CLOZE TESTING AS AN ALTERNATIVE TO THE CONVENTIONAL EXAM IN E.B.E.

Honesto HERRERA SOLER
Universidad Complutense

INTRODUCTION

Since Wilson Taylor (1953, 1956) designed the cloze technique to measure the reading capability of native English speakers much has been written on cloze tests. Its use has spread not only among students of General English but also among students of English for Specific Purposes. In testing literature this technique is classified as an integrative test and considered as a reliable and valid tool to assess the readibility of texts, although Alderson (1980) has questioned the exactness of cloze as a diagnostic instrument and, especially, its purported predictive capacity. Others researchers like Hatch (1979), Valmont (1983) and Brown (1986) have also discussed its potential as a teaching reading device.

During these four decades many hypotheses to study the different variables that intervene in the cloze procedure have been examined. Most of them take the deletion and assessment criteria as points of reference: whether one has to be "strict" or "flexible" in either of them. In spite of that, every other day new studies appear and the deletion and assessment systems are still open to new approaches. And it is one of them we are concerned with in this article.

There also exists considerable literature, Bachman (1982), Brown (1983), Alderson (1983) and Porter (1983) on whether there is sensitivity in cloze items across sentences or not. As in this study sequential and not scrambled texts are considered, the points of view of Chavez-Oller (1985) and Jonz (1991) are taken into account. The former demonstrated that about 10% of the items in each of the two texts examined were highly sensitive to constraints ranging across sentence boundaries while the latter confirmed that textually cohesive cloze-items showed a significant sensitivity to sentence scrambling.

Empirical findings on the item analysis and selection processes are described in Brown (1988). He investigates the possibility that the reliability and validity of a cloze procedure be improved with a "taylored cloze," taking into account the facility and discrimination indices. Brown (1989) is also concerned with the cloze item difficulty. On this occasion, he studies the linguistic features of cloze tests items in every 12th deleted word of randomly selected passages. Brown explores the effects on the scores of content/function words according to their appearance in the text. In this study the ease/difficulty of each item is also considered but from a different point of view, so two cloze tests using the rational deletion procedure rather than the fixed-ratio one were administered, trying to find out in each case whether students of English for Business and Economics (EBE) not only had more or less dificulty with content or function words but also within the content words with nouns or verbs, and within the function words between determiners or connectors.

The last research to be mentioned is Butler's contribution (1991) to the use of authentic material with the concordance-generated tests. It seems to be very relevant although Taylor's original idea of working with a complete text as a whole fades away if only scrambled texts are tested. In spite of that, his contribution is seen as a new device on which further research must be done in the near future. For the time being, the investigation reported is considered to be very useful for the classroom teacher since it deals with four individual sentences or propositions and that means more information and consequently more probabilities to recover the suitable term. As this article deals with testing rather than with methodological tools and the access to an EBE Corpora is unavailable at the moment, Taylor's patterns are followed.

Summarising then, the research done by Bachman (1992), Valmont (1983), Brown (1983, 88, 89), Henk and Helfeldt (1985), Chavez-Oller et al. (1985), Chapelle et al. (1990) and Butler (1991) is taken into account. But whereas all of them take texts to work with from General English, in this investigation the texts are taken from English for Business and Economics.

The overall purpose of this study is to explore the possibility of developing alternative tests to the conventional exams, with the purpose of avoiding, on the one hand, biased interpretations, as happens when we work with comprehension questions, and on the other, to find a faster, more accurate and fairer way to correct the exams, especially if there are hundreds of students taking the same exam and several teachers marking them. To that end the following research questions are examined:

1. To what extent can a cloze be considered an alternative to the conventional exams in EBE?

2. To what degree are content words easier or more difficult than function words? The same question can be asked about nouns and verbs within the lexical items. And what happens when the student deals either with determiners or connectors within the grammatical items?

3. Does it make any difference to give the first letter of the word being sought in a classical fixed-ratio procedure on an EBE intermediate level cloze test performance?

METHOD

1. Subjects

The sample consisted of 69 non-native students (27 males and 42 females) in their first EBE course at the Faculty of Economics in Complutense University, Madrid. The group of students weren't volunteers nor randomly selected from the enrolled students in this course but naturally-occurring groups since they were the students who on the date that the cloze tests were administered, attended their class without any previous notice.

2. Materials

Much of the work is done in the classroom with English for Academic Purposes Series Books, readings from the *International Management,* and *The European'* s business pages since the objectives of this course are focused on technical and semi-technical lexical items, collocations and premodifications in EBE rather than on text organization. So the cloze tests were based on passages taken from *Economics* by Yates C.St.J. (1989), (Appendix). The headline and the first sentence/paragraph of each passage were left intact in each cloze before starting the chosen deletion pattern. In the first two, cloze A (CLA) and cloze B (CLB), as the aim was to study the effect on scoring of either content or function words, a rational deletion scheme was carried out with blank spaces within a range of 4 to 14 words. Whereas in the third and the fourth ones, cloze C (CLC) and cloze D (CLD), a fixed-ratio deletion pattern was followed and every eighth word was

removed. In the latter, the first letter was written down in each blank. So a set of four cloze tests was delivered.

A total of 25 items were replaced by standard length blanks in each test. If in the first one, (CLA), the students were supposed to recover either nouns, verbs or modifiers, in the second one, (CLB), the attention was centered on determiners, pronouns, connectors, auxiliary and modal verbs, that is, function words. In the other two, (CLC) and (CLD), the students had to deal with the deleted terms no matter what kind of word it was (Table 4).

3. Procedure

Two weeks were allowed between the administration of the cloze tests and the first term conventional exam (CEXAM) at the Faculty of Economics, which took place in February, and included a reading comprehension text (TEX), and exercises on synonyms / antonyms (SYN), partnership words (PART) and fill in the blanks (FILL). This design has to be taken into consideration since the first research question settled above is to study if there is a correlation between this sort of standardised exam and the cloze testing presented as an alternative.

Two models of presentation were designed changing the order of each type of cloze, and were distributed to every other student to avoid cheating. They were allowed 1 hour to take the four 25-item tests.

The scoring was done by taking into account a flexible interpretation of the terms with which the students had filled in the blanks. Any word which could be considered suitable to recover the text was accepted, as there was not much concern with stylistic EBE texts. Thus, several options to express the same concept were admitted, a decision that can be easily understood whenever one comes across any EBE topic. There are different ways, all of them aceptable, to express the same idea as it happens, for instance, talking of the Stock Exchange with the idea of "rise and fall".

Instructions: The students were told

1. That taking the cloze tests was voluntarily.
2. That their scoring would be counted only if it were better than the one of the conventional exam.

4. Analysis

In the descriptive statistics central tendency and variability measures are primarily considered.

The degree of relationship between each cloze test and each conventional subtest was calculated using the Pearson product-moment correlation coefficient. The same was done with the subtests of the four types of clozes presented.

The T-test was used to determine whether or not there were significant differences within cloze tests and between the cloze tests taken as a whole and the conventional exam as such. Significant differences in the following pairs: content versus function words, nouns versus verbs, and determiners versus connectors were also calculated with the t-test.

The significance level was set at $\alpha<0.05$.

The data were studied with the DBASE III PLUS and with the BMDP Statistical Software.

RESULTS

A description of the results, which shows the central tendency and dispersion in the cloze tests and in the subtests of the conventional exam, is required in the first approach to the issue.

TABLE 1. DESCRIPTIVE STATISTICS

V. NAME	MEAN	MEDIAN	MODE	SD	SKEWNESS
CLA	5.4261	5.1999	4.8000	1.5530	-0.28
CLB	4.8986	5.1999	5.1999	1.9090	-0.35
CLC	5.5246	5.5999	5.1999	1.6063	-0.25
CLD	7.4899	7.5999	N.UNIQUE	1.4595	-0.64
TEX	6.1232	6.5000	6.5000	1.5235	-1.05
SYN	5.9565	6.0000	N.UNIQUE	2.3542	-0.42
PART	5.4348	5.0000	5.0000	1.9590	-0.06
FILL	5.7391	6.0000	6.0000	2.2206	-0.33
CLOZE	5.8478	6.0000	N.UNIQUE	1.3866	-0.47
CEXAM	5.9145	6.4000	6.8000	1.3482	-0.53

It is easy to appreciate in Table 1, that the analysed indicators of central tendency have almost the same score value and all but two are found near the center, slightly negatively skewed, in a normal distribution. Leaving aside the cloze where we gave the first letter (CLD), the means of the other cloze tests and the four subtests range from 4.8986 in the CLB to 6.1232 in the TEX. Similar ranges are found in the median and mode values. Taking the mean of all cloze tests "CLOZE" and the four subtests which form the conventional exam (CEXAM), there is only a difference of 0.0667.

The standard deviation in all but the TEX subtest is higher in the conventional exam than in each of the cloze tests. That means that the dispersion is greater in the conventional one.

As it has been mentioned above, the first issue to investigate in this study is whether there is some sort of relationship between the different subtests of the conventional exam and the different cloze tests. The results shown in the correlation matrix (Table 2) confirm that there is a strong significant relationship, on the one hand, among all types of cloze tests and subtests of the CEXAM and, on the other, among all cloze tests. This relationship is higher than that which can be read among the different subtests of the CEXAM, in which the "PARTnership word exercise" shows the lowest correlations. Nevertheless if the cloze tests are taken together and the same is done with the four subtests of the CEXAM then, the empirical finding, 0.7884, confirms a strong relationship between the conventional exam and the alternative presented.

TABLE 2. CORRELATION MATRIX

	CLA	CLB	CLC	CLD	TEX	SYN	PART	FILL
CLA	1.0000							
CLB	0.6884	1.0000						
CLC	0.7144	0.5772	1.0000					
CLD	0.6034	0.4946	0.6041	1.0000				
TEX	0.5214	0.3310	0.5219	0.4063	1.0000			
SYN	0.4935	0.3557	0.4720	0.6269	0.5120	1.0000		
PART	0.4390	0.2982	0.2069	0.2598	0.1198	0.0456	1.0000	
FILL	0.6656	0.5723	0.6112	0.5519	0.4965	0.4535	0.2868	1.0000

*Critical value 1-tailed Signif, $p<0.05= .2319$

The purpose of the second research question set at the beginning was to study if it was easier or more difficult for the student of EBE to fill in the

blanks in which they had to deal either with content (Nouns —NOU—, verbs —VER—, modifiers —MOD—) or with function (Determiners —DET—, connectors —CON—, auxiliary and modals —AMO—) words. So a comparison between the first two types of cloze is required (Table 3).

The statistical analysis indicates that the observed differences in means in the lexical cloze test versus the grammatical cloze text were due to factors other than chance with an $\alpha<0.05$.

TABLE 3. SIGNIFICANCE OF DIFFERENCES IN THE FIRST TWO CLOZE TESTS

CLA vs CLB

TEST STATISTICS		P-VALUE	DF
MATCHED T	3.12	0.0027	68
CORRELATION	0.6884	0.0000	68

We find for a p-value of 0.0027 a "t" test = 3.12. Thus the null hypothesis that the content and function words have the same level of difficulty is rejected.

Further analysis is shown in the following table. The number of items in each category, the means of errors, since it is thought to be more enlightening, and their percentages are presented.

TABLE 4. DISTRIBUTION OF ITEMS IN EACH CLOZE. MEAN OF ERRORS IN EACH CATEGORY

	CLA			CLB			CLC		CLD	
	NOU	VER	MOD	DET	CON	AMO	CON	FUN	CON	FUN
_										

N	14	6	5	11	10	4	18	7	13	11
$\bar{X}$	6	1	2	3	4	2	6	2	2	2
%	43	17	40	27	40	50	33	29	15	18

T-test carried out among the different pairs has proved that there are significant differences between the recovery of nouns and verbs (A-NOU vs A-VER), determiners and connectors (B-DET vs B-CON), content and function words (C-CON vs D-CON, and C-FUN vs D-FUN) according to the system of deletion as can be appreciated in table 5.

TABLE 5. SIGNIFICANCE OF DIFFERENCES BETWEEN THE CATEGORIES STUDIED

	TEST STATISTICS	p-VALUE	DF
A-NOU vs A-VER	13.90	0.0000	68
B-DET vs B-CON	-7.60	0.0000	68
C-CON vs D-CON	11.38	0.0000	68

C-FUN vs D-FUN	4.86	0.0000	68

While the last two pairs, CCON vs DCON and CFUN vs DFUN, established in this table show that there were significant differences among them, results that back up the ones that can be read in Table 6, in which scores not errors are considered. No significant differences, perhaps because of the sample, are found between lexical and grammatical performance within each of the fixed-ratio cloze (Table 4). These results lead us to consider other factors than the difficulty of function or content words as the explanation of the significant difference between the CLC and the CLD tests.

TABLE 6. SIGNIFICANCE OF DIFFERENCES IN THE CLC AND CLD.

	CLC VS. CLD	
TEST STATISTICS	P-VALUE	DF
MATCHED T -11.91	0.0000	68
CORRELATION 0.6041	0.0000	68

The statistical analysis confirms in this case that the differences in means between the fixed-ratio deleted cloze tests with or without cue letters, were due to factors other than chance with an $\alpha<0.05$.

Finally, I want to point out that considerable significant differences are found if we compare CLD versus CLA or CLB. They are even stronger if we take CLC versus CLD, T-test = -11, which leads us to say that it makes a difference if we give the initial letter in a fixed-ratio deletion procedure whether errors or scores are considered, and that this sort of cloze, the cued one stands quite apart from the conventional clozes.

DISCUSSION

1. Correlation between the conventional comprehension reading test and the alternative presented.

The results reported above in the correlation matrix support the existence of a good relationship between the alternative presented, that is, the cloze battery and the conventional exam, among students of EBE. The figures read in Table 2 confirm that the correlation is stronger among the different sorts of clozes than that found among the subtests of the conventional exam. It is the PARTnership words test which offers the poorest correlations with the cloze tests and doesn't reach the level of significance with the other subtests. A possible explanation for these anomalous figures could be that this subtest was introduced for the first time in the CEXAM and the students didn't master the item-matching technique.

On the whole, this battery of EBE cloze tests with some modification could fulfil the same objectives, even in a better way, than the conventional exam. It mustn't be forgotten that the purpose of the latter is to measure the students' capability of reading something for which the cloze was originally designed.

2. Is the students' performance better in the lexical or in the grammatical items?

The means of CLA and CLB are compared and the differences between them are found to be significant. A synonymic rather than a verbatim scoring has been preferred in the assessment, not so much for methodological reasons, but for the characteristics of these sorts of texts. They are open to several synonyms, especially in the content words.

The students' performance in the CLA, in which lexical items had been eliminated, was better than in the CLB, in which we weren't as strict as Levenston et al. (1982) in their "discourse cloze," and all kinds of grammatical words were deleted. That supports Cohen's findings (1984) on this issue in his studies with economic and business students. These results should lead us to consider that EBE students have more problems with grammar than with content words.

The organisation of the discourse in which determiners, prepositions, conjuctions and to some extent auxiliary verbs are needed, was more difficult to cope with than trying to recover items such as: "business, goods, money, accounts, prefer or satisfy." As a possible explanation of this performance it

could be pointed out that the students have a good deal of information on the different topics that usually appear in the EBE texts. So it must be assumed that they use the knowledge they have of the subject matter to make inferences.

—Which items create more problems to EBE students? To what extent?

To answer these questions it is necessary to refer to Tables 4 and 5. In a first approach to Table 4 it seems that the distribution of items to the different categories is quite normal in all cases except the CLC, in which the fixed-ratio deletion system has given 18 content words and 7 function words. In order to know whether these distributions are normal or not, further research on EBE Corpora is needed. In the second issue the means of errors is quite different. Leaving aside the auxiliary and modals because the number of items in its distribution is quite low, the differences in the percentages among nouns and verbs, connectors and determiners point out that EBE students have fewer problems with verbs in the first cloze and with determiners in the second one.

Going through the different categories in each cloze, it is in CLA, perhaps because of its design, where more options for each blank have been found. There are more synonyms within the verbs and fewer mistakes. Only on one occasion when we read, "it might be more profitable to *sell* hamburgers...," no other options than the original term are registered. It does not happen the same in the other blanks to be filled with verbs, as in "... suppose you *buy* a hamburger," where *choose, have, take, like, want* and *get* are registered as acceptable options. Synonyms are not so numerous with nouns and modifiers. Only in the following case: "If rents were higher, it might be more profitable to sell hamburgers in a cheaper area or to switch to luxury *lunches* for rich executives on expense accounts," such items as *areas, restaurants, cafes, meals* have been taken into consideration.

On the whole it can be said that in the responses related to any part of the nominal phrases the range of options is far narrower, some of them are bound because of the lexical collocation, than if the item to be recovered is connected with the verbal phrase.

If CLB is taken into consideration the data also show that connectors, whether conjunctions or prepositions, offer more difficulty than determiners. It's worth noting in this cloze test that in the students' performance in the grammatical items almost no other options are found. There is only room for some considerations in the "determiners": the, this for "that," or among some disjuncive or additive connectors.

3. Does it make any difference to give the initial letter?

The study of the third research question CLC vs CLD has been done from two different points of view. On a first approach, taking the cloze tests as such the T-test = -11.91, shows a significant difference. Obviously the help the student has with the initial letter is quite remarkable. Even if the CLD is compared to CLA and CLB, in which a different deletion procedure has been followed, the differences are still significant. Besides if the distribution of scores is examined in the CLD, then a considerable negative skewness -0.64 is observed. The figure has to be interpreted as an indicator of a quite high facility index for a norm referenced test. So it is quite evident that this deletion system, not far from other proposals: Braley-Klein (1985a), Grotjahn (1987) and Kokkota (1988), makes a difference if it is compared to the rest of the cloze test battery.

The second approach to the third research question is done from the errors perspective. If the content words of CLC and CLD are examined, and the same is done with the function words, there also appear significant differences. Consequently, giving the initial letter also has an effect. In this comparison it is also observed there is no difference between lexical and grammatical items. Apparently, this finding isn't consistent with the one of the second research question, in which it was said that EBE students had more problems with function than with content words.

Nevertheless, the deletion procedure could be one explanation. A rational elimination pattern affects the reading comprehension, both from a lexical or a grammatical point of view, more than a fixed-ratio deletion does, since there is a tendency in the former system to eliminate most of the lexical items which are usually labelled as semi-technical or function words. Whereas if the student has to cope with a fixed-ratio deletion pattern therer are fewer problems with comprehension because not all deleted items are either lexical or grammatical, so the economic discourse is a little more balanced and consequently it is easier to follow. Furthermore, the difficulty of the grammatical items, above mentioned, almost disappears since the discourse is a little more organised and it is not as difficult as before to infer the missing function words.

CONCLUSIONS

However, although more research is needed to study other possible cloze test batteries for EBE students, there seems to be no reason, if the correlation coefficients are examined, why the first three cloze tests cannot be considered an alternative to the conventional exam. Another relevant insight in this investigation is that EBE students have more problems with grammatical than with lexical performance and within the former with connectors rather than with determiners, whereas in the lexical performance they fail more in their noun-related responses. And finally, the last finding to be pointed out is that a cued cloze test is far easier than the other three.

For these reasons, we can conclude that a battery of cloze tests formed by a combination of the two types of rational deletion and another with a random deletion procedure represent a promising alternative to the conventional comprehension reading test.

Acknowledgments

I am indebted and grateful to Rosario Martínez Arias, Methodology and Research Chair at the Faculty of Psycology, who has been of great help with statistics.

REFERENCES

ALDERSON, J. C., and A. H. URQUHART. 1984. "ESP Tests: the Problem of Student Background Discipline." *Occasional Papers* 24 (U of Essex, Dpt. of Language and Linguistics; autumn): 1-13.

BACHMAN, L. F. 1985. "Performance on Cloze Tests with Fixed-ratio and Rational Deletions." *TESOL Quaterly* 19: 535-55.

BROWN, J. D. 1986. "Cloze Procedure: A Tool for Teaching Reading." *TESOL Newsletter* 20.5: 1, 7.

- - -. 1988. "Tailored Cloze: Improved with Classical Item Analysis Techniques."*Language Testing* 5.1 (January): 19-31.

- - -. 1989. "Cloze Item Difficulty." *JALT Journal* 11.1 (May): 46-67.

BUTLER, J. 1991. "Cloze Procedures and Concordance: The Advantages of Discourse Level Authenticity in Testing Expectancy Grammar." *System* 19.1/2: 29-38.

CHAPELLE, Carol A., and Roberta G. ABRAHAM. 1990. "Cloze Method: What Difference Does It Make?" *Language Testing* 7.2: 121-146.

CHAVEZ-OLLER, M. A., T. CHIHARA, K. A. WEAVER and J. W. OLLER, Jr. 1985. "When Are Cloze Items Sensitive to Constraints across Sentences?" *Language Learning* 35: 181-206.

COHEN, A. D. 1984. "On Taking Language Tests: What the Students Report." *Language Testing* 1: 70-81.
GROTJHAN, R. 1987. "How to Construct and Evaluate a C-Test: A Discussion of Some Problems and Some Statistical Analyses." *Quantitative Linguistics* 34: 219-253.
HATCH, E. 1979. "Reading a Second Language." In *Teaching English as a Second or Foreign Language.* Ed. M.Celce-Murcia and L. McIntosh. Rowley (MA): Neubury House. 128-44.
HENK, W. A. and J. P. HELFELDT 1985. "Effects of Alternative Deletion Patterns, Blank Conditions, and Scoring Systems on Intermediate Level Cloze Test Performance." *Reading Psychology* 6.1-2: 85-96.
JONZ, J. 1991. "Cloze Item Types and Second Language Comprehension." *Language Testing* 1: 1-22.
KLEIN-BRALEY, C. 1985. "A Cloze-up on the C-test: A Study in the Construct Validation of Authentic Tests." *Language Testing* 2:76-104.
KOKKOTA, V. 1988. "Letter-deletion Procedure: A Flexible Way of Reducing Text Redundancy." *Language Testing* 5.1: 115-119.
LEVENSTON, E. A., R. NIR, and S. BLUM-KULKA. 1982. "Discourse Analysis and the Testing of Reading Comprehension by Cloze Technique." Paper presented at the International Symposium of Language for Special Purposes, Eindhoven, The Netherlands.
TAYLOR, W. L. 1953. "Cloze Procedure: A New Tool for Measuring Readability." *Journalism Quaterly* 30: 414-38.
VALMONT, W. J. 1983. "Cloze Deletion Patterns: How Deletions are Made Makes a Big Difference." *The Reading Teacher* 37.2: 172-5.

APPENDIX

Two excerpts of the four passages.

1.- Cloze A, CLA. Rational deletion procedure.

MARKETS

Markets bring together buyers and sellers of goods and services.

Prices of goods and of (1) RESOURCES , such as labour, machinery and (2) LAND , adjust to ensure that (3) SCARCE resources are used to produce those (4) GOODS and services that society (5) DEMANDS . Much of economics is devoted to the (6) STUDY of how markets and prices enable (7) SOCIETY to solve the problems of what, (8) HOW , and for whom to produce. Suppose you

(9) BUY a hamburger for your lunch. What does this have to do with (10) MARKETS and prices?...

Cloze C, CLC. Fixed-ratio deletion.

SUPPLY AND DEMAND

In Chapter 1 we defined markets in a very general way as arrangements through which prices guide resource allocation. We now adopt a narrower definition.

A market is a set of arrangements by which buyers and sellers are in contact (1) TO exchange goods or services.

Some markets ((2) SHOPS and fruit stalls) physically bring together the (3) BUYER and the seller. Other markets (the London (4) STOCK Exchange) operate chiefly through intermediaries (stockbrokers) who (5) CONTRACT business on behalf of clients. In supermarkets, (6) SELLERS choose the price, stock the shelves, and (7) LEAVE customers to choose whether or not to (8) MAKE a purchase. Antique auctions force buyers to (9) BID against each other with the seller taking (10) A passive role.