. To find out if there were any order effects on the scores of the two types of tests taken on the same content, e.g. cloze test first, and multiple-choice test second, and vice versa.

3. To find out if the differences between sexes effects cloze test and multiple-choice test scores.

4. To provide a frame of reference for interpreting the cloze test scores in terms of multiple-choice test scores.

Significance of the Study

The author of this study expect the following benefits from the study:

1. It will introduce the use of cloze procedure to test English structure as an alternative to multiple-choice tests. The former is much more economical in time, expense and energy.

2. It will provide a practical frame of reference for interpreting cloze test scores in terms of multiple-choice test scores. It will

be very useful for English teachers at the Certification of Education Level, especially for those who are in Nakornsawan Teachers' College

3. It will be a pioneer research in utilizing cloze procedure for testing the structure of a language

4. It will provide the ideas of cloze procedure in measuring other abilities to those who are interested in it, and those who want to do more research in this field.

Delimitation of the Study

The scope of this study was as follows:

1. It was confined especially to 200 Thai second-year students (120 boys; 80 girls) in Nakornsawan Teachers' College in the academic year of 1973. They all had had at least 8-year's experience in learning English as a foreign language compulsorily and were drawn by the random sampling method. They were learning "English 5".

2. The selection of subjects was made on the basis of the above criteria, with no regard to the subjects' socio-economic background, learning achievement, intelligence quotient, health, motivation or any other psychological variables.

3. The study was purposely aimed at measuring of students' English structure proficiency by utilizing a multiple-choice structure test and a cloze test of English structure.

Basic Assumptions

This study has the following basic assumptions:

1. The multiple-choice structure test and the cloze test utilized in this study were suitable to the subjects' abilities in the areas of grammatical points, structures, vocabulary, cultural background of English language, and readability.
2. The multiple-choice structure test utilized in this study was reliable because it had been tested and all items were analyzed statistically in the pilot study.
3. The testing and administration were reliable because the examinees and the proctors followed the same test instructions.

Hypotheses

The author hypothesized that:

1. The scores of the multiple-choice structure test and the cloze test would be significantly correlated at the .01 level of confidence.
2. There would be no order effect on the scores of the multiple-choice structure test or the cloze test taken on the same matters.
3. The differences between sexes would not effect on the multiple-choice structure test or the cloze test scores.

Imperfection of the Study

It was possible that this study was imperfect. The imperfection might due to partly the subjects' unfamiliarity with the cloze test, and

partly the short span of cognitive abilities. The tests were based upon the grammatical points learned by the subjects in "English 1, 2, 3, 4 and 5". Only "English 5" was being learned.

Besides, the imperfection might possibly be caused from the lack of very serious intention of some subjects in performing the tests in the pilot and in the main studies.

Definition of Terms

- Multiple-choice structure test : refers to a 75-item with 4 alternatives for each item. It is a research-made test for measuring the subjects' English structure proficiency.
- Cloze test : refers to a 690-word passage of which only 75 grammatical points based on the multiple-choice test, were deleted and replaced with a blank ten spaces in length for each. It is a research-made test for measuring the subjects' English structure proficiency by using reading comprehension as a means.
- Validation : refers to the method of correlating the scores of the cloze test with the scores of multiple-choice structure test that has been proved valid.

- Teachers' College : refers to an officially owned institution where offers a 2-year and a 4-year studies to teacher students in Thailand.
- Certification of Education level: refers to a 2-year study in a teacher college
- Second-year student : refers to a teacher-student studying in the second year in a teacher college at the Certification of Education Level.
- English structure proficiency : refers to the scores of the subjects obtained from choosing the only correct alternative in each item of the multiple-choice structure test and/or from filling up the gaps of the cloze test grammatically and contextually.