OPEN-ENDED CLOZE PROCEDURE AND MULTIPLE-CHOICE CLOZE PROCEDURE: SIMILARITIES AND DIFFERENCES

BY

Dr Haifa A. Al-Buanain
Qatar University
Department of English

ABSTRACT

This paper is a validation of two cloze procedures: Open-Ended and Multiple-Choice (henceforth OE and MC respectively). Open-Ended and Multiple-Choice tests on the same text were administered to 60 English majors in their 2nd year of university level, in Qatar. The subjects were divided randomly into two groups. The first group took **Format 1**, while the second group took **Format 2**. The order of the procedures were different on the two formats. Each subject attempted 40 multiple-choice items and 40 open-ended items.

Statistics were calculated for the entire groups of 60 subjects as well as separately for the 30 subjects taking **Format 1** and the other 30 taking **Format 2**.

I. INTRODUCTION

Cloze procedure, which was first applied as a readability measure with native speakers (Taylor, 1953), has since been demonstrated in many studies to have substantial concurrent validity as an integrative test of overall proficiency in English as an SL (e.g. Alderson, 1979; Darnell, 1970; Oller, 1973; Irvine *et al*, 1974; Stubbs and Tucker, 1974 and Hinofotis and Snow, 1980). In these studies, high correlation were obtained between cloze scores and corresponding scores on an establishing measure of language proficiency.

The principle is based on the Gestalt psychology theory of "closure". It claims that by filling in something which has been deleted the brain is completing an incomplete structure i.e. closing gaps in patterns subconsciously. It has been argued that learning various kinds of basic cognitive skills and procedural knowledge involves acquisition of complex processing rules of which the subject is not aware (Lewicki *et al.*, 1987). It is also suggested that ability to fill in gaps in imperfect patterns may be related to ability to construct the same patterns. "It would seem that the ability to fill in gaps in prose is characterisible in that way" (Oller, 1973: 342). On the basis of incomplete information, the subject is required to project a word to fill in a blank and thereby complete a sequence. In so doing both the "productive/active" and "receptive/passive" skills of language must be utilized. It is interesting to note that the process of taking a cloze test involves more than "passive" reading. By sampling the information that is present the subject formulates hypotheses, or expectations about information that is to follow. The subject is making predictions which he either confirms or refutes. If the expectations are refuted, they must be revised and new hypotheses must be formed. Thus, cloze tests measure the readers' ability to decode "interrupted" or "mutilated" messages by making the most acceptable subsituations from all the contextual clues available. Oller (1973) referred to "grammar of expectancy" as a major factor which cloze tests measure. He argued that a grammar of expectancy is the "chief mechanism underlying the skills of thinking, understanding, speaking, reading and writing."

Anderson (1971, 1972), Oller and Conrad (1971) and Porter (1976) constructed tests by modifying the cloze procedure originally invented by Taylor. The procedure as established by Taylor is the so-called fixed ratio method. This is a continuous passage of prose in which every nth word has been systematically deleted. While marking the test, only replacement of the correct item was accepted as the principle of scoring. With this basic framework, cloze tests will vary from one another on one or more of the following variables: deletion, rate, length of passage, type of passage (style, content, syntactic and semantic complexity) and the marking method employed in correcting the test.

Evidence about the effect of all these variables on test results is not yet conclusive. Porter (1978) believes that "it cannot be assumed that any 2 cloze tests will yield equivalent results: blank-position and possibly style may be important variables" (p. 333). Thus, we find Davies (1978) warns that "cloze must be treated with caution."

6

II. THE PRESENT STUDY

2.1 Rationale

What most language tests do is to place one student in relation to a group of students, in terms of some particular language ability. The comparison for the test is therefore between students. The comparison for the experiment, on the other hand, is between effects or treatments of methods.

Occasionally it is suggested that the problem of language testing could be solved by following one's students around for some considerable time and observing how they perform in a very wide range of actual communicative situations, thus allowing powerful generalisations to be made about proficiency. However, this proposal is impractical (Clark, 1978).

Skehan (1989: 213) suggests that language proficiency consists of a general factor plus specific factors concerned with oral/aural skills, literary skills and then more specific aspects still of test material. Therefore, the combination of different types of tests may serve not merely to tap different aspects of proficiency but also to reduce bias which may arise from learner characteristics (Farhady, 1982). For instance, Hanani and Shikhani (1986) investigate the interrelationships among 3 types of measures. They sought to determine whether the addition of cloze component to the standardized ESL test would improve the predictability of students' communicative proficiency as reflected in their performance on a writing test. High correlations were found between the tests. Also, there was a "substantial residual correlation between the cloze and writing tests which suggests that these tests may measure in common some aspects of language ability beyond those that they share with other tests" (P. 97).

2.2 Statement of the problem

The main research questions asked in this experiment are:

a) Are there any variations in cloze procedures on the following two variables

  i) on cloze procedure?;

 ii) on order of procedure?;

      and whether their interactions significantly affected cloze scores on the same cloze test given to the same group of subjects?

b) Are the two procedures OE and MC similar or different?

c) Does the multiple choice procedure correlate with the open ended procedure?

d) Does the MC procedure on Format 1 correlate with the MC procedure on Format 2?

e) Similarly, does the OE procedure on Format 1 correlate with the OE procedure on Format 2?

2.3 Hypotheses

Our alternative hypothesis (H_1) is that there is a significant difference between the two procedures: i.e. there is a difference between OE cloze procedure and MC cloze procedure.

Our second alternative hypothesis (H_2) in this experiment is that, the independent variable (order of the procedure i.e. OE-MC versus MC-OE) will not make any significant difference which may influence the dependent variable (scores).

2.4 Construction of the test

In order to provide results which would be comparative in relation to particular texts and the effect of different procedures,

a) Two cloze procedures were applied:

    i) Open-Ended

    ii) Multiple Choice

b) Two formats were used:

    i) Format 1 (OE before MC)

    ii) Format 2 (MC before OE)

On format 1 the students were given the Open-Ended (OE) version first and Multiple Choice (MC) version second. The order was reversed on format 2.

c) Two groups (each of 30 students) were tested:

    i) First group took Format 1.

    ii) Second group took Format 2.

2.5 Subjects

The sample for whom we are reporting data consisted of 60 native speakers of Arabic studying English as their major at Qatar University. The subjects were drawn from 3 classes of 2nd year. The three groups could be regarded as a homogeneous group, if we allowed for individual variability. The examiner had been teaching the students for almost two years now and "knew them very well."

The subjects in the 3 classes are female students, aged 20-26. The subjects were divided randomly into two groups of 30.

2.6. Text

The text deals with "Atlantis", a subject with which most of the students would be familiar, although none of them had encountered the particular text. It was taken from "Practice Tests for First Certificate English", Book 1. The text contains

approximately 270 words, 10 sentences. The mean sentence length is 27 words. The text is written mainly in the present tense. The proportion of "co-ordinate sentence" and "complex sentence" is approximately equal. While the proportion of "simple sentence" is less than the two. Every sixth word was deleted resulting in 40 missing words.[1] A line of uniform length was replaced for the deleted word on the OE procedure, while 4 choices were listed on the MC procedure. The first two sentences of the text were left intact to provide lead-in context. The last three lines were left as lead-out from the context. (See Appendices 1 and 2).

2.7. Method of Scoring

The Exact Word (EW) method was employed in this study. It involves counting as correct only the words which were actually deleted. Stubbs and Tucker (1974) believe that the cloze exact scoring and the cloze acceptable scoring yielded a very high correlation which would eliminate the time consuming problem evaluating acceptable responses. It would result in more reliable scores since the element of subjectivity which enters into the evaluation of contextually acceptable responses would be eliminated entirely. Similarly, Lado (1986: 137) states that "The advantage of easy preparation and scoring of cloze tests is cancelled if we use acceptable word scoring, which turns out to be difficult and time consuming and requires technical testing and linguistic knowledge." Another advantage of using the EW method in this study is that the marking will be as objective as the marking on the Multiple Choice close procedure.

On the OE procedure, spelling mistakes were disregarded except in so far as the identity of the word might be in doubt. One point was given for each correct response.

2.8 Administration of the test

The test was administered at the end of Autumn 1988 Semester. It was given during normal class hours. The teacher had allayed any apprehension students might have felt by emphasizing some days before the test, that their scores on the test would remain confidential and that the purpose of the test was to investigate a testing procedure, not the standard of the students. This was felt to be necessary in view of the fact that students might otherwise have thought that their results would affect their marks. However, to avoid Hawthorne effect, the word "experiment" was not mentioned to the students.

Tests were distributed among the two groups so that the first group students took (OE) procedure, while the second group were taking (MC) procedure. Despite the fact that the students knew that they were not really being 'tested' they responded well to the test.

---

[1]. We can either delete selected words or delete words at regular intervals (every 7th word is often chosen). Tests in which selected words have been deleted (e.g. grammatically significant words) have not been shown to be superior in performance to tests where a regular deletion intervals have been used. (Baker 1989: 74).

Five minutes were spent for instructions about the test. For the open-ended items the students were asked to follow the standard fill-in the blank procedure. For the multiple-choice items, the four alternatives were listed in the appropriate places and the students were instructed to underline the best choice in the blank. They were also instructed to pay close attention to the grammatical structure, as this would help them towards a clearer understanding of both the structural and thematic development of the text. The administrator then asked whether there were any questions. Subjects were allowed not more than 60 minutes to complete each procedure. Oller suggested that 50 items in 50 minutes is reasonable for a native speaker, (here we are dealing with non-native speakers of the language).

2.9 Analysis

The scores for the test were computerized for data analysis using SPSSX programme (SPSSX User's Guide, 1983). Statistics were computed over the entire group of 60 subjects and separately for the 30 subjects taking Format 1 and the 30 subjects taking Format 2. The raw scores (see Appendix 3) resulting from the two procedures and formats described above, were converted into mean scores and standard deviations to facilitate inter-procedure and inter-format comparison. Also, t-tests were computed to test for significant differences between individual test formats. Product-moment coefficients of correlation were computed for the comparison of the two procedures and formats. An item analysis was carried out for Format 1, i.e. OE-MC.

III. RESULTS AND DISCUSSION

The summary test statistics are reported in the following table:

TABLE ONE: Mean Scores, Standard Deviations, Medians and Range for Procedures and both Formats

| | Format 1 (OE-MC) | | Format 2 (MC-OE) | |
	Open Ended	Multiple Choice	Open Ended	Multiple Choice
Mean	17.37	23.23	23.6	23.07
S.D.	3.68	4.27	3.76	3.88
Median	18	23.5	23.5	23.5
Range	16	19	14	15
No	30	30	30	30

Where No = number of subjects

SD = standard deviation

Inspection of Table one points to the probability of a significant difference between the mean scores of the order of the procedures in the two formats. The mean for the two procedures in Format 2 were almost the same (MC = 23.06, OE = 23.6). However, on Format 1 : (OE — MC), there was a great difference between the two means. The

Open-Ended mean score was 17.37, while the Multiple Choice mean score was 23.23. The higher mean scores for the MC test indicate that the MC task was easier.

As far as the standard deviation is concerned, the scores were more spread on the MC procedure than on the OE procedure.

The MC procedure could mask students' incompetence. In the MC test the student has approximately 25% (1 in four) chance of guessing the right answer. This, I think, brings us to the crux of the matter. MC cuts down the risk of making mistakes because it sometimes aims to test only a very narrow slice of language. The chance element correspondingly increases. It tends to cover up differences between the best and the worst. Thus performance may be more uniform and it becomes difficult to draw any conclusions. Oller (1974) believes that cloze technique challenges the efficiency of the developing SL grammar of the learner in a way that reflects natural language processes rather than contrived pedagogy. He also believes that MC is generally the distilled product of an evolutionary process which clearly emphasizes administrative, pedagogical and procedural considerations. According to him, the former is a contextualized challenge to learner grammar efficiency, displaying an enviable simplicity.

This is in accord with Pickett (1968) who argues that those who become fluent through blank-filling may become equally fluent through another method because they have something which is independent of either; i.e. 'linguistic' ability'. Krzyanowski (1976) writes that "cloze is considered an integrative rather than a discrete-point test" because it draws at once on the overall grammatical, semantic and rhetorical knowledge of the language.

However, in this study, the MC procedure is used in a passage rather than in isolation. In order to reconstruct the textual message, students have to understand key ideas and perceive interrelationships within a stretch of continuous discourse and they have to recognize an appropriate word for each blank. On the OE procedure the students' task is more difficult since they have to produce, rather than simply recognize, an appropriate word for each blank. The focus of the task involved is more communicative than formal in nature and it is therefore considered to reflect the students' ability to function in the language. In other words OE items test productive skills while MC items test receptive skills only (Ingram 1974).

3.1 Procedure

3.1.1 Comparison of Mean Scores and Standard Deviations

Table two: Mean Scores and Standard Deviations

	Format 1 OE - MC		Format 2 MC - OE		All Subjects	
	Mean	SD	Mean	SD	Mean	SD
OE	17.37	3.68	23.6	3.76	20.5	3.71
MC	23.23	4.27	23.07	3.88	23.15	4.07

Table two above, confirms the expectation mentioned in section 2.3 concerning the differences between the two procedures. The higher mean scores for the MC test across the board indicate that the MC task was easier.

If we consider Format (1): OE - MC, we will notice a great difference between the two means. The students performed better on the MC test. One would expect a procedure giving clues to possible answers to be substantially easier than a procedure requiring exact-word replacement of a blank. And, indeed, the mean score of OE procedure is considerably lower than the mean scores of MC procedure.

The increase in means of the OE procedure in Format 2 (i.e. MC - OE), as well as for all subjects, has to be interpreted with caution. These two values were clearly influenced, first by the design of the multiple choice procedure and in particular by the choice of distractors. An attempt was made to incorporate into the distractors stated for each deleted item at least one contextually appropriate item and one wrong (inappropriate) answer (see Appendix 2). Naturally, this was not always feasible; when the deleted item was a functional word, e.g. a preposition, an article, a conjunction, etcetra, the choice of distractors given quite often consisted of three unacceptable inappropriate items. The distractor analysis given in Appendix 4, shows that the quality and effectiveness of the distractors also differed from format to format (different groups of subjects and different order of procedures), with the distractors being slightly more effective on Format 2.

Second, the increase in the means, was also influenced by the order of the procedures. On Format 1, the students did not have any clues for filling in the blanks. On the other hand, on Format 2, they were given strong clues by taking the MC procedure first. Therefore, their chances of filling in the blanks correctly were more than those who took Format 1. Obviously, memory was an extremely important factor in ability to predict the correct word. Oller (1973), concluded his paper "Cloze Tests of Second Language Proficiency and what they Measure" by stating that "If as I have argued in this paper, and elsewhere, language competence is best characterized by a grammar of expectancy, then memory constraints are clearly an aspect of competence. It is my belief that this is the sort of competence measured by cloze test."

Clearly the order of the two procedures, the choice of distractors and the characteristics of the multiple choice procedure heavily biased the mean score obtained for the OE procedure on Format 2 and for all subjects. This is also borne out by the standard deviation values obtained. Whilst the standard deviation was 3.68 on Format 1, it increased to 3.71 and 3.76 for all subjects, and on Format 2, respectively.

Thus, I believe it is correct to conclude that when we come to the level of procedure (OE versus MC), OE task was more difficult than MC at least for this particular sample.

3.1.2. t-test between Procedures on both Formats

The t-values tabulated below in Table Three (derived from t-tests set out in Appendix 5) reveal that there is a significant difference (even on the 0.001 level of significance with 29 degrees of freedom) between the two procedures (i.e. MC and OE) on Format 1. While the difference is insignificant on Format 2 (MC before OE). This is mainly due (as mentioned before) to the order of the procedures. In other words, the students were benefitted from taking the multiple choice procedure first. (This will be discussed in detail in section 3.1.4.).

TABLE THREE: t-tests between Procedures on both Formats

	Procedure	Observed t	Tabulated t	Significance level with 29 degrees of freedom (two Tail test)
Format (1)	OE - MC	4.40	2.045	0.001
Format (2)	MC - OE	1.00	2.045	Not significant

Therefore, from the t-test on Format 1, we can safely retain the alternative hypothesis (H_1), mentioned earlier in section 2.3. Our two tail hypothesis stated that there is a difference between the two procedures.

However, the difference between the two procedures might be due to the fact that the distractors on the multiple choice procedure were not powerful enough. So, the students did not have a lot of trouble in choosing the correct option. In order to investigate the appropriateness of the above mentioned suggestion, an item analysis was carried out.

3.1.3. Item Analysis

The item analysis was carried out only for Format 1 (OE - MC), since Format 2 will not reveal unbiased results due to the order of the procedures.

The item facility indices (tabulated in Appendix 7) revealed that in general both MC items and OE items were of the same difficulty. Few of the items on either the OE or MC tests, however, approached 100% facility level, but a number of them came close to the 15% facility level, indicating that they were quite difficult for this sample of students. The discrimination of both procedures was not very high, however, it was slightly better with the multiple choice items than with the open-ended task (see Appendix 6). This may be due to the fact that in taking an open-ended cloze test the student has to generate his own alternatives much as he would in normal communication allowing less chance for correct guessing. Also, this may be due to the fact that although some of the multiple choice items, especially in the options given for content words, more than one item was applicable. However, with regard to the scoring method of the exact word, the other should be considered wrong even if they were contextually appropriate.

13

The item seems to indicate that the difference between the two procedures was not mainly due to the choice of the multiple choice distractors. Therefore, from the t-test given in the previous section (3.1.2), and from the item analysis, a conclusion could be drawn, that there is a significant difference between the two procedures (i.e. MC and OE). This is consistent with Davies (1978: 217) who argues that a multiple choice test is not a cloze.

3.1.4. Correlations between Procedures on both Formats

From Table Four below, we notice that on Format 1 (OE - MC), the correlation was lower than that on Format 2. On Format 1 it was 0.47, while on Format 2 the correlation was 0.73, which is considered to be a high correlation. Some might argue that this is a moderate correlation rather than a high one, since it indicates 53% shared variance, yet in comparison with the correlation on Format 1 (0.47), which indicates a mere 22% shared variance, it is certainly a high correlation.

However, the increase in the correlation values should be interpreted with caution, bearing in mind the considerations which were discussed earlier under section 3.1.1. Clearly, students who started with multiple choice version of the test had clues of the deleted items. Therefore, when filling in the blanks on the OE task they supplied the words they chose for the multiple choice procedure. Thus, the two procedures correlated highly since students who did badly on the multiple choice procedure also did badly on the open-ended procedure.

On the other hand, students who took Format 1 started with the open-ended procedure, followed by the multiple choice. Those students did not have any clues to pick the correct option in the multiple choice test. Even if they supplied acceptable words from the deleted items they were marked as incorrect (due to the scoring method). Therefore, since most of the students who scored badly on the open-ended test, scored highly on the multiple choice test,[1] the two procedures were of lower correlation.

Although the difference between the correlations of the OE cloze scores and the MC cloze scores on the two Formats appear to be fairly large, the two obtained correlations are, in fact, statistically significant even at the 0.01 probability level. Since the Pearson coefficient is significant to the 0.01 level when $r_{xy} = 0.44$ with 28 degrees of freedom. This suggests that the Open-Ended and Multiple Choice tests are providing similar information concerning the students' level since those who scored the highest marks on one procedure were also the best on the other. The hypothesis that language skills being integrated rather than separate (Scholz *et al* 1980) could give us some explanation here. Whenever we are testing the students no matter what kind of test we are giving, the best students will score the highest marks on different tests. This experiment confirmed the findings of Pike (1973); Jonz (1976) and Hinofotis and Snow (1980). The results also support Bormuth's view (1967: 295) in which he claims that "cloze tests ... measure skills closely related or identical to those measured by conventional multiple choice reading comprehension test."

---

[1]. See Appendix 3, Raw Scores on Both Formats.

TABLE FOUR: Correlations between Procedures on both Formats

	Procedure	r	Significance Level[1]
Format 1	OE - MC	0.47	0.01
Format 2	MC - OE	0.73	0.01

A comparison may now be made between significance levels for t-tests and correlations of the two procedures on both formats. Table Five below shows that on Format 1 the t-test between the two procedures revealed highly significant differences between means. Furthermore, the corresponding correlation was also highly significant. However, on Format 2, the t-test indicated insignificant difference between the two procedures, while the correlation was highly significant.

TABLE FIVE: Significance Levels for t-tests and Correlations

	Format 1 (OE - MC)	Format 2 (MC - OE)
t-test	0.001	Not significant
Correlation	0.01	0.01

3.2 Formats

3.2.1. Increase in Mean Scores for OE procedure on Format 2

TABLE SIX: Mean Scores, Standard Deviations, and Raw Scores

Format	Raw Scores	Mean	S.D.
1	521	17.37	3.68
2	708	23.6	3.76

Table Six above, shows the increase in the raw scores, mean and standard deviation for the OE procedure on Format 2. The mean increase in score is noticeably greater than that on Format 1. The reason for the large increase in mean score was explained by the fact that students had clues to fill in the blanks. It is well established (Oller, 1974) that predictability is inversely proportional to the size of grammatical class. Open class words will therefore, be more difficult to predict than closed class words. This was certainly the case with the students who started with the OE procedure but not for those who began with the MC procedure. The latter had the advantage of being familiar with the text and with its syntactic and semantic constraints. Oller (1973) quoted Jongsma (1971) who quite rightly claimed that "It is known that if a student is familiar with the text of a cloze passage, he will perform

---

[1] See Guilford and Fruchter (1978:531).

significantly better, and in proportion to his familiarity with the passage: (This point was discussed earlier in sections 3.1.1 and 3.1.4).

3.2.2. Increase in Mean Scores for MC procedure on Format 1

From Table Seven below, a comparison could be made between the raw scores, means and standard deviations of the MC procedure on the two Formats. On the whole, we will notice that Format 1 was easier for this sample of students than Format 2, although the difference is so small. This is due to the MC technique. The raw scores, mean and standard deviation were slightly decreased on Format 2.

It may be recalled that in section 2.5, the claim was made that the two groups were of the same population. Therefore, although the two Formats were not given to the same individuals, the two groups were considered as one group. So, if the differences between the two groups were disregarded for the above mentioned reason, the increase in the scores was mainly influenced by the order of the procedure.

TABLE SEVEN: Mean Scores, Standard Deviations and Raw Scores

Format	Raw Scores	Mean	S.D.
1	697	23.23	4.27
2	692	23.07	3.88

3.2.3. t-tests between same Procedure on Different Formats

t-tests for independent samples were conducted in order to investigate the null hypothesis, reads as the order of procedure would have no significant effect on the scores (see section 2.3.).

TABLE EIGHT: t-tests between same Procedure on Different Formats

Procedure	Observed t	Tabulated t	Significance Level with 58 Degrees of Freedom (Two Tail Test)
Multiple Choice	0.16	2.000	Not significant
Open Ended	6.36	2.000	0.05

From the t-values tabulated above in Table Eight (derived from t-tests set out in Appendix 7), two remarks could be made. First, on the MC procedure there was no significant difference for the observed t. Since its value was 0.16, a very low figure, this indicates that the independent variable (IV) i.e. the order of procedures, had no effect on the dependent variable (DV), i.e. the scores. Therefore, the null hypothesis was safely retained. This of course, should have been expected because the students who were given the OE procedure first did not have any advantage, i.e. clues for choosing the correct option on the MC

procedure. The students might have benefited from being familiar with the text, but this did not help them in choosing the correct alternative.

Second, from Table Eight, we notice very clearly that there is a significant difference for the observed t on the OE procedure which was significant even to the 0.001 level of significance with 58 degrees of freedom. Thus, the null hypothesis was rejected at 0.001 level. This indicates that the order of procedure had an effect on the scores. Obviously, this shows that the subjects benefited from taking the MC procedure first, through which they had clues to fill in the blanks (This was discussed earlier in sections 3.1.1., 3.1.4., and 3.2.1.).

3.2.4. Correlations between Same Procedures on Different Formats

TABLE NINE: Correlations between same Procedures on Different Formats

Procedure		Significance level with 28 Degrees of Freedom
OE	0.69	0.01
MC	0.45	0.01

Table Nine above shows that the correlation between OE Format 1 and OE Format 2 is 0.69, while the correlation between MC Format 1 and MC Format 2 is 0.45.

Although the differences between the two correlations seem to be quite large, both correlations are statistically significant at the 0.01 probability level with 28 degrees of freedom. This would suggest that even when presented in different order, it is unlikely that the strength of these correlations is the product of chance. The same procedure provided similar information about the level of students.

An interesting comparison may now be made between significant levels for t-tests and correlations for the procedures between formats. Table Ten below shows that while t-test between MC Format 1 and MC Format 2 revealed significant difference between means, the corresponding correlation was highly significant. On the other hand, highly significant difference between means and highly significant correlation were found between OE Format 1 and OE Format 2.

TABLE TEN: Significance Levels for t-tests and Correlations between same Procedure

	MC	OE
t-tests	Not significant	0.001
Correlations	0.01	0.01

It was felt that such an inverse relationship between the correlation coefficient and the t-value, for the MC procedure, would be elucidated by the reasons

discussed more than once in this paper. Mainly, that the order of the procedures did not influence the scores on the MC task as it did with the OE task.

IV. CONCLUSIONS

In conclusion, the experiment largely fufilled the aims made in section 2.3 and also raised further issues. Some of our predictions set out in 2.4 were supported by this experiment. The findings could be summarized as follows:

1) The two procedures, i.e. MC and OE were significantly different.

2) The order of procedures did not affect the scores on the multiple choice procedure.

3) The prediction that the scores would not be influenced by the order of procedures, was not demonstrated at all by the open-ended procedure. On the contrary, the dependent variable was highly affected by the independent variable. To which the null hypothesis should be rejected.

4) The correlation between the two procedures indicated that the two procedures were measuring skills which are closely related.

5) Finally, the results in this experiment suggest that MC cloze tests have some promise and may be preferred over OE cloze because of ease of scoring.

However, a caution to keep in mind is that constructing an MC cloze test is a considerably more complicated procedure than constructing an open-ended cloze test.

The value of this study would be greater when considering this experiment as a pretext for the MC cloze version in order to check item facility and discrimination. It will be worth while to modify or eliminate items with lower discrimination and re-test the MC cloze procedure. This will certainly reveal new and more likely different results. Once a reliable MC cloze procedure is obtained, the time involved in administering and scoring the test should be less.

REFERENCES

Allen, J. and A. Davies (eds.) (1977). *Edinburgh Course in Applied Linguistics: Testing and Experimental Methods*. Vol. 4. London: Oxford University Press.

Anderson, J. (1971). 'A Technique for Measuring Reading Comprehension and Readability. *English Language Teaching*. 25:178-82.

Anderson, J. (1972). 'The application of Cloze Procedure to English Learned as a Foreign Language in Papua and New Guinea'. *English Language Teaching*. 27:66-72.

Baker, D. (1989). *Language Testing: A Critical Survey and Practical Guide*. Edward Arnold.

Bormuth, J. (1967). 'Comparable Cloze and Multiple Choice Comprehension Test Scores'. *Journal of Reading*. 10:291-99.

Clark, J. (1978). *Direct Testing of Speaking Proficiency: Theory and Application*. Educational Testing Service.

Darnell, D. (1970). 'Clozentropy: A Procedure for Testing English Language Proficiency of Foreign Students. *Speech Monographs*. 37:36-46.

Davies, A. (1978). Language Testing: Survey Article 1 and 2. *Language Teaching and Linguistics Abstracts*. 2:145-59 and 215-31.

Davies, A. and Widdowson, H. G. (1975). 'Reading and Writing'. *Edinburgh Course in Applied Linguistics, Vol. 3*. London: Oxford University Press.

Farhady, H. (1982). 'Measures of Language Proficiency from the Learner's Perspective'. *TESOL Q*. 16:43-59.

Guilford, J. and B. Fruchter (1978). *Fundamental Statistics in Psychology and Education*. 6th edition. McGraw-Hill, New York.

Hanani, E. and M. Shikhani (1986). 'Interrelationships among Three Tests of Language Proficiency: Standardized ESL, Cloze and Writing'. *TESOL Q*. 20:97-109.

Heaton, J.B. (1975). *Writing English Language Tests*. London: Longman.

Hinofotis, F. and B. Snow (1980). 'An Alternative Cloze Testing Procedure: Multiple-Choice Format'. In J. Oller and K. Perkins (eds.). (1980): 129-33.

Ingram, E. (1968). 'Item Analysis.' In A. Davies (ed.). *Language Testing Symposium*. London: Oxford University Press.

Ingram, E. (1974). 'Language Testing.' In J. Allen and S. Corder, *The Edinburgh Course in Applied Linguistics: Techniques in Applied Linguistics*. Oxford.

Irvine, P.; P. Atai and J.W. Oller (1974). 'Cloze, Dictation and the Test of English as a Foreign Language. *Language Learning*. 24:245-52.

Jonz, J. (1976). 'Improving on the Basic Egg: the M-C Cloze'. *Language Learning*. 26:255-65.

Krzyzanowski, H. (1976). 'Cloze Tests as Indicators of General Language Proficiency'. *Studia Anglica Posnaniensia*. 7:29-43.

Lado, R. (1986). 'Analysis of Native Speaker Performance on a Cloze Test'. *Language Testing*. 3:130-46.

Lewicki, P.; H. Haffman and M. Czyzewska (1987). 'Unconscious Acquisition of Complex Procedural Knowledge. *Journal of Experimental Psychology: Learning, Memory and Cognition,* 13:523-30.

Oller, J. (1973). 'Cloze Tests of Second Language Proficiency and What they Measure'. *Language Learning*. 23:105-118.

Oller, J. (1974). 'Expectancy for Successive Elements: Key Ingredient to Language Use'. *Foreign Language Annuals*. 7:443-52.

Oller, J. and C. Conrad (1971). 'The Cloze Technique and ESL Proficiency'. *Language Learning*. 21:183-95.

Oller, J. and K. Perkins (eds.) (1980). *Research in Language Testing*. Newbury House.

Pike, L. (1973). 'An Evaluation of Present and Alternative Item Formats for Use in the Test of English as a Foreign Language.' Mimeograph. Princeton, NJ: Educational Testing Service.

Porter, D. (1976). 'Modified Cloze Procedure: A More Valid Reading Comprehensive Test.' *ELTJ*. 15:151-55.

Porter, D. (1978). 'Cloze Procedure and Equivalence'. *Language Learning*. 28:333-41.

Scholz, G. Hendricks, R. Spurling; M. Johnson and L. Vandenbury (1980). 'Is Language Ability Divisible or Unitary? A Factor Analysis of 22 English Language Proficiency Tests' in Oller and K. Perkins (eds.). *Research in Language Testing*. Newbury House.

Skehan, P. (1989). 'Language Testing Part II'. *Language Teaching*. 22:1-13.

SPSS Inc. (1983). *SPSSX User's Guide*. McGraw Hill.

Stubbs & Tucker (1974). The Cloze Test as a Measure of English Proficiency'. *Modern Language Journal*. 58:239-41.

Taylor, W. (1953). 'Cloze Procedure: A New Tool for Measuring Readability'. *Journalism Quarterly*. 30:415-33.

Widdowson, H. (1978). *Teaching Language as Communication*. Oxford University Press.

NAME: ....

YEAR: ....

AGE: ....

In the following passage there are 40 blanks left to be filled in. First read over the entire passage and try to understand what it is about. Then try to fill in the blanks. It takes exactly ONE word to fill in each blank. If you are not sure of the word that has been left out, leave it and come back to it. If you still do not know what it is, GUESS. Please fill in ALL the blanks. Note that the words vary in length. If you have any question, please DO ASK.

The lost city of Atlantis remains one of the world's greatest mysteries. Its existence

or non-existence has intrigued mankind throughout history. Scientists and

archaeologists have devoted .... time to researching the

problem, ... to date no really conclusive .... has been

produced. If Atlantis .... exist, the first question that

.... to be answered is where .... its geographical location?

According to .... Greek philosopher, Plato, Atlantis was

.... beyond the "Pillars of Hercules", ... the Straits of

Gibraltar were .... called, in a part of .... is now the Atlantic

Ocean. .... was in a period before .... melting of the Polar

ice- .... when the two great continents .... Africa and America

were joined .... land. Plato based his information .... a report

made to his ....-grandfather by Solon, who visited .... temple in

Egypt where the .... showed him written evidence of ....

existence of Atlantis. Plato relates .... "... in the land of Atlantis

21

.... was a great and wonderful .... which had ruled over the

.... island and several others as .... as over part of the

.... ". He describes in detail the .... wealth possessed by the

Atlanteans — .... temple, for example, covered in .... with pinnacles

of gold — and .... highly efficient organisation of their

...........................and trade and the general .... of their society.

It is .... curious fact that many ancient .... across the world share

the .... legends including that of the .... Flood from which only a

.... of specially-chosen persons survived. .... are similarities of

language and .... and religious customs that cannot be satisfactorily

explained, many of which support the theory that a continent or large group of

islands existed, the inhabitants of which established contact with Europe, the

Middle East and the Americas.

NAME: ....

YEAR: ....

AGE: ....

In the following passage four choices are given between the brackets. First, read over the entire passage and try to understand what it is about. Then choose the correct word by UNDERLINING it. If you are not sure of the correct word, leave it and come back to it later. If you still do not know the correct word, GUESS. Note that the sign (∅) means NOTHING. If you have any question, raise your hand and please don't hesitate to ask.

The lost city of Atlantis remains one of the world's greatest mysteries. Its existence or non-existence has intrigued mankind throughout history. Scientists and archaeologists have devoted (more/ much/ many/ some) time to researching the problem, (and/ however/ therefore/ but) to date no really conclusive (research/ evidence/ work/ answer) has been produced. If Atlantis (were/ did/ does/ will) exist, the first question that (ought/ has/ have/ had) to be answered is where (been/ be/ is/ was) its geographical location?

According to (one/ this/ the/ a) Greek philosopher, Plato, Atlantis was (placed/ situated/ located/ sited) beyond the "Pillars of Hercules", (as/ for/ like/ so) the Straits of Gibraltar were (originally/ then/ later/ often) called, in a part of (which/ that/ what/ where) is now the Atlantic Ocean. (That/ This/ There/ It) was in a period before (a/ this/ the/ that) melting of the Polar ice- (fall/ breaker/ cap/ field) when the two great continents (from/ with/ of/ between) Africa and America were joined

(through/ throughout/ with/ by) land. Plato based his information (in/ on/ upon/ with) a report made to his (old/ grand/ great/ dead) -grandfather by Solon, who visited (the/ one/ every/ a) temple in Egypt where the (priests/ Egyptians/ citizens/ inhabitants) showed him written evidence of (an/ this/ the/ non-) existence of Atlantis. Plato relates (this/ the/ that/ it) ".... in the land of Atlantis (where/ there/ which/ that) was a great and wonderful (kingdom/ country/ empire/ state) which had ruled over the (Ø/ whole/ main/ best) island and several others as (far/ much/ long/ well) as over part of the (world/ land/ sea/ continent) ..." He describes in detail the (enormous/ smallest/ great/ biggest) wealth possessed by the Atlanteans — (the/ a/ one/ every) temple, for example, covered in (steel/ marble/ silver/ pearl) with pinnacles of gold — and (also/ the/ their/ that) highly efficient organisation of their (market/ shopping/ work/ shipping) and trade and the general (success/ prosperity/ wealth/ generosity) of their society.

It is (that/ the/ a/ this) curious fact that many ancient (societies/ civilizations/ inhabitants/ cultures) across the world share the (various/ rare/ same/ famous) legends including that of the (Great/ Huge/ Big/ Old) Flood from which only a (dozen/ number/ handful) few) of specially chosen persons survived. (They/ These/ Those/ There) are similarities of language and (belief/ faith/ thought/ behaviour) and religious customs that cannot be satisfactorily explained, many of which support the theory that a continent or large group of islands existed, the inhabitants of which established contact with Europe, the Middle East and the Americas.

RAW SCORES ON BOTH FORMATS

Subjects	Format 1 (OE-MC)		Subjects	Format 2 (MC-OE)	
	Multiple Choice	Open-Ended		Multiple Choice	Open-Ended
1	22	10	31	21	22
2	20	14	32	24	25
3	19	12	33	20	20
4	26	17	34	26	25
5	26	22	35	28	28
6	30	18	36	27	27
7	12	12	37	24	25
8	24	17	38	25	16
9	25	15	39	20	19
10	26	18	40	26	27
11	14	18	41	16	18
12	25	14	42	16	17
13	22	20	43	29	30
14	25	18	44	21	23
15	25	22	45	22	25
16	22	23	46	20	24
17	20	19	47	18	21
18	30	26	48	16	22
19	21	19	49	20	22
20	29	21	50	22	21
21	25	18	51	26	28
22	21	18	52	21	21
23	25	13	53	27	29
24	18	15	54	23	20
25	22	12	55	24	22
26	26	22	56	24	27
27	23	15	57	31	26
28	24	16	58	24	28
29	31	20	59	22	21
30	19	17	60	29	29
$X =$	23.23	17.37		23.07	23.6
$SD =$	4.27	3.68		3.88	3.76
$>X =$	697	521		692	708

Distractor Analysis

	Format 1 (OE-MC)				Format 2 (MC-OE)			
	Choice				Choice			
Blank	A	B	C	D	A	B	C	D
1								
2								
3								
4								
5								
6								
7								
8								
9								
10								
11								
12								
13								
14								
15								
16								
17								
18								
19								
20								
21								
22								
23								
24								
25								
26								
27								
28								
29								
30								
31								
32								
33								
34								
35								
36								
37								
38								
39								
40								

Total: 503   Total: 509

T-TESTS between PROCEDURES (Format 1: OE-MC)

Formula employed: t for differences between correlated pairs of means

$$t = \frac{X_d}{\sqrt{\dfrac{> x_d 2}{N\,(N\text{-}1)}}}$$

Where X_d = mean of the N differences of paired observations

x_d = deviation of a difference from the mean of the differences

Subject	OE	MC	d	d²	Subject	OE	MC	d	d²
1	10	22	−12	144	16	23	22	1	1
2	14	20	− 6	36	17	19	20	− 1	1
3	12	19	− 7	49	18	26	30	− 4	16
4	17	26	− 9	81	19	19	21	− 2	4
5	22	26	− 4	16	20	21	29	− 8	64
6	18	30	−12	144	21	18	25	− 7	49
7	12	12	0	0	22	18	21	− 3	9
8	17	24	− 7	49	23	13	25	−12	144
9	15	25	−10	100	24	15	18	− 3	9
10	18	26	− 8	64	25	12	22	−10	100
11	18	14	4	16	26	22	26	− 4	16
12	14	25	−11	121	27	15	23	− 8	64
13	20	22	− 2	4	28	16	24	− 8	64
14	18	25	− 7	49	29	26	31	−11	121
15	22	25	− 3	9	30	17	19	− 2	4

$\leq$ d = 176 $\leq x_d^2$ = 1548 X_d = 5.87 Degrees of Freedom (df) = 29
obtained t = 4.40 tabulated t = 2.045

Conclusion

Alternative hypothesis retained at 0.001 level for Format 1.
Alternative hypothesis: There is a difference between the two procedures.

27

T-TESTS between PROCEDURES (Format 2: MC-OE)

Formula employed: t for difference between correlated pairs of means
(See the previous page)

Subject	OE	MC	d	d^2	Subject	OE	MC	d	d^2
1	26	31	−5	25	16	20	23	−3	9
2	29	29	0	0	17	21	22	−1	1
3	30	29	1	1	18	21	22	−1	1
4	28	28	0	0	19	25	22	3	9
5	29	27	2	4	20	23	21	2	4
6	28	26	2	4	21	21	21	0	0
7	27	26	1	1	22	22	21	1	1
8	25	26	−1	1	23	20	20	0	0
9	27	27	0	0	24	22	20	2	4
10	16	25	−9	81	25	24	20	4	16
11	25	24	1	1	26	19	20	−1	1
12	28	24	4	16	27	21	18	3	9
13	27	24	3	9	28	18	16	2	4
14	22	24	−2	4	29	17	16	1	1
15	25	24	1	1	30	22	16	6	36

$\Sigma d = 16 \quad \Sigma x_d^2 = 244 \quad X_d = 0.53 \quad df = 29$
obtained t = 1.01   tabulated t = 2.045

Conclusion:

Alternative hypothesis rejected for Format 2.

Alternative hypothesis: There is a difference between the two procedures.

ITEM ANALYSIS (FORMAT 1:OE-MC)

	Multiple Choice						Open-Ended					
Item	U	L	U-L	ID	U+L	FV	U	L	U-L	ID	U+L	FV
1	8	8	0	0	16	0.8	8	6	2	0.2	14	0.7
2	6	6	0	0	12	0.6	7	7	0	0	14	0.7
3	8	8	0	0	16	0.8	9	8	1	0.1	17	0.85
4	5	3	2	0.2	8	0.4	6	1	5	0.5	7	0.35
5	8	2	6	0.6	10	0.5	7	4	3	0.3	11	0.55
6	6	5	1	0.1	11	0.55	7	3	4	0.4	10	0.5
7	4	4	0	0	8	0.4	5	5	0	0	10	0.5
8	7	3	4	0.4	10	0.5	6	3	3	0.3	9	0.45
9	8	4	4	0.4	12	0.6	5	1	4	0.4	6	0.3
10	2	2	0	0	4	0.2	4	2	2	0.2	6	0.3
11	7	5	2	0.2	12	0.6	6	5	1	0.1	11	0.55
12	7	7	0	0	14	0.7	9	5	4	0.4	14	0.7
13	10	8	2	0.2	18	0.9	10	8	2	0.2	18	0.9
14	10	6	4	0.4	16	0.8	8	3	5	0.5	11	0.55
15	6	6	0	0	12	0.6	10	7	3	0.3	17	0.85
16	10	10	0	0	20	1	10	10	0	0	20	1
17	8	5	3	0.3	13	0.65	9	7	2	0.2	16	0.8
18	9	7	2	0.2	16	0.8	10	9	1	0.1	19	0.95
19	9	2	7	0.7	11	0.55	10	7	3	0.3	17	0.85
20	6	4	2	0.2	10	0.5	6	4	2	0.2	10	0.5
21	10	7	3	0.3	17	0.85	10	10	0	0	20	1
22	9	6	3	0.3	15	0.75	5	5	0	0	10	0.5
23	7	6	2	0.2	13	0.65	4	4	0	0	8	0.4
24	5	1	4	0.4	6	0.3	2	2	0	0	4	0.2
25	7	3	4	0.4	10	0.5	5	5	0	0	10	0.5
26	9	6	3	0.3	15	0.75	10	4	6	0.6	14	0.7
27	3	3	0	0	6	0.3	5	2	3	0.3	7	0.35
28	2	1	1	0.1	3	0.15	2	1	1	0.1	3	0.15
29	6	1	5	0.5	7	0.35	6	4	2	0.2	10	0.5
30	5	1	4	0.4	6	0.3	5	3	2	0.2	8	0.4
31	7	2	5	0.5	9	0.45	6	2	4	0.4	8	0.4
32	4	3	1	0.1	7	0.35	5	2	3	0.3	7	0.35
33	4	3	1	0.1	7	0.35	7	4	3	0.3	11	0.55
34	8	4	4	0.4	12	0.6	7	4	3	0.3	11	0.55
35	7	2	5	0.5	9	0.45	9	5	4	0.4	14	0.7
36	8	3	5	0.5	11	0.55	7	7	0	0	14	0.7
37	10	8	2	0.2	18	0.9	9	8	1	0.1	17	0.85
38	6	3	3	0.3	9	0.45	7	3	4	0.4	10	0.5
39	7	6	1	0.1	13	0.65	10	8	2	0.2	18	0.9
40	8	7	1	0.1	15	0.75	6	5	1	0.1	11	0.55

U = Upper Group L = Lower Group ID = Item Discrimination FV = Facility Value

11 MAR 89   SPSSK-X RELEASE 1.0   PRIME
              Univ. of Qatar Computer Centre Prime 2250   Rev

GROUP: 1.00

NUMBER OF VALID OBSERVATIONS (LISTWISE) = 30.00

VARIABLE	MEAN	STD DEV	MINIMUM	MAXIMUM	VALID N
MC	23.233	4.345	12.000	31.000	30
OE	17.367	3.746	10.000	26.000	30

11 MAR 89   SPSS-X RELEASE 1.0   PRIME
              Univ. of Qatar Computer Centre Prime 2250   Rev
19.1.A

GROUP: 2.00

NUMBER OF VALID OBSERVATIONS (LISTWISE) = 30.00

VARIABLE	MEAN	STD DEC	MINIMUM	MAXIMUM	VALID N LABEL
MC	23.067	3.947	16.000	31.000	30
OE	23.600	3.820	16.000	30.000	30

App. (7)

T-TESTS between OPEN-ENDED PROCEDURE ON DIFFERENT FORMATS

FORMULA EMPLOYED: t for differences between uncorrelated pairs of means.

$$\left[\frac{\bar{X}_A - \bar{X}_B}{\dfrac{\{\Sigma X_A^2 - (\Sigma X_A)^2/N_A\} + \{\Sigma X_B^2 - (\Sigma X_B)^2/N_B\}}{(N_A-1) + (N_B-1)} \times \left(\dfrac{1}{N_A} + \dfrac{1}{N_B} \right)} \right]$$

Null hypothesis: There is no difference between the same procedure on different formats with different order.

Summary of Input:

Procedure	N	$\underset{\sim}{\Sigma} X$	$\underset{\sim}{\Sigma} X^2$	$(\underset{\sim}{\Sigma} X)^2$	X
OE Format 1	30	521	9455	271441	17.37
OE Format 2	30	708	17132	501264	23.6

30

Result:

Obtained t = 6.36
DF = 58

Tabulated t = 2.000 at the 0.05 level of significance

Conclusion:

The null hypothesis was rejected at 0.001 level of significance for the OE procedure. Therefore, the order of the procedures (IV) *had an effect* on the scores (DV).

<div align="right">

App. (7)

</div>

T-TESTS between MULTIPLE CHOICE PROCEDURE on DIFFERENT FORMATS

Formula employed: t for differences between uncorrelated pairs of means. (See the previous page)

Null hypothesis: No difference between the same procedure on different formats with different orders.

Summary of Input:

Procedure	N	$\sum X$	$\sum X^2$	$(\sum X)^2$	X
MC Format 1	30	697	16741	485809	23.23
MC Format 2	30	692	16414	478864	23.07

Result:

Obtained t = 0.158
DF = 58

Tabulated t = 2.000 at the 0.05 level of significance

Conclusion:

The null hypothesis was retained for the MC procedure. The order of the procedures (IV) *had no effect* on the scores (DV).

31

11 MAR 89   SPSS-X RELEASE 1.0   PRIME
Univ. of Qatar Computer Centre Prime 2250   Rev. 19.1.A

GROUP: 1.00

T-TEST

VARIABLE	NUMBER OF CASES	MEAN	STANDARD DEVIATION	STANDARD ERROR	(DIFFERENCE) MEAN	STANDARD DEVIATION	STANDARD ERROR	CORR.	2-TAIL PROB.	T VALUE	DEGREES OF FREEDOM	2-TAIL PROB.
MC		23.2333	4.345	0.793								
	30				5.8667	4.216	0.770	0.465	0.010	7.62	29	0.000
		17.3667	3.746	0.684								
OE												

11 MAR 89   SPSS-X RELEASE 1.0   PRIME
Univ. of Qatar Computer Centre Prime 2250   Rev 19.1.A.

GROUP: 2.00

T-TEST

VARIABLE	NUMBER OF CASES	MEAN	STANDARD DEVIATION	STANDARD ERROR	(DIFFERENCE) MEAN	STANDARD DEVIATION	STANDARD ERROR	CORR.	2-TAIL PROB.	T VALUE	DEGREES OF FREEDOM	2-TAIL PROB.
MC		23.0667	3.947	0.721								
	30				−0.5333	2.849	0.520	0.731	0.000	−1.03	29	0.314
		23.6000	3.820	0.697								
OE												