# A COMPARATIVE STUDY OF TESTING GRAMMAR KNOWLEDGE OF IRANIAN STUDENTS BETWEEN CLOZE AND MULTIPLE – CHOICE TESTS

By

**MINOO ALEMI ***                                    **APAMA MIRAGHAEE ****

* Sharif University of Technology, Tehran, Iran.
** Islamic Azad University, Tehran, Iran.

*ABSTRACT*

*The present study was carried out to find out whether regular administration of cloze test improved the students' knowledge of grammar more than the multiple choice one. Subjects participating in this study were 84 Iranian pre-university students of Allameh-Gotb-e Ravandi University, aged between 18 and 35 and enrolled in a grammar course. To investigate the research, the candidates were screened through a NELSON test. This test enabled the researchers to have homogeneous subjects; therefore, among 84 subjects, 75 students were selected to take part in this research. Among this number of students, 39 students were to be in one group and the other 36 were the other group members. The two groups were provided with 10 consecutive weeks, each session 100 minutes long. Every week one experimental group was treated by a 25-item cloze test of grammar while the other experimental group was provided with a 25-item multiple choice one. The time allowed for each multiple choice test was about 25 minutes. Each week the grammatical point which had been taught the previous session, was tested. After the treatment, a 60-item test consisting of a 30-item cloze passage and a 30-item multiple-choice grammar section were administered to find out whether there was any difference in the grammatical knowledge of both experimental groups. The analysis of data indicated that there was no significant difference between the average performance of the two groups. In another words, the findings proved that the cloze procedure did not improve the subjects' knowledge of grammar any more than the multiple-choice grammar test did.*

*Keywords: Cloze Test, Multiple-Choice Test, Grammar Knowledge.*

## INTRODUCTION

Throughout the past decades, various theoretical arguments have been proposed to improve the existing demerits of contemporary language teaching methods and quite naturally language teaching procedures. Some methods were temporarily stopped, while others continued their existence and are still extensively utilized in many countries. One of these widely used testing procedures which has taken its roots from holistic approaches of teaching is the 'cloze test'. It is easy to prepare and rather easy to score. Teachers like it too, because it is integrative, that is, requiring students to possess the components of language simultaneously, much like what happens when people communicate. Moreover studies have shown that it relates well to various language measures from listening comprehension to overall performance on battery of language tests.

In 1977, Randall Jones commented that the cloze procedure has been shown to be an efficient, reliable and valid method of measuring (second) language proficiency. However, recent studies indicate that the use of cloze test as a proficiency measure can yield parallel and reliable forms, with content validity, if certain steps in the preparation of the testing materials are carefully followed. They are the careful selection and preparation of several alternative tests at appropriate levels of reading difficulty.

Cloze test criterion related validity and its two components, concurrent validity and predictive validity, have been extensively examined in the literature by researchers such as Bachman (1982) and Brown (1983,1988,1989), who carefully prepared different rational and random deletion

cloze test versions and then correlated the results with standard proficiency tests administered at the same time, finding strong positive relations.

There also exist a number of studies about different methods of testing grammatical structures more effectively. Harris (1968) states that the preparation of a structure test should always begin with setting up a detailed outline of the proposed test content. The outline should specify not only the structures that are to be tested, but the percentage of items to be written around each problem. This outline may have to be modified somewhat on the basis of the result of pre-testing, but great care must be taken to ensure that the final form of the test includes a broad range of relevant grammatical problems in proportions which reflects their relative importance. Selection of structures to be included in a proficiency test must be made subjectively. Perhaps the test approach is to examine a number of standard tests prepared for students at the level of the intended test population, listing the structures commonly presented in these works and noting the emphasis which each receives.

Heaton (1988) believes that carefully constructed completion items are a means of testing a student's ability to produce acceptable and appropriate forms of language. Completion items can not be machine marked but they are very useful for inclusion in classroom tests and for exercise purposes. There are two major advantages in using a passage of continuous prose rather than separate sentences when giving a completion type test. Firstly, the use of context often avoids the kinds of ambiguity. Secondly, the students experience the use of grammar in context, being required to use all the context clues available in order to guess many of the missing words.

Oller (1979) argues that the cloze tests generally satisfy the first pragmatic naturalness requirement: They require the learner to process temporal sequences of elements in the language that conform to normal contextual constraints. He also believes that the cloze tests require the utilization of discourse level constraints as well as structural constraints within sentences. Probably, it is the distinguishing characteristic which makes cloze tests so robust and which generates their surprisingly strong validity coefficients in relation to other pragmatic testing procedures.

However, what is of great importance is whether it is justified to assume that these standard tests can be used as appropriate devices in application of some grammatical elements. In other words, it might be possible for a student to develop his grammar proficiency through understanding and performing the befitting use of grammatical items. Hence, this study is designed to investigate the effectiveness of cloze procedure in connection with grammar tests.

## Background and Purpose

In the present post-communicative age there appears to be a discernible shift towards a more structural approach to second language learning. More and more, one hears about consciousness raising (C-R), which is defined as "the drawing of the learner's attention to features of the target language" (Rutherford, 1987). Although C-R comes under the rubric of formal instruction, it is not quite the same thing as traditional grammar teaching.

Since all language is 'an effort meaning', it is obviously necessary for the learner to surface constituents. In other words, learners should be given ample opportunities to explore relationships between form and function. That is they have to operate the two linguistic systems in tandem.

In the present study, it was intended

- To determine whether the use of "cloze tests" is more effective than the "multiple-choice tests" in developing the EFL learners' knowledge of grammar at pre-university level,

- To provide some guidelines for curricula developers'

- To prepare some insights for both material preparation and test design, and finally

- To provide the EFL teachers with some insights to teaching as well as testing grammar.

## Statement of problem

Much of research about cloze procedure has been based on theoretical views of cloze tests suggested by Alderson and Brown (1978) and Oller (1978). Research studies, however, have shown that performance on cloze tests correlates highly with the listening, writing, and speaking abilities. In other words, cloze testing is a good indicator of

general linguistic ability, including the ability to use language appropriately according to particular linguistic and situational contexts. It is argued that three types of knowledge are required in order to perform successfully on a cloze test: linguistic knowledge, textual knowledge, and knowledge of the world. As a result of such findings, cloze tests are now used not only in general achievement and proficiency tests but also in some classroom placement tests and diagnostic tests. However, it might be of great importance to bridge the gap between testing methods and teaching different main skills and sub-skills of language. Therefore, there exists a need to apply the cloze procedure in developing proficiency skills.

## Significance of the Study

The importance of cloze procedure is now widely recognized. In other words, cloze procedure could be considered as a facilitator that aims to bridge the gap between teaching and testing. Hence it seems quite suitable to administer the cloze procedure in developing EFL students' knowledge of grammatical structure.

This study gave both language teachers and learners an opportunity to work with teaching and testing activities due to the fact that teaching and testing are closely related to each other. In fact, what makes students perform better in their grammar tests is their ability to use context as well as their grammatical knowledge in order to act more communicatively within their teaching curriculum; therefore, a deep study over the practicality of the cloze tests of grammar was highly required. In addition, by involving the students in testing activities, learners may consider language learning as an active and enjoyable activity not as a boring or tedious one. Since the students' attention is on using their grammatical knowledge through taking part in various test session, their mental barriers and anxiety will be minimized; on the other hand, they will learn to take responsibility for their own progress.

## Research Question

In order to investigate the purpose of study, the following research question is proposed: "Does cloze test procedure develop knowledge of grammar more effectively than multiple-choice tests?"

## Definition of Important Terms

For the sake of clarification, lucid definitions for some key terms are given as follows: (Standard) cloze test is "a passage of appropriate difficulty and approximately 220-250 words with every 11 word deleted. The first and the last sentences of the passage are to be intact.

Grammar is "a description of the structure of a language and the way in which linguistic units such as words and phrases are combined to produce sentences in the language. It usually takes into account the meanings and the functions these sentences have in the overall system of language. It may or may not conclude the description of the sounds of a language" (Richards & Heidi Platt, 1992, p.161).

Multiple-choice test is "a test in which the examinee is presented with a stem and four or five possible answers from which one must be selected" (Mousavi, 1997, p.82).

## Delimitations and limitations of the Study

Because of some limitations, conducting a random selection was not possible in this study. That is why the researcher used four intact groups taught by four different teachers to randomly select the two experimental groups. The researcher was not also able to assign a control group for the two experimental groups in her research.

There were also certain delimitations that had to be imposed on the study. First, this study was concerned with one of the two major sub-skills, that is, grammar. Teaching only one grammatical point per week was possible. Second, the students in this study were all at pre-university level as far as their knowledge of proficiency was concerned.

## Methodology

This research was carried out to study the relationship between the use of cloze tests and the development of Iranian EFL learner's knowledge of grammar. In other words, it was a contrastive study over the cloze and multiple-choice tests to see if the cloze procedure improves the students' knowledge of grammar any more than multiple-choice grammar test does. What will follow is a detailed description of how the study was conducted.

## Design of the study

According to Best and Kahn (1989), because of some limitation, conducting a true experimental research may not be possible. Therefore, quasi-experimental method was adopted to approximate the standards of the true experimental method as much as possible. Of the many quasi-experimental designs, the pretest-posttest non equivalent-groups design best fit the situation.

The schematic representation of this design is as follows:

01 x 02

03 x 04

01, 03 = Pretest. 02, 04 = Posttest,

X = Experimental group one,

C = Experimental group two.

## Subjects

Subjects participating in this study were 84 Iranian pre-university students of Allameh-Gotb-e Ravandi University, both male and female majoring in English language, aged above 18. The maximum age was 35 and enrolled in a grammar course. The classes were held in 17 consecutive weeks, two sessions a week, each session 100 minutes long. This number of students was given a NELSON TEST, and on the basis of the results, out of 84 subjects, 75 students were selected as explained later. The number of subjects remained to go under treatment for 12 consecutive weeks. They were all then dichotomized to two experimental groups of students called "group I" (38 students) and "group II" (37 students).

## Instrumentation and materials

This research took advantage of three types of tests for data collection at different stages of the procedure. The tests were used to equate the experimental and control groups. They are enumerated as follows:

- A standard test namely "NELSON, number OSOA, (1976)" with reliability index of .95 . The candidates were screened through this test as a pre-test which included 50 multiple choice items. The test functioned for the purpose of making the subjects homogeneous to their level of English proficiency.

- Two teacher-made test types with reliability of .84 and

.82 administered during 10 consecutive weeks as the treatment of the study. Each session, "group I" received a 25-item multiple-choice test, while "group II" received a variable-ratio cloze test again with 25 blanks. Both tests were mainly prepared in order to test the grammatical points which were taught in each class during the course.

- A 60-item post-test with reliability of .94 consisted of two main parts. The first part was a cloze passage with 30 blanks followed by its 30 multiple-choice answers. The second part, namely the multiple-choice part, consisted of 30 multiple-choice items. The post test was administered to evaluate the subjects' knowledge of grammar after the 10-week test treatment.

It is worth mentioning that a pilot study had been carried out to validate the post test. Through administering the post test and a standard proficiency grammar test (NELSON, number 050 A) to 31 freshman Iranian EFL students and calculating the correlation between the two.

About the Nelson proficiency test, it is worth indicating that seven items (items 21, 22, 28, 30, 31, 46, and 47) of the afore-mentioned test were deleted to make it a test with a pure grammatical content.

## Procedure

To accomplish the purpose of the study and to verify the hypothesis, the following steps were taken:

First, a Nelson test of English Proficiency was given to 84 candidates. From among these subjects those who scored between 25 and 45 were selected and divided into two experimental groups of students namely "group I" and "group II". To ensure the homogeneity of the variances an F-test for the homogeneity of the variances and then a T-test for comparing the means were used. The results proved the homogeneity of the two groups.

In this study, "group I" was provided with ten 25-item multiple-choice grammar tests for ten consecutive weeks, while "group II" benefited from ten cloze passages with the same 25 items and more or less the same grammatical content. It is worth mentioning that three passages of the students' reading book were selected in advance. The readability calculated for the first, the fifth, and the ninth

passages of the reading book "Reading Attack Skills for Adults", showed that texts with difficulties between 6.53 and 19.17 were acceptable for this study. The readability of each cloze passage was then estimated in advance through Fog's formula and proved to be at the right level for the subjects.

It is worth indicating that each test assessed the students' knowledge of grammatical points taught the previous session. The followings are a list of the grammatical points tested during the treatment period:

- First week: simple present, present continuous, and present tense of verb 'to be'

- Second week: past simple, past continuous, and past tense of verb 'to be'

- Third week: a revision on the materials of the last two weeks

- Fourth week: present perfect tense

- Fifth week: present perfect continuous

- Sixth week: future tense and the format of 'to be going to'

- Seventh week: a revision on the materials of the last six sessions

- Eighth week: 'wh' words such as where, why, who whom etc.

- Ninth week: articles

- Tenth week: prepositions

In this study, the treatment period which was actually a testing treatment period, lasted for 12 sessions. At the end of this period, a post test was administered to both groups. The post test, which was an achievement test of grammar, consisted of two main parts: (i) a cloze passage followed by its 30-item multiple-choice answers; and (ii) a 30-item multiple-choice test. The students in each group had 80 minutes to mark off the answers on their answer sheets.

Later on, through the application of a T-test, the gain scores of the subjects in the experimental group I were compared with those of the subjects in group II. The reason was to see if the difference between the two groups in terms of the experimental subjects'

development of grammatical knowledge was significant or not. In other words, the performance of the two groups was compared and analyzed to decide whether the null hypothesis should be rejected or accepted. About the validity of the weekly tests, it is worth mentioning that the validity was checked to be acceptable through correlating the test's mean score with the Nelson total mean score the two groups took a pre-test.

The reliability of the post test was checked by running a pilot study and then calculating the test reliability through 'KR-21, formula. The validity of the post test was also examined through administering number '050 A' Nelson test with 43 grammatical items and calculating Pearson's 'criterion-related' correlation coefficient formula.

The next coming section will deal with the analysis of the tests results in detail.

### Data Analysis and Discussion

In this section, the method of analyzing the collected data is to be described in detail. The steps taken in analyzing the test results will be fully elaborated while the researchers will interpret the results with regard to the central purpose of the study.

### Statistical Analysis

To examine whether there was any significant difference between the mean scores of the two experimental groups, the data obtained were analyzed as· follows:

After scoring the tests, the results were put under some statistical analyses to provide an appropriate, accurate and reasonable answer to the research question.

First, an F-test and then a T-test were conducted to determine the homogeneity of the two groups. In order to complete the significance of the difference between the results of the post-test of the two groups another T -test was applied.

About each grammar test, either multiple choice or cloze ones, the validity was calculated through correlational procedures. The reliability of the post test was also checked using KR-21 formula. The next section deals with the analysis of the data in detail.

### Homogeneity of the subjects

To determine the level of subjects' homogeneity, a NELSON

test of English language proficiency was administered at the beginning of the study. As Table 1 depicts, through the application of an F-test and a T-test, it was proved that the students whose scores in NELSON test fell between 25 to 45 were homogeneous.

Where    X= Mean

       V= Variance

       DF= Degree of Freedom

       Fo= F-observed

       Fc= F-critical

       To= T-observed Tc= T-critical

       N= Number of students

       G1 = Group I (group working with cloze tests)

       G2= Group II (group working with multiple-choice tests)

Since the value of the F-observed was smaller than the value of F-critical at both 0.05 and 0.01 level of significance, it was concluded that the variances fulfilled the condition of homogeneity. On the other hand, the value of t-observed was also smaller than the value of t-critical at the same level of significance. Therefore, it can be concluded that there was no significant difference between the mean scores of the two groups. These two figures, the F-value and the T-value proved that the subjects of the two groups were more or less homogeneous, enjoying the same level of grammatical knowledge.

## Reliability and Validity of the Post Test

The reliability and validity of the post-test were checked through applying a pilot study and using correlational procedures. From Table 2, the reliability was estimated through KR-21 formula. Validity was checked by correlating the post test mean scores with the official NELSON total scores of grammar.

A coefficient of correlation of 0.41 indicated that there was

a moderate relationship between the set of post test scores and the official NELSON total scores of grammar.

Then the researchers conducted a t-test to show whether the difference between the scores of the subjects in the two groups was significant or not. Table3 shows the results.

Since the t-observed does not equate or exceed the t-critical value for rejection of the null hypothesis at both 0.01 and 0.05 levels of significance for 73 degrees of freedom, the hypothesis is not rejected. In other words, there is no significant difference between the mean scores of the students who worked with grammar cloze tests and those who worked with grammar multiple choice ones.

## Conclusion and Implication

It is quite axiomatic that knowing and using a foreign language is not solely a sudden effort. A foreign language learner learns more and more about the language. He must learn the skills and gradually use them in order to develop his knowledge of that particular language. Grammar, as one of the components of a language, plays a crucial role in learning. Therefore, developing the learner's ability to apply his knowledge of grammar is possible through employing both testing and teaching activities and they seem to be an important component of a language course.

In this study, the central purpose was to answer the question whether cloze or multiple-choice tests give the best result as far as the development of the grammar knowledge among the Iranian EFL students is concerned. To this aim, a NELSON test was administered to the students to ensure their homogeneity as far as their proficiency level of English was concerned. As soon as the homogeneity of the subjects was ensured, they were divided to two

| | x- | FO | FC | N | TO | TC | OF |
|---|---|---|---|---|---|---|---|
| G1 | 36 | 1.09 | (0.05 LEVEL) | 39 | 0.15 | (0.05 LEVEL) | 73 |
| | V1=242 | | 1.53 | | | 2.000.1180 | |
| G2 | 22.91 | 1.09 | (0.01LEVEL) | 36 | 0.15 | (0.01 LEVEL) | 73 |
| | V2=221 | | 1.84 | | | 2.660 2.617 | |

Table 1. Results of the NELSON Test

| Reliability | Validity |
|---|---|
| KR-21 | r with NELSON Total |
| 0.94 | 0.41 |

Table 2. Reliability and Validity of the post-test

| | -x | Level Of Significance | T-Observed | T-Critical |
|---|---|---|---|---|
| G1 | 32.5 | 0.05 | 0.03 | 2.000 |
| G2 | 32.5 | 0.01 | 0.03 | 2.660 |

Table 3. A two-tailed test between the mean scores of the two experimental groups' performance at 73 degree of freedom

experimental groups. Then the treatment started. The subjects in one group received multiple choice tests per week while the subjects in the other group benefited from the cloze ones. At the end of the treatment period, both groups received a post test which consisted of both cloze and multiple choice tests of grammar. The data gathered from the subjects' scores were statistically compared and analyzed. The results indicated that there was no significant difference between the mean scores of the two groups. In other words, the effect of cloze tests in the development of the subjects' knowledge of grammar was the same as that of the multiple-choice ones.

In this case, the independent variable was the use of both cloze and multiple-choice tests and the dependent variable was the students' scores on the post-test of grammar.

## Pedagogical Implications

This study, having focused on an issue in language testing, seems capable of introducing new horizons in the field of language teaching as well as testing.

Although the findings proved that the use of cloze tests had no significant effect on the students' knowledge of grammar, language teachers can easily benefit from these tests to educate their students more effectively. Instead of wasting time and energy on only one test type, they can concentrate and embark upon more tests in order to change the serious and boring mood of grammar classes. This research also encourages language teachers who are reluctant to change their methods and techniques of teaching and those who still stick to the routine approaches to try the new ways in both teaching and testing foreign languages and to change their view points in favor of the more up-to-date and useful ones in this field. In this way, students also learn to pay attention not only to grammatical rules, but also to different techniques in learning these rules. That is, using cloze passages as a means of improving the knowledge of grammar teaches students to take more responsibility for their own learning. This leads them to be active participants rather than passive recipients. Using cloze passages with some grammatical deletions helps learners to enjoy a great amount of diversity during their grammar class. Different

types of tests encourage the students to be more autonomous and self-confident language learners. Their reading comprehension will also improve through the application of cloze tests. Teachers can also use cloze passages to provide the class with vocabulary and grammar exercises appropriate for the students' level. The findings of this research also help teachers and test writers to use and apply different teaching tools in designing and writing both teaching and testing materials.

## Suggestions for Further Research

A research is a recurring sequence of events. It is cyclical. Its nature is such that the more answers you obtain the more questions arise.

The study presented during this research is only a preliminary one, and requires many modifications to provide more generalized conclusions. What follows are some of the recommendations which need to be considered for further research:

- This study can be replicated by a large number of Iranian EFL learners at different levels of language proficiency.

- Cloze tests can be used in developing the knowledge of different skills and subskills.

- The effect of cloze tests can be investigated on the overall proficiency level of Iranian EFL learners.

- This study can be completed by using a control group (with no test treatment period) to check whether there is any significant difference among the students' mean scores in any of the three groups.

At the end, the researchers hope that this study would be beneficial to both teachers and learners and help them improve their knowledge and experience in both fields of teaching and testing.

## References

[1]. Alderson, J. C. (1978). 'The Effect of Certain Methodological Variables on Cloze Test Performance and its Implications for the Issue of the Cloze Procedure in EFL Testing. Paper presented at the Fifth International Congress of Applied Linguistics, Montreal.

[2]. Bachman, L. F. (1982). 'The Trait Structure of Cloze Test scores'. TESOL Quarterly, 1, 61-67.

[3]. Best, J. W., & Kahn, J, V. (1989). *Research in Education* (6th ed.). New Delhi: Prentice Hall, Inc.

[4]. Brown, J. D. (1978). '*Correlational Study of Four Methods for Scoring Cloze Tests*'. Master's thesis, University of California, Los Angeles.

[5]. Brown, J. D. (1983). 'Conventional Cloze Tests and Conversational Ability'. *ELT Journal* 37(2): 158-161.

[6]. Brown, J. D. (1988). 'Tailored cloze: improved with classical item analysis techniques'. *Language Testing, 1,* 19-30.

[7]. Bourke, J. M. (1989). 'The grammar gap'. *English Teaching Forum,* 27 (3), 20-23.

[8]. Harris, D. P. (1968). *Testing English as A Second Language*. Mc Graw Hill, New York.

[9]. Heaton, J. B. (1988). *Writing English Tests*. Longman Handbooks for Language Teachers.

[10]. Jones, Randall. (1977). 'Testing: a vital connection', 237-261 in June K. Phillips (ed.). *The Language Connection:* From the Classroom to the World. Skokie, Illinois: National Testbook Company.

[11]. Mousavi, A. (1997). *A Dictionary of Language Testing*. Tehran: Rahnama Publications.

[12]. Oller, J. W. Jr., & Perkins, Kyle. (1978)*. Language in Education: Testing The Tests*. Rowley, MA: Newbury House.

[13]. Oller, J. W. Jr. (1979). *Language Tests at School: A Pragmatic Approach.* New York: Longman Group Ltd.

[14]. Richards, J. C., Platt J., & Platt, H. (1992). *Dictionary of Language Teaching and Applied Linguistics.* New York: Longman Group Ltd.

[15]. Rutherford, W. (1987). *Second Language Teaching and Learning*. London: Longman.

## ABOUT THE AUTHORS

*Minoo Alemi is the Faculty member of Languages and Linguistics Department at Sharif University of Technology, Iran. Her main areas of research interest in the field of Language Teaching are Second language Acquisition, ESP, Discourse Analysis, and Syllabus Design. She has published about 16 textbooks in General English and ESP, a large number of papers in different areas in International Journals, and given presentations on TEFL at many International Conferences. She is also the member of editorial committees of a few International Journals.*

*Aparna Miraghaee is an M.A holder of TEFL teaching at Allame Tabatabaee University, Tehran, Iran. She has more than 15 years of experience as an EFL teacher teaching at Iranian Kids English Departments in Tehran, Iran. She has also been teaching in Azad University West-Tehran and Sama Branches for about six years. Her main areas of research in the field of Language Teaching are Teaching Methodology, Testing and ESP. She has published a book on the assessment of Knowledge of Grammar with Dr Minoo Alemi and is planning to continue her research on both fields of teaching and testing EFL.*