The Relationship between the Cloze Testing Procedure and the Skills Tests of EFL: A Correlational study

Ibrahim Saleh Al-Fallay*

* Ph.D. in Applied Linguistics, University of New Mexico, U.S.A, 1994.
* Assistant Professor of Applied Linguistics, King Saud University - College of Arts English Department.

Abstract

Since the cloze testing procedure was first used by Taylor in 1953 as a means of providing a readability index of passage difficulty level for native speakers, much research has been conducted on it. Some researchers, for instance, have used cloze tests not only with native speakers, but also with non-native speakers. Others have investigated the validity, reliability and practicality of cloze tests. Still others have tried to find a relationship between the cloze testing procedure, and well known proficiency and placement tests. This study, however, treats a different, but important issue. It investigates the relationship between cloze tests and tests of actual courses taught in the Intensive English Program (IEP) of the English Language Center (ELC) of the Institute of Public Administration (IPA) in Saudi Arabia. The rationale is that such research will be more practical and will yield results and recommendations which could be used in English as a Foreign Language (EFL) classrooms. For this purpose 432 students enrolled in the IEP of the IPA, who represented four different proficiency levels were selected. Subjects were asked to perform a completion task for fifty item cloze tests. Their scores on other measurements, tests of grammar, reading, writing, speaking and listening, were also obtained. Statistical analysis was conducted and the results of this study are in harmony with previous research. There was a significant relationship between cloze tests and tests of the EFL skills and subjects. It was also found that the relationship between cloze and other tests was stronger in the case of advanced students. This study concludes that the cloze testing procedure is a promising assessment tool in the EFL classroom. Cloze tests could be short and used for diagnostic purposes, as well as be major tests of aptitude, achievement and proficiency. Although the cloze testing procedure seems useful at any proficiency level, the results are more reliable when learners are of a high proficiency level.

I. Introduction

Language assessment is considered the criterion against which teaching methodology and procedures, textbooks, and learners' achievements are tested. It is a major step in the teaching and learning processes. With this in mind, a great amount of research is needed in order to discover some of the crucial elements of the field of language testing, which can later be utilized to improve and develop that field, and then the whole learning process. Language teachers generally use evaluation tools to measure how much the students have gained after being introduced to a lesson or a unit of their textbooks. Moreover, the tools are also used to get a clear picture of students' overall proficiency level, or to measure their readiness and aptitude to absorb more information.

The aim of this study, then, is to present the cloze testing procedure and suggest it as an evaluation tool which can be used in a continuum fashion, ranging from measuring students' mastery of specific elements such as a short vocabulary list or a few grammatical structures, to evaluating the overall proficiency of learners. The issues of the cloze tests' validity, reliability and practicality will not be discussed here, since a great deal of studies have discussed these issues.[1] The scope of this study will be limited to investigating whether cloze tests will give similar results to those obtained by other tests used to measure language skills and subjects, such as listening, reading, writing, speaking and grammar. If this study demonstrates that cloze tests are related to these tests which measure various aspects of language, then the possible interpretation is that the cloze testing procedure measures the same aspects of language skills as done by these different tests. This will lead to the consideration of cloze tests as a quick exam to diagnose certain weaknesses and strengths of the students, and test their achievement in specific elements of language, as well as a major test which truly determines learners' proficiency.

II. Review of selected literature

The cloze testing procedure was originally used by Wilson Taylor (1953) to determine the difficulty level of prose passages for native English speakers.[2] The average score of a language group of subjects provides a readability index of passage difficulty level. The term "cloze" was coined by Taylor as a spelling corruption of the word "close". It originates from the notion of closure in "Gestalt Psychology" and is used to refer to the human psychological tendency to fill in gaps in a given text. The cloze procedure is based on the assumption that given contextual clues of a text, a native speaker of the language or reasonably proficient non-native speaker, say, ESL student should be able to anticipate the words that belong in the blanks. Taylor described his procedure as follows:

> A cloze unit may be defined as: Any single occurrence of a successful attempt to reproduce accurately a part deleted from a "message" (any language product), by deciding from the context that remains, what the missing part should be. (1953:416)[3]

Since then, the cloze test has been under extensive investigation in order to check its reliability in assessing language proficiency. Carroll et al., (1959) used an English cloze test to compare the performance of adult native speakers of English to that of adult English-French and English-German bilinguals.[4] They concluded that the cloze test was not able to either assess language proficiency or to distinguish between a German speaker or French. In addition, Alderson (1980) stated that the cloze test was not sensitive to differences in the performance of native and non-native speakers.[5] Other criticisms of the cloze test were cited by Hughes (1989, 1990), Davies (1990a, 1990b), Spolksy (1990), and Weir (1988).[6] Hughes (1989), for instance, deemed that "... cloze could not deliver all that was promised on its behalf... [and] it is not possible to predict accurately from this what is people's ability with respect to the variety of separate skills (speaking, writing etc.)..." (1989:65).[7]

This view, however, has not been held by all researchers in the field of language testing. There are other testing specialists who do believe in the importance of the cloze test. Indeed, it is believed that the cloze test can assess language proficiency whether it is the first, second, or third language proficiency being assessed.[8] Others believe that the cloze test is so powerful that it will yield similar results even if the text is a translation from another language.[9] In addition, there is a high positive correlation between cloze test scores and those of discrete point tests.[10] Moreover, the claims of Carroll et al., (1959) were refuted by Briere and Hinofotis (1979) who demonstrated that the cloze test is capable of distinguishing between first, second, and third semester students studying German, Japanese, French, and Russian.[11]

Concerning the length of the test, Taylor (1956) claims that a 50 item test provides a sufficient sample for a stable score with sufficient provision for easy and hard words to cancel each other out.[12] Bormuth (1964), however, feels that even with 50 item tests, there were significant differences between mean scores for over half of twenty cloze test forms.[13] He notes that the mechanical selection of words for deletion tends to produce a number of non-discriminating items which lower reliability. As a result of Bormuth's conclusion it is suggested that cloze passages be long, at least for testing purposes.

Two methods have been widely used to score the cloze test namely, the exact, and the acceptable word criterion. The former allows for counting as corrects, only the exact words which have been deleted from the text. The latter accepts both the exact as well as any grammatically and contextually appropriate words. In 1969 Darnell developed a cloze test scoring method called clozentrophy and provided a measure of comparability of foreign students' language patterns with those of their native English speaking peers.[14] The method reveals absurd results, and because of its impracticality and complexity, it has hardly been used or referred to. Another cloze test scoring method is multiple choice scoring, involving a choice of words provided for each blank, where students have to select the appropriate word/response. On the

other hand, Oller (1972) applies five methods to score cloze tests: (i) the exact word criterion; (ii) acceptable word criterion; (iii) responses violating long-range constraints; (iv) responses violating short-range constraints, and (v) entirely incorrect fill-ins or items left blank, indicating complete lack of comprehension.[15] He concludes that the best cloze scoring method is the acceptable-word criterion, that is the one that counts as correct any contextually acceptable responses, even though the exact-word criterion is superior in terms of item discrimination and validating corrections regardless of the level of difficulty of the test.

Hinofotis (1976) supports Oller's conclusion by asserting that the acceptable-word scoring method yields more reliable scores and provides more accurate information about ESL [or any other second language] proficiency levels.[16] Nevertheless, it was noted that there seems to be very little difference when the exact-word method is substituted for the acceptable-word method.[17] This thus makes the reasons for choosing between the two methods unclear. After analyzing and comparing Oller's five cloze scoring methods, Brown (1978) asserts that the decision about which method to use should be dictated by the testing situation.[18]

Reporting on the pertinence/accuracy of the cloze procedure, many studies indicate that this procedure correlates with measures of English as a Foreign Language (EFL) proficiency.[19] Oller (1976a) asserts that this evidence seems to indicate that a cloze test score maybe more representative of general language ability than of single skills.[20] Prior to the above evidence Oller also found that cloze tests relate more to dictation and reading comprehension tests than to "discrete point" tests of grammar and vocabulary.[21] He thus concluded that cloze and dictation [as well as reading comprehension] were integrative tests, and so very useful measures of global skills. In 1979 Oller wrote that pragmatic tests (such as cloze, dictation, reading comprehension and oral interview) challenge the students' expectancy grammar; and thus force them to mobilize their knowledge of the language system and apply it to a given social situation.[22]

Oller (1983a) used a total of 376 foreign students attending the University of California in Los Angles (UCLA) to investigate the relationship between cloze tests and standardized tests used by UCLA (UCLA ESLPE 2A Revised test) to screen them. He was also interested in investigating the role of the difficulty level of the cloze tests. Hence he assigned the subjects to three conditions: Easy cloze test (N = 129), medium cloze test (N = 128) and difficult cloze test (N = 119). All subjects also took the standardized test. Oller pointed out that the correlation between cloze tests and tests of vocabulary, grammar, reading, and dictation in the UCLA ESLPE 2A Revised test was statistically significant.[23] In fact, the correlation was significant regardless of the difficulty level of the cloze tests. The results of the correlation between the various cloze tests and the four components of UCLA ESLPE 2A Revised test, namely vocabulary synonym matching, grammar, reading comprehension and dictation tests are summarized as follows:

Table (1)

	Vocabulary	Grammar	Reading	Dictation
Cloze (easy)	0.55	0.70	0.65	0.73
Cloze (med.)	0.69	0.66	0.70	0.78
Cloze (diff.)	0.71	0.76	0.72	0.78

Source: Oller 1983a:8-10

Alderson (1983) compared the performance of 360 non-native speakers of English seeking admission to British and Scottish universities in cloze tests, with their performance in standardized tests of proficiency of English as a foreign language (ELBA) used by several of those universities to screen their prospective foreign students. He reached almost the same conclusion regarding the relationship between the cloze test and other discrete point tests.[24] By using the same design as Oller's, all correlations were significant at $\alpha < 0.01$ level. In addition, Alderson's study contributes more to the question at hand. For example, he tested the correlation between the cloze test and four other tests which could be included in the phonology panel. A partial duplication of his study is given below:

Table (2)

Exact word scoring	**Cloze (Diff.)**	**Cloze (Med.)**	**Cloze (easy)**
Sound Recognition	0.36	0.49	0.47
Intonation	0.49	0.36	0.47
Sentence stress	0.25	0.36	0.29
Listening Comp.	0.55	0.56	0.63
Acceptable Word scoring			
Sound Recognition	0.50	0.57	0.55
Intonation	0.57	0.47	0.53
Sentence stress	0.30	0.43	0.45
Listening comp.	0.67	0.70	0.70

Source: Alderson 1983:208

As was mentioned before regarding the scoring criterion, the correlation coefficients are highly significant in both criteria. We can conclude that the cloze test is potentially able to assess linguistic competence. Other correlations

between the cloze test and tests of grammatical vocabulary yield almost the same results as Oller's.

Shohamy (1983) used a sample of 106 students attending Hebrew classes at the University of Minnesota in an attempt to investigate the reliability of an oral interview test. She reported significant correlation coefficients between cloze, regardless of used scoring procedure, and components of the oral interview test.[25] She demonstrated that the correlation between cloze tests and vocabulary, grammar, pronunciation and fluency scores were significant as shown in table (3):

Table (3)

	Easy Cloze (Acceptable word)	Diff. Cloze (Accept able Word)
Vocabulary	0.80	0.82
Grammar	0.82	0.86
Pronunciation	0.74	0.78
Fluency	0.77	0.78

Source: Shohamy 1983:234

In her conclusion, she mentioned this positive high correlation and states that "The findings also suggest a high concurrent validity with cloze tests..." (1983:234).[26] Devine (1987) also concluded that, "Gains in language competence as measured by holistic tests correlate positively with increasing reading proficiency." (1987:83)[27]

One version of the cloze procedure is the multiple choice (MC) cloze test. In such a test, examinees are supposed to select from among a number of alternatives the choice which best fills in the blank. Hinofotis and Snow (1978) reported a significant correlation between the MC cloze test and the traditional cloze procedure.[28] Therefore, it seems that both of these tests have something in common, which is their assessment of the overall language proficiency of native speakers, and second/foreign language learners. They also reported significant correlations, ranging from 0.63 to 0.71, between the cloze tests and ESL placement tests in Southern Illinois University. Similar results were obtained by Brown who reported a correlation coefficient of 0.90 between UCLA ESL placement tests and completion cloze tests using the acceptable-word criterion for scoring, and a correlation coefficient of 0.89 between MC cloze tests and the placement tests.[29]

While investigating the relation between cloze tests and the Test of English as a Foreign Language (TOEFL), Pike (1979) reported a significant correlation of $\alpha < 0.05$ among scores on the different parts of the TOEFL and the MC

cloze tests.[30] A more comprehensive study related to the relationship between the scores of the TOEFL supposedly assessing various skills, and cloze tests was presented by Hale et al., (1989).[31] For the purpose of the study, they presented cloze tests as part of the TOEFL which was administered to 11,290 examinees from nine countries. The cloze test consisted of fifty items divided into four parts designed to assess a combination of skills. These parts were a combination of reading and grammar, reading and vocabulary, grammar and reading, and vocabulary and reading with the first mentioned skill or subject being the primary focus of the assessment. The parts of the TOEFL, as might be known, are designed to assess listening comprehension, grammatical structure, written expressions, vocabulary, and reading comprehension. Since there was a discrepancy between the number of cloze items and the number of TOEFL parts, they used the Spearman-Brown formula to check for reliability. They reported a high correlation which was significant at $\alpha < 0.001$ between the cloze test and the total of the TOEFL scores which ranged from 0.86 to 0.94, with a median correlation of 0.89.

The reported correlation coefficients among the parts of the cloze test and the scores of the various parts of the TOEFL deserve thorough consideration. In the case of listening comprehension, the highest correlation was 0.91 for the reading/grammar part of the cloze test, and the lowest was 0.67 for listening comprehension and reading/vocabulary. For the grammatical structure part, the highest correlation coefficient was with reading/grammar which reached 0.99 and the lowest correlation was 0.76 for reading/vocabulary. There were also high correlations between written expressions and reading/grammar where the highest correlation was 0.99 and the lowest was 0.75 between the written expression part and the part of reading/vocabulary. In the case of the vocabulary assessment part, the highest correlation was 0.99 and it was with reading/vocabulary, and the lowest was that of the vocabulary and grammar/ reading which was 0.72. Finally, the highest correlation of reading comprehension was with reading/vocabulary, and the lowest was with vocabulary/reading at 0.78. In the discussion of their results, they contended that "... the relationship between these two measures [cloze test and TOEFL] was quite high, indicating a considerable amount of overlap in the abilities measured by them." (Hale et al., 1989:60).[32]

III. The present study

A. The research problem

As previously mentioned, second/foreign language instructors always need testing tools to enable them to objectively and comprehensively evaluate their learners, the teaching process in general, and other related matters. This

evaluation is not limited to tests used to check the final outcome of the learning process, which are usually thought of as finals and midterms. In fact, tests which are administered during the learning process in the form of short quizzes seem to be more important since these not only work as an information source for the final outcome, but also provide a means of diagnosis and treatment. For example, their use keeps instructors up-dated regarding the current proficiency level of learners and the obstacles which they may face. Unlike finals and midterms, which are long and time consuming, quizzes should be shorter in both content and administration time. This has led to the search for a testing instrument of reported validity and reliability and which has a strong relationship with tests used to separately assess the learners' command of language skills and subjects. This relationship, if found, indicates that this testing instrument measures the same elements tested by various other tests.

This study, then, considers whether the cloze testing procedure could be a candidate which may achieve the aforementioned goals. A great amount of research has recently been conducted on cloze testing, either to demonstrate its validity and reliability, or to discover any relationships between the cloze procedure and other tests such as the TOEFL, and placement tests used in the American universities. They have also been used to investigate the sensitivity of the cloze procedure to contextual and textual constraints. The subjects used in many of these studies have been native speakers of English, or foreign students of college level, learning and speaking English as a second language. The research has been conducted in the United States and Britain. Hence, there is a need for research in a different setting. The goal of this study is to investigate the relationship between cloze tests and actual tests of reading, writing, listening, speaking, and grammar skills and subjects used in an EFL program. If a significant correlation is found, this will suggest that all of these tests, including cloze tests, share a general underlying factor and that cloze testing is capable of assessing learners' skills now measured by more specific tests. The present study differs from previous research in that it is conducted in Saudi Arabia where English is a foreign language. Moreover, the subjects of this study are high school graduates who learned English as a school subject and are now learning it in the Institute of Public Administration as a foreign language. Another point of interest is that where previous studies used subjects of equal, or at least believed to be so, language proficiency, this study used subjects who differ in their command of English. In other words, the subjects of this study came from four different proficiency levels. This difference in subjects and location may yield different yet significant results compared to the results of the previous studies.

B. The hypotheses

The hypotheses of this study were set to be nondirectional to capture any significance through indicating both positive and negative relationships. The study was designed to test the following hypotheses:

H1: There is a significant statistical correlation of $\alpha < 0.05$, between the cloze test and other tests of English as a Foreign Language.

H2: There is a significant difference in the correlation values between the cloze test and each test of English as a Foreign Language (skills and subjects) as the level of learners' proficiency changes at $\alpha < 0.05$.

C. Methods & Procedures

a) Subjects

The subjects of this study were 432 Saudi Arabian students who were high school graduates enrolled in Private Sector programs in the headquarters of the Institute of Public Administration in Riyadh, Saudi Arabia. These programs lead to diplomas in fields such as insurance, banking, sales, and accounting. As part of these two and a half year programs, students must spend one full academic year in the Intensive English Program offered by the English Language Center of the Institute. The sample consisted of 108 subjects from each level of the program's four levels and hence their English proficiency levels varied. Data was collected during the training years of 1993/1994 and 1994/1995. Table (4) shows the distribution of subjects according to their levels and programs.

Table (4)
The distribution of subjects according to levels and programs

Program	Insurance	Banking	Sales	Accounting
Level				
1	22	40	17	29
2	26	31	24	27
3	31	13	40	24
4	26	27	27	28

b) The Intensive English Program (IEP)

The purpose of the IEP of the English Language Center is to improve the students' command of English in order to enable them to pursue their education through various programs of the Private Sector Department which are taught in English. The IEP has four levels: preparatory, elementary, intermediate, and

advanced. It teaches five courses, which are the four skills of English writing, reading, speaking, and listening, in addition to grammar. The English Language Center uses the quarter system where the academic year is divided into four sessions, each lasting for nine weeks. Students have to attend the IEP for 24 hours per week. These 24 hours are divided as follows: 6 hours for grammar, 5 hours for writing, 5 hours for reading, 5 hours for speaking, and 3 hours for listening. The highest mark in any course is 100 points, with 90 points used for quizzes, a midterm, and a final, whereas the remaining 10 points are assigned for participation and homework. By the end of the ninth week of each session, students are tested in the foregoing courses in order to determine the learners' readiness to move to the next level. The learners pass the level and are eligible to move to the next one if they receive at least 60 points in all courses. If they get less than 60 points in any course, they are allowed to repeat the level only once during the whole program.

c) Materials

The materials used in this study are of two kinds. First, subjects' scores on the various tests of the IEP were collected, and then points of participation and homework were deducted. The scores after deduction are the actual scores of learners on language tests, without being contaminated by the points of participation and homework. Hence, the highest mark became 90 points. After that, each score was multiplied by 1.11 in order to transform it to a scale of 100 points and also to avoid any unnecessary correction for reliability. The second type of material was four reading passages of different levels of difficulty, adapted from four different sources. The reason for varying the difficulty level was to make the passages compatible with the students' current levels of proficiency, since it was reported in previous research that highly proficient students usually use contextual and textual information and hence score better than students with low proficiency levels. Four cloze tests, one for each level, were then constructed, following the usual procedure of deleting every fifth word and replacing the removed words with blanks of uniform length. The first and last sentences were left intact and each passage made a cloze test of 50 items.[33]

d) Procedures of Administration & Scoring

Tests of cloze procedure were administered during the eighth week of randomly selected sessions, that is one week before the final examination. Subjects were allowed 90 minutes to complete the cloze tests. The acceptable-word criterion was used in scoring. A value of (0) was assigned to each wrong response and (2) points were granted to each acceptable answer with the highest mark being 100 points. This was done in order to have an identical scale which ranged from 0 to 100 points for all the cloze and the IEP tests. After

the final examinations of the IEP were administered and learners' scores became available, these scores were collected and transformed as mentioned in the previous section.

e) Statistical techniques & formulae

The data was entered in an Apple Macintosh LCII computer for statistical analysis and a statistical software package, Statistica, was used. In addition to the calculation of the mean and the standard deviation, the simple Product Moment Correlation (Pearson) (r) was calculated between the cloze tests and the IEP's subjects and skills tests for each level in order to test the first hypothesis. Then, t values were calculated to test ρ, the estimate of the population correlation coefficients based on the values of the sample correlation coefficients (r), between the cloze tests and the IEP's subjects, to see if ρ equals or does not equal (0) (see note 36 for formula and explanation). The confidence intervals for ρ were also calculated (see note 37 for formula and explanation). Finally, the differences between correlation coefficients of the sample and the population, between subjects' scores on cloze test (for each level separately), and the scores of the tests of grammar, reading, writing, speaking and listening were calculated (see note 38 for formula and explanation).

D. Results and Discussions

A summary of mean and standard deviations for the students' scores (N = 108 for each level) on the IEP tests are given in table (5):

Table (5) Mean and standard deviations of IEP tests

	grammar		reading		writing		speaking		listening	
	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s
level										
1	70.08	16.61	72.91	14.41	66.62	21.12	70.41	15.95	70.07	15.66
2	73.99	13.98	73.90	12.04	72.13	15.04	73.08	11.33	73.32	11.62
3	69.80	14.46	70.35	12.41	75.59	13.23	73.48	11.15	72.48	11.28
4	70.29	13.58	74.94	09.15	74.95	11.70	76.10	09.39	74.17	09.41

N = 108 for every level with a total of 432 subjects. where $\bar{x}$ is the mean and s is the standard deviation.

Since the aim of the IEP tests is to check learners' readiness to move to the next level, each level has its own tests which are based on materials presented in class and that is why they take the form of achievement tests. Moreover, this explains why intermediate and advanced students' scores resembled those of the preparatory and elementary levels. In table (5), the most

important point worth mentioning is the consistent decrease in the standard deviation values as the proficiency levels of learners increase, with the exception of the standard deviation values of grammar, reading, and writing of the second level. The huge difference in standard deviation values between the first and fourth levels indicates that as learners improve in their command of English, they move towards a desired normalization. In other words, the advanced learners form a more homogeneous group than those of the lower levels. Since this issue is beyond the scope of this paper, no further statistical analysis regarding the difference in standard deviation values will be offered here, although this is an interesting question which could be tackled in future studies.

Table (6) also gives mean and standard deviations for the cloze tests, in addition to reliability coefficients. Reliability was calculated by using the Kuder-Richardson formula 21 (KR-21) which gives conservative estimates of reliability.

Table (6) Mean, standard deviations and reliability coefficients for the four different cloze tests

Level	$\bar{x}$	s	KR-21
1	41.70	15.02	0.84
2	56.22	14.53	0.87
3	60.32	09.75	0.86
4	72.76	08.24	0.89

Since the four cloze tests are considered tests of overall English proficiency, the difference in the mean among the four groups is justifiable. Although subjects had different cloze tests which varied in their reading difficulty levels and which seemed to be compatible with their current levels of proficiency, advanced students scored better than the other three groups and this agrees with the conclusions of Oller et al., that advanced ESL learners usually use contextual and textual information to fill in the gaps.[34] Moreover, the difference in the values of standard deviation among the four groups becomes even clearer. As proficiency levels of learners increase, the standard deviation value decreases, and this is in harmony with the results of Al-Fallay's study of 1994.[35]

To test the first hypothesis set by this study, Pearson product-moment correlation coefficients among the cloze tests and the tests of IEP's subjects were calculated. These coefficients are showed in table (7).

Table (7) Correlation coefficients among cloze tests and tests of the IEP

Cloze & level	grammar	reading	writing	speaking	listening
1	0.73	0.71	0.65	0.63	0.61
$\alpha <$	.0001	.0001	.0001	.0001	.0001
2	0.75	0.76	0.72	0.72	0.67
$\alpha <$	.0001	.0001	.0001	.0001	.0001
3	0.78	.76	.82	0.84	.073
$\alpha <$	.0001	.0001	.0001	.0001	.0001
4	0.90	0.90	0.88	0.84	0.78
$\alpha <$	.0001	.0001	.0001	.0001	.0001

All coefficients are significant at $\alpha < 0.05$. In fact all correlation coefficients reported in the table are significant at $\alpha < .0001$.

It is clear from table (7) that the correlation coefficients ranged from 0.90, the correlation value between reading and cloze tests for the fourth (advanced) level, to 0.61, the correlation of cloze and listening tests in the first (preparatory) level. All observed correlations are statistically significant at $\alpha < 0.05$. Hence we may conclude that according to our sample level, there is a relationship between the cloze testing procedure and various other tests assessing EFL skills and subjects. The interest of this study is not limited to the sample used, but rather directed toward the population of EFL Arab learners. This population could be defined as those Arab students who having completed six years of studying EFL in Saudi public schools, are now enrolled in intensive programs of EFL. The limits of the population may also be extended to include Arab students learning EFL, who share similar background knowledge with the Saudi students, and who attended and are now attending similar learning programs. In order to generalize to the aforementioned population where Arabic is the L1 possible, additional statistical tests have to be carried out. The first step in these additional statistical tests was to determine if ρ, the estimates of the population correlation coefficients based on the values of the sample correlation coefficients (r) between the cloze tests and the IEP's subjects, equal or do not equal (0). This was done by calculating (t) values based on the values of (r) as is shown in table (8).[36] If the values of (t) equal or exceed 1.96, a conclusion is made that the value of ρ does not equal (0) at $\alpha < 0.05$, two-tailed.

Table (8): t values of the correlation coefficients among cloze tests and tests of the IEP

level	cloze & df	grammar	reading	writing	speaking	listening
1	107	10.96	10.38	08.90	08.33	07.83
	α <	.0001	.0001	.0001	.0001	.0001
2	107	11.75	12.20	10.78	10.71	09.19
	α <	.0001	.0001	.0001	.0001	.0001
3	107	12.93	12.15	14.75	15.94	10.90
	α <	.0001	.0001	.0001	.0001	.0001
4	107	20.79	21.40	19.26	15.89	12.93
	α <	.0001	.0001	.0001	.0001	.0001

All values of t are statistically significant at $\alpha < 0.05$, two-tailed. In fact all t values are significant at $\alpha < .0001$.

As table (8) indicates, all values of t exceeded the critical value of 1.96. Hence we may conclude that the value of ρ does not equal (0) and that there is a relationship between the cloze tests and the tests of EFL subjects on the population level. In order to construct a confidence interval which would include the values of ρ, all r values were transformed to Fisher Z'. Then the upper and lower limits were calculated.[37] These lower and upper limits were as follows:

Table (9): Lower and upper limits of the correlation coefficients for the population among cloze tests and tests of the IEP

cloze &	grammar		reading		writing		speaking		listening	
level	LL	UL	LL	UL	LL	UL	LL	UL	LL	UL
1	0.63	.080	0.61	0.79	.053	0.75	.50	0.73	0.48	0.72
α <	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0002	.0001
2	0.65	0.82	0.67	0.83	0.62	0.80	0.62	0.80	0.55	0.76
α <	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001
3	0.70	0.85	0.67	0.82	0.75	0.87	0.78	0.89	0.74	0.80
α <	.0001	.0001	.0001	.0001	.0001	.0001	.0001	,0001	.0001	.0001
4	0.86	0.93	0.83	0.93	0.83	0.92	0.78	0.89	0.70	0.85
α <	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001

All coefficients are significant at $\alpha < 0.05$. In fact all correlation coefficients reported in the table are significant at $\alpha < .0001$.

As table (9) indicates, all lower and upper limits of the correlation coefficients between the cloze testing procedure and the tests of EFL skills and subjects for the population are statistically significant at $\alpha < 0.05$. (In fact they are significant at $\alpha < .0001$). The value of the correlation ranged from 0.49 as the coefficient of the cloze test and the test of listening for the preparatory level, to 0.86, the correlation coefficient between the cloze test and the test of reading for the advanced level. The correlations mentioned are for the cloze testing procedure and the tests of EFL skills and subjects for the population of Arab students learning EFL. Therefore, the first null hypothesis is rejected and the alternative hypothesis is accepted and it may be concluded that there is a statistically significant relationship between the cloze testing procedure and the various tests assessing different skills and subjects of EFL. It seems that the cloze testing procedure is capable of assessing these skills and subjects regardless of their modalities. One cloze test was actually able to determine the proficiency level of learners of different levels.

In order to test the second hypothesis of this study, the values of Z(r1-r2) for the correlation coefficients of the sample were calculated.[38] This indicates whether the correlation coefficients among the cloze test and the other tests differ with the level of students' proficiency. In other words, the hypothesis tries to determine whether the correlations for the advanced level are statistically higher than those of other levels. The values of Z (r1-r2) are given in table (10). Since the difference direction is known, the hypothesis was set as nondirectional. The values of Z(r1-r2) should equal or exceed 1.65 in order that the difference be statistically significant at $\alpha < 0.05$, one-tailed.

Table (10): Values of Z (r1 - r2) for the sample

cloze and levels	grammar	reading	writing	speaking	listening
1&2	0.32	0.88	0.97	1.21	0.73
$\alpha <$	.63	.27	.27	.15	.28
1&3	0.84	0.88	2.70*	3.48*	1.57
$\alpha <$	.28	.27	.005	.0004	.06
1&4	3.75*	4.24*	4.29*	3.48*	2.49*
$\alpha <$	.0002	.00003	.00003	.0004	.009
2&3	0.52	0.00	1.73*	2.27*	0.84
$\alpha <$	.25	0.00	.05	.02	.28
2&4	3.44*	3.36*	3.32*	2.27*	1.76*
$\alpha <$	.0004	.0005	.0005	.02	.05
3&4	2.91*	3.36*	1.59	0.00	0.92
$\alpha <$	.0025	.0005	.06	0.00	.28

Correlations significant at $\alpha < 0.05$ one-tailed

It is clear from the table that there is no significant difference among the correlations of the first and second levels. There was also no significant difference between the correlations of the second and third levels, except the difference in the correlation coefficients of cloze, writing and speaking tests. In the case of the difference between the correlations of the first and third levels, the differences between the cloze and the skills and subjects tests were only significant when the correlations were between cloze and writing tests, and cloze and speaking tests. The difference between the first, second and fourth levels was more interesting. The correlations of the cloze test with the other tests were significantly higher in the advanced level than in the preparatory and elementary levels. Moreover, the correlations of the advanced level were significantly higher than those of the second, elementary level. Finally, only the correlation between cloze and reading tests was significantly higher for the advanced level than for the third level. As a conclusion, it seems that the advanced students used the text's information and employed their skills in completing the cloze test.

In order to generalize the foregoing results to the population of Arab students learning EFL, the Z (ρ LL1 - ρ LL2) and Z(ρ UL1 - ρ UL2), that is the difference between the lower limits and the upper limits for each correlation of the population between cloze and subjects tests for each level as obtained from table (9), were calculated.[39] Table (11) gives these values.

Table (11) Values of Z (ρ_1 - ρ_2) of lower and upper limits for the population

cloze and	grammar		reading		writing		speaking		listening	
levels	LL	UL	LL	UL	LL	UL	LL	UL	LL	UL
1&2	0.29	0.44	0.72	0.87	1.01	0.93	1.30	0.94	0.72	0.62
α <	.63	.25	.28	.27	.28	.28	.10	.27	.28	.25
1&3	0.94	1.17	0.72	0.65	2.75*	2.61*	3.62*	3.55*	1.59	1.37
α <	.28	.14	.28	.25	.005	.006	.0003	.0004	.06	.09
1&4	3.99*	4.07*	4.20*	4.28*	4.35*	4.49*	3.62*	3.55*	2.54*	2.54*
α <	.0001	.0001	.0001	.0001	.0001	.00003	.0003	.0004	.009	.009
2&3	0.65	0.72	0.00	-0.22	1.74*	1.67*	2.32*	2.33*	0.87	0.75
α <	.28	.25	0.00	.60	.04	.05	.02	.02	.27	.25
2&4	3.70*	3.62*	3.48*	3.41*	3.33*	3.56*	2.32*	2.33*	1.81*	1.91*
α <	.0003	.0003	.0004	.0004	.0005	.0004	.02	.02	.04	.03
3&4	3.04*	2.90*	3.48*	3.62*	1.59	1.88*	0.00	0.00	0.94	1.17
α <	.002	.0025	.0005	.0003	.06	.03	0.00	0.00	.28	.14

Correlations significant at α < 0.05 one-tailed

Table (11) shows the values of Z (ρ_1 - ρ_2) for lower and upper limits. The same observations mentioned regarding the values of Z (r1-r2) of the sample are true here. The differences between the first and fourth levels and the second and fourth levels are also interesting since they show the sharp contrast between the low and high proficiency learners. Moreover, the difference between the third and fourth levels in the correlation of cloze and reading tests indicates that advanced learners were capable of dealing with the cloze test as they used reading strategies and knowledge of grammar to fill in the gaps. In addition, only the difference between the upper limits of the correlation between cloze and writing was significant. There were also no significant differences between the lower and upper limits of the first and second levels. Only the differences between the limits of the correlations of the first and third levels between cloze and writing, and cloze and speaking were significant. There were also no significant differences between the correlations of the first and second levels. Moreover, only the limits differences between cloze and writing and cloze and speaking were significant in the case of the comparison between the second and third level. Therefore, it is possible to reject the second null hypothesis of this study and conclude that the correlation between cloze tests and the EFL subject tests of advanced students is significantly higher than those of lower level learners which shows that cloze testing is capable of assessing learners' proficiency especially in advanced levels. In addition, this demonstrates that higher proficiency students are using their reading strategies, and knowledge of grammar, and learned language skills to deal with tasks. This seems to assess the learner's overall language proficiency. This kind of knowledge seems to be unavailable to lower proficiency students.

E. Conclusions and Recommendations

The results reported in this study are clear. The Cloze testing procedure is a promising assessment tool. It has a variety of uses since it significantly correlates with other tests used to assess the different skills and subjects usually taught in EFL institutes. This correlation indicates two points. First, cloze tests will yield almost similar results to those of tests of grammar, reading, writing, speaking, and listening. Second, cloze testing is actually assessing the underlying factors which control language proficiency levels. With this in mind, it seems that cloze tests can be used as short diagnostic tests which will assist the teacher in determining the students' strengths and weaknesses. This will help in any necessary amendment or modification in the materials used, or in teaching procedures. Moreover, cloze testing can be used as an instrument to determine the aptitude, proficiency level, and the achievement of learners.

The Cloze testing procedure also has more advantages than the currently used tests. First, it is easy to construct and administer. The teacher can adopt

any suitable text as a cloze test and this eventually saves the teacher time. Cloze tests are also interesting. For example, the teacher can use narrative texts related to the students' background and culture. This, in fact, has showed that learners do not only score better on such texts, but also improve their overall language proficiency.[40] Furthermore, these texts and the constructed cloze tests differ from currently used tests which usually take the form of unrelated vignettes in that they present continuous and relevant information.[41] Moreover, cloze tests challenge the learners and motivate them to interact with the text since they view the examinee as active and productive.

The difference between correlations either on the sample or population level is also of interest. It seems that high proficiency students were able to employ their knowledge of English structure and the language skills they learned in dealing with the cloze testing procedure. Lower proficiency students, however, were not able to use such resources. This could be ascribed either to their lack of such abilities, or their inability to invest those resources. The generalization to the population should be carried out carefully. The outcomes of this study might only be applicable to the population of Arab students learning EFL in intensive English programs, and who completed six years of learning English as a subject in public schools.

With these advantages and the reported significant correlations of this study, we can conclude that the cloze testing procedure is a test which assesses all modalities, and which has various uses ranging from short diagnostic tests to major tests which determine the proficiency level of learners. Although cloze testing seems to better reflect the proficiency of advanced learners, it is also useful for learners of lower proficiency levels.

Notes and references

1. For a comprehensive review refer to

J. Brown, "What are the characteristics of natural cloze tests", **Language Testing,** 10, (1993), pp. 93-116.

J. Brown, "A closer look at cloze validity", In Oller, J. Jr. and Jonz, J. (eds.), **Cloze and Coherence,** (London University Press, 1994, pp. 189-196).

J. Jonz, "Textual sequence and second-language comprehension", **Language Learning,** 37, (1989), pp. 207-249.

J. Jonz, "Cloze item types and second language comprehension", **Language Testing**, 8, (1991), pp. 1-22.

J. Oller, Jr, **Language tests at school**, (London: Longman, 1979).

J. Oller, Jr, "Cloze, discourse, and the approximation to English", In Oller, J. Jr. and Jonz, J. (eds.), **Cloze and Coherence**, (London University Press, 1994, pp. 119-133).

E. Rankin and S. Thomas, "Contextual constraints and the construct validity of the cloze procedure", In Oller, J. Jr. and Jonz, J. (eds.), **Cloze and Coherence**, (London University Press, 1994, pp. 165-175).

2. W. Taylor, "Cloze procedure: A new tool for measuring readability", **Journalism Quarterly**, 30, (1953), pp. 415-33.
3. Ibid, p. 416.
4. J. Carroll, A. Carton and C. Wilds, **An investigation of cloze items in the measurement of achievement in foreign languages,** (Cambridge, Mass: LRI, Graduate School of Education, Harvard University, 1959).
5. J. Alderson, "Native and non-native speaker performance in cloze tests", **Langauge Learning,** 30, (1980), 59-76.
6. For a detailed discussion see the following sources:

   A. Hughes, **Testing for language teachers,** (New York: Cambridge University Press, 1989).

   A. Hughes, "Response to Spolsky", In J. de Long and D. Stevenson (eds.), **Individualizing the assessment of language abilities,** (Clevedon, Philadelphia: Multilingual Matters Ltd., 1990, pp. 16-20).

   A. Davies, **Principles of language testing,** (Cambridge, MA: Basil Blackwell, 1990).

   A. Davies, "Operationalising uncertainty in language testing: An argument in favor of content validity", In J. de Long and D. Stevenson (eds.), **Individualizing the assessment of language abilities,** (Clevedon, Philadelphia: Multilingual Matters Ltd, 1990, pp. 179-195).

   B. Spolsky, "Social aspects of individual assessment," In J. de Long and D. Stevenson (eds.), **Individualizing the assessment of language abilities,** (Clevedon, Philadelphia: Multilingual Matters Ltd, 1990, pp. 3-15).

   C. Weir, **Communicative language testing, with special reference to English as a foreign language,** (Exeter University Linguistic series No. 11, 1988).
7. A. Hughes, 1989, p. 65.
8. W. Taylor, "Recent developments in the use of cloze procedure", Jounalism Quarterly, 33, (1956), pp. 24-48.

   J. Oller, Jr., "Some psycholinguistic controversies". In J. Oller and J. Richards (eds.), **Focus on the learner: Pragmatic perspectives for language teachers.** (Rowley, MA: Newbury House, 1973, pp. 36-52).

   J. Oller, Jr, and C. Conrad, "The cloze procedure and ESL proficiency", **Language Learning,** 21, (1971), pp. 183-196.

   J. Oller, Jr, and N. Inal, "A cloze test of English prepositions", **Tesol Quarterly,** 5, (1971), pp. 315-326.

   J. Stubbs and G. Tucker, "The cloze test as a measure of English proficiency", **Modern Language Journal,** 58, (1974), pp. 239-242.

   V. Streiff, "Relationships among oral and written cloze scores and achievement test scores in bilingual setting", In J. Oller, Jr. and K. Perkins (eds.) **Language in education: Testing the tests,** (Rowley, Mass: Newbury House, 1978, pp. 65-102).

   K. Aitken, "Using cloze procedure as an overall language proficiency test.", **TESOL Quarterly,** 11,(1977), pp. 59-67.
9. J. Oller, Jr., T. Bowen, T. Dien and V. Manson, "Cloze tests in English, Thai and Vietnamese: native and non-native performance", **Language Learning,** 22, (1972), pp. 1-15.

   S. Xiao and J. Oller, Jr., "Can relatively perfect translation between English and Chinese be achieved?", **Language Testing,** 11, (1994), pp. 267-289.
10. L. Bachman, **Fundamental consideration in language testing,** (Reading, MA: Oxford University Press, 1990)

    P. Oltman, L., Stricker and T. Barrows, "Analyzing test structure by multidimensional scaling", Journal of Applied Psychology, 75, (1990), pp. 1-7.

11. E. Briere and F. Hinofotis, "Cloze test cut off points for placing students in ESL classes", In J. Briere and F. Hinofotis (eds.) **New concepts in language testing: Some recent studies,** (Washington, D.C.: TESOL, 1979, pp. 12-20).
12. Taylor, 1956.
13. J. Bormuth, **Cloze Tests as measures of readability and comprehension ability.** Unpublished Ph.D. Dissertation, (University of Indiana, 1962).
14. D. Darnell, **The development of an English language proficiency test of foreign students using a cloze-entropy procedure,** (ERIC ED 024039, 1968).
15. J. Oller, Jr., "Scoring methods and difficulty levels for cloze tests of proficiency in ESL", **Modern Language Journal,** 56, (1972), pp. 151-158.
16. F. Hinofotis, **An investigation of the concurrent validity of cloze testing as a measure of overall proficiency in English as a second language.** Unpublished Ph.D. Dissertation, (Southern Illinois University, 1976).
17. P. Irvine, P. Atai and J. Oller, Jr., "Cloze, dictation, and the test of English as a Foreign Language", **Language Learning,** 24, (1974), pp. 245-252.
18. J. Brown, **Correlation study of four methods for scoring cloze tests.** Unpublished MA thesis. (University of California at Los Angeles, 1978).
19. Taylor, 1956, pp. 24-48; Oller, 1973, pp. 36-52; Oller, 1979; Oller and Conrad, 1971, pp. 183-196; Irvine, Atai and Oller, 1974, pp. 245-252; Stubbs and Tucker, 1974, pp. 239-242; Aitken, 1977, pp. 59-67; Streiff, 1978, pp. 65-102.
20. J. Oller, Jr., "Evidence for a general language proficiency factor: an expectancy grammar", **Die Neuren Sprachen,** 2, (1976), pp. 165-174.
21. Oller, 1972, pp. 151-158.
22. Oller, 1979, pp. 303-305.
23. J. Oller, Jr., "Discrete-point tests versus tests of integrative skills", in J. Oller, Jr. and J. Richards (eds.), **Focus on the learner: Pragmatic perspectives for langauge teachers,** (Rowley, MA: Newbury House, 1983, pp. 184-199).
24. J. Alderson, "The cloze procedure and proficiency in English as a second language", in J. Oller Jr. and J. Richards (eds.). **Focus on the learner: Pragmatic perspectives for language teachers,** (Rowley, MA: Newbury House, 1983, pp. 205-217).
25. E. Shohamy, "Interrrater and intrarater reliability of the oral interview and concurrent validity with cloze procedure", in J. Oller, Jr. and J. Richards (eds.). **Focus on the learner: Pragmatic perspectives for language teachers,** (Rowley, MA: Newbury House, 1983, pp. 229-236).
26. Ibid, p. 234.
27. J. Devine, "General language competency and adult second language reading", in J. Devine, P. Carrel, and D. Eskey (eds.), **Research in reading in English as a second language,** (Washington, D.C.: TES OL, 1987, p. 83).
28. F. Hinofois and B. Snow, "An alternative cloze testing procedure: multiple-choice format", In Oller, J. Jr. and Perkins, K. (eds), **Research in language testing,** (Rowley, MA. Newbury House, 1978, pp. 234-257.
29. J. Brown, "Relative merits of four methods for scoring cloze tests", **Modern Language Journal,** 64, (1980), pp. 311-317.
30. L. Pike, **An evaluation of alternative item formats for testing English as a foreign language,** TOEFL Research Report No. 2, (Princeton, New Jersey: Educational Testing Service, 1979).
31. G. Hale, C. Stansfield, D. Rock, M. Hicks, F. Butler and J. Oller, Jr., "The relation of multiple-choice cloze items to the Test of English as a Foreign Language", **Language Testing,** 1, (1989), pp. 47-76.

32. Ibid, p. 60.
33. Four different texts varied in their level of reading difficulty were selected from the following sources and they are arranged from the advanced to the preparatory levels:

    A. Sonka, **Skillful reading,** (Englewood Cliffs, NJ: Prentice-Hall International, 1986, pp. 91-92).

    G. Mosback and V. Mosback, **Practical faster reading,** (Cambridge, England: Cambridge University Press, 1981, p. 15).

    S. Robsertshaw, R. Hamblen and R. Fledman, **Reading first: Building reading competence,** (Boston, MA: Heinle & Heinle Publishers, 1990, pp. 134-135).

    P. Ackert, **Insights and ideas,** (New York, New York: Holt, Rinehart and Winston, Inc. 1982, pp. 13-14).
34. Oller, J., Jr., YÜ, G., Geenberg, L. and Hurtado de Vivas, R., "The learning effect from textual coherence measured with cloze", In Oller, J. Jr. and Jonz, J, (eds.), **Cloze and coherence,** (London University Press, 1994, pp. 247-268).
35. Al-Fallay, I. **The impact of background knowledge on the proficiency of Saudi Arabian students learning English as a Foreign Language,** Unpublished Ph.D. Dissertation. (The University of New Mexico, Albuquerque, NM, 1994).
36. The t values were calculated using the following formula:

$$t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}}$$

    The formula was mentioned in A. Odah and K. Al-Khalili, **Statistics for the researcher in education and humanities,** (Amman, Jordan: Dar Al-Fekr, 1988, pp. 304-305) (in Arabic).
37. The formula used was:

    $Z' \pm Z\rho\ \sigma z'$

    where Z' = the value of Fisher's Z' corresponding to the sample r.

    $Z\rho$ = the magnitude of normally distributed for which the probability ρ of obtaining a value so deviant or more so (in either direction).

    ρ = (1 - C), where C is the confidence coefficient.

    $\sigma_{z'}$ = the standard error of Fisher's Z'. This was calculated as

$$\sigma_{z'} = \frac{1}{\sqrt{n-3}}$$

    The formula was mentioned in E Minium, B. King and G. Bear, **Statistical reasoning in Psychology and Education**, 3rd ed. (New York; NY: John Wiley & Sons, Inc., 1993, p. 446).
38. Ibid, p. 569. The formula used was

$$Z(r_1-r_2)\ obs. = \frac{Zr_1 - Zr_2}{\sqrt{\frac{1}{n_1-3} + \frac{1}{n_2-3}}}$$

39. Ibid, p. 569.
40. Al-Fallay, 1994.
41. T. Taira, **Episodic organization and CALL: A pragmatic approach.** Unpublished Ph.D. Dissertation. (The University of New Mexico, Albuquerque, NM, 1992, pp. 103-106).

❊ ❊ ❊